Off-the-grid variational sparse spike recovery: methods and algorithms

Bastien Laville, Laure Blanc-Féraud, Gilles Aubert

Journal of *Imaging*

MDPI

Review

Off-the-grid variational sparse spike recovery: methods and algorithms.

Bastien Laville[1], Laure Blanc-Féraud[1] and Gilles Aubert [1,2].

[1] Université Côte d'Azur, CNRS, Inria, I3S, Morpheme project, France.
[2] Université Côte d'Azur, CNRS, LJAD, France.

Abstract: Gridless sparse spike reconstruction is a rather new research field with significant results for the super-resolution problem, where we want to retrieve fine-scale details from a noisy and filtered acquisition. To tackle this problem, we are interested in optimisation under some prior, typically the sparsity *i.e.* the source is composed of spikes. Following the seminal work [1–4] on the generalised LASSO for measures called the *Beurling-Lasso* (BLASSO), we will give a review on the chief theoretical and numerical breakthrough of the off-the-grid inverse problem, as we illustrate its usefulness to the super-resolution problem in *Single Molecule Localisation Microscopy* (SMLM) through new reconstruction metrics and test on synthetic and real SMLM data we performed for this review.

Keywords: Off-the-grid optimisation review; inverse problems; sparse spike localisation; super-resolution; fluorescence microscopy; SMLM; functional analysis.

Citation: Laville, B.; Blanc-Féraud, L.; Aubert, G. Off-the-grid variational sparse spike recovery: methods and algorithms.. *J. Imaging* **2021**, *7*, 266. https://doi.org/10.3390/jimaging7120266

Academic Editor: Fabiana Zama, Elena Loli Piccolomini

Received: 21 October 2021
Accepted: 28 November 2021
Published: 6 December 2021

1. Introduction

In this paper, we propose to conduct a comprehensive review on the so-called *off-the-grid variational methods* to solve the sparse spike recovery problem. We will exhibit the main theoretical and numerical results in the literature, underlining the interest of these methods for various domains dealing with inverse problems. As part of this review and our former work on gridless methods, we developed an implementation of the more consistent numerical methods with a focus on efficiency and computation time. With this implementation, we were able to apply off-the-grid method to fluorescence microscopy super-resolution problem. The codes and the computed result are an addition to the off-the-grid literature, and constitute further evidence supporting the relevance of this domain in inverse problem field.

Loosely speaking, inverse problems consist in the reconstruction of the causes from the consequences. The problem is generally ill-posed, meaning that existence, uniqueness, and stability of a solution(s) is (are) not guaranteed. A case arising in numerous fields such as image or signal processing, telecommunications, machine learning, super-resolution, *etc.* is the *sparse spike problem*. It consists in the reconstruction of spikes located on a domain $\mathcal{X}$ from an acquisition y, with the prior of sparsity on the cause; or in layman terms, the source is composed of a few spikes. This includes sources such as stars in astronomy, fractures in seismology, *etc.* A spike is typically modelled by a Dirac measure $a\delta_x$ with amplitude $a \in \mathbb{C}$ and position $x \in \mathcal{X}$. All the difficulty lies in the estimation of the number N of spikes, of their amplitudes $(a_i)_{i=1}^N$ and their positions $(x_i)_{i=1}^N$. Hence, the goal is to reconstruct the measure $m = \sum_{i=1}^N a_i \delta_{x_i}$ only from a few number of observations y in a Hilbert space $\mathcal{H}$ (typically $\mathrm{L}^2(\mathcal{X})$) linked to m through an operator Φ accounting for deterioration of the input (blur, downsizing by the sampling) such as $y \overset{\text{def.}}{=} \Phi m + w$ where $w \in \mathcal{H}$ is an additive noise. The reconstruction of the spikes may be *off-the-grid i.e.* the positions $(x_i)_{i=1}^N$ are not constrained on a grid hence $(x_i)_{i=1}^N$ are not limited to a finite set of values: this allows interesting new mathematical insights and guarantees for the reconstruction, at the cost of some challenges for the numerical implementation. The general sparse spike problem is encountered in many situations, such as:

- compressed sensing domain [5], where one wants to recover a s-sparse vector $v \in \mathbb{C}^N$ from M measurements Av_0 where $A \in \mathbb{C}^{M\times N}$;
- machine learning, sketching mixtures, *etc*. For example, we desire to fit a probability distribution with respect to given data. The point is to estimate parameters $(a_i) \in \mathbb{R}^N$ and $(x_i) \in \mathcal{X}^N$ of a mixture $\sum_{i=1}^N a_i \varphi(x_i)$ of N elementary distributions described by φ. For instance, one wants to retrieve the means $\mu_i \in \mathbb{R}$ and standard deviations $\sigma_i \in \mathbb{R}^+$ of a Gaussian mixture, see [6] for more insights on this question;
- deep learning such as training neural networks with a single hidden layer [7];
- signal processing, for instance low rank tensor decomposition for *Direction of Arrival* estimation through sensor array (multiple sampling points);
- super-resolution, a rather central problem in image processing. Roughly speaking, it consists in the reconstruction of details from an altered input of signal/image. It includes classic physical operator of acquisition such as Fourier measurements, Laplace transform or Gaussian convolution.

The latter item will be our case of interest in the sparse spike problem for this paper. All the difficulty stems from the degradation in the acquisition process, which entails in general two things: a deterioration by the system of acquisition, typically modelled by the *Point Spread Function* in imagery which acts as a low-pass filter sensor acquisition which results in sampling and pollution by noise of different types, characterised by densities such as Gaussian, Poisson, *etc*. To sum-up, we want to reconstruct the correct number of spikes with correct amplitudes and positions in the continuous setting from a noisy and filtered discrete acquisition. It can be tackled from the theoretical point of view by either the variational approach or the Prony's method:

- Prony's method and its variants[1] which recover the signal source from Fourier measurements in a noiseless 1D setting. It consists in the decomposition of the signal onto a basis of exponentials with different amplitudes, damping factors, frequencies and phase angles to match the observed data. The results are compelling in the 1D noiseless case, and can be extended to a multivariate and noisy context; but still these methods lack of versatility since they cannot be sometimes extended to the context of interest. Thus, we will not consider this approach in this paper;
- variational approach which does not impose any particular structure on the acquisition operator, which can be adapted to any type of noise and does not need any prior on the number of point sources [8]. The key idea is to solve the inverse problem by finding among all possible signal sources the one minimising an objective function called *the energy*, formulated as a trade-off between a fidelity data term and a regularisation term, typically enforcing the sparsity prior here.

Then, there are two types of variational approaches: the discrete and the off-the-grid. In the discrete setting, one seeks to recover the spikes on a prescribed fine grid, typically with more points than the acquisition image. Indeed, we call coarse grid for the low-resolved acquisition, and fine grid for the finer (by a so-called super-resolution factor $q \in \mathbb{N}^*$) grid of the reconstruction. Thus, it consists in a finite dimensional problem, where the positions of the spikes must lie on a grid $\mathcal{G}$ of L points meshing the domain $\mathcal{X}$. This problem is a problem of sparse vectors reconstruction, and it can be tackled by enforcing sparsity through minimisation of the ℓ_1 norm of the unknown vector. This is known as the LASSO [9] or the Basis-pursuit problem, defined as the variational problem with tuning parameter $\lambda > 0$ controlling the trade-off between fidelity to the data and enforcement of the prior:

$$\min_{a\in\mathbb{R}^L} \underbrace{\|y - \Phi_L a\|_{\mathcal{H}}^2}_{\text{data term}} + \underbrace{\lambda \|a\|_1}_{\text{sparsity prior}} \tag{LASSO}$$

[1] such as MUSIC (MUltiple SIgnal Classification), ESPRIT (Estimation of Signal Parameters by Rotational Invariance Techniques) or Matrix Pencil.

where $\Phi_L : \mathbb{R}^L \to \mathcal{H}$ is the acquisition operator with a vector of size L as an input and $\mathcal{H}$ is a Hilbert space. A grid is useful to epitomise the concept of sparsity in the case of spikes: indeed, sparsity is just the fact that only a few points of the L grid have a non-zero value. Moreover, since a computer can only store array and vector quantities, it seems rather fair to work with finite dimensional problem; even for the theoretical analysis. However, how does one choose the discretisation? A grid with a step-size too small yields numerical instabilities [10] while choosing the step-size too large leads to round-off errors. Moreover, one would like to localise the spikes as precisely as possible without having to rely on a grid: a discretisation of positions would necessarily convey approximation on positions. The appropriate mathematical objects to get rid of these discretisation drawbacks is to represent a collection of spikes with Dirac measures, an element of the space of Radon measures $\mathcal{M}(\mathcal{X})$. The operator of acquisition is now $\Phi : \mathcal{M}(\mathcal{X}) \to \mathcal{H}$, the sparsity is enforced by a norm on $\mathcal{M}(\mathcal{X})$ called the TV-norm. This variational problem is called the BLASSO (for Beurling LASSO):

$$\min_{m \in \mathcal{M}(\mathcal{X})} \underbrace{\|y - \Phi m\|_{\mathcal{H}}^2}_{\text{data term}} + \underbrace{\lambda \|m\|_{\mathrm{TV}}}_{\text{sparsity prior}} . \tag{BLASSO}$$

In this latter setting, the spikes can move continuously on the domain $\mathcal{X}$: a comparison between the discrete and the off-the-grid reconstruction is given in Figure 1. The off-the-grid setting can be seen as the limit of the discrete case with a finer and finer grid [11].

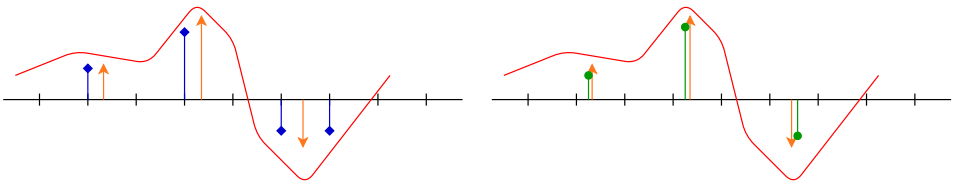

Figure 1. (**a**) Discrete reconstruction, which can be seen as spikes with support constrained on a grid (**b**) Off-the-grid reconstruction, the spikes can move continuously on the line. The red line is the acquisition y, orange spikes are the source (the cause we want to retrieve), blue spikes are discrete reconstruction constrained on a grid and green can move freely since it is off-the-grid. Note that when a source spike is between two grid points, two spikes will be recovered in the discrete reconstruction.

This shift from the discrete domain to the continuous setting called *off-the-grid* or *gridless* leads to some crucial mathematical insights, in particular a sharp signal-dependent criterion for stable spikes recovery [4], the *minimum* separation distance (see the next section). Obviously, some difficulties arise also due to the infinite dimension and the lack of algebraic properties of the set of optimisation. The comparison between discrete and gridless settings may be summed up by:

- the discrete problem is tackled by LASSO, through the minimisation of a convex function defined on a fine grid *i.e.* a convenient finite dimension Hilbert $\mathbb{R}^L$ space. Due to the ℓ_1 norm, there are some cases where the sparsity is not properly enforced: one can then replace the ℓ_1 norm by the non-continuous pseudo-norm ℓ_0, but this yields a NP-hard combinatory non-convex problem. There exists some continuous relaxation of ℓ_0 such as CEL0 [12], but due to the non-convex aspect the problem is still hard from the theoretical and numerical point of view. Despite the lack of guarantees, there are numerous algorithms to compute the solution of LASSO or its ℓ_0 relaxed variant;
- the off-the-grid problem is treated by BLASSO, a convex functional defined on $\mathcal{M}(\mathcal{X})$. The convex property is handy from the theoretical point of views as it leads to some crucial insights on the existence/uniqueness/support estimation w.r.t. noise, at the cost of the set of optimisation namely $\mathcal{M}(\mathcal{X})$ a Banach (no Hilbertian structure so no

straightforward proximal algorithm) infinite dimensional and non-reflexive space for the strong topology (convergence results are then essentially on the weak-$*$ topology). Despite these lack of algebraic properties, one has currently a wide range of algorithms to tackle this problem, such as root-finding or greedy algorithms.

Gridless reconstruction can then be evaluated through suitable metrics, namely the Flat Metric based on optimal transport of measures. This metric assesses the quality of the reconstruction and can be applied straightforwardly to off-the-grid and even discrete reconstruction outputs.

In the following, we give a review on the key results in the variational off-the-grid domain. The paper is organised in 3 sections, namely:

- the variational analysis of the space $\mathcal{M}(\mathcal{X})$, the properties and the guarantees of reconstruction concerning the sparse spike problem are now quite well-documented [1–4] and will be recalled in the theoretical section 2;
- multiple strategies were considered to numerically tackle BLASSO, the more compelling will be presented and put into context in the numerical section 3;
- interesting practical applications and new metrics have been considered for the gridless method, such as the SMLM super-resolution; these results are shown and discussed in section 4.

> At the end of each paragraph, a grey box (beginning either with 'summary' or 'shorthand') like this one will recall the main results highlighted in the section. Please refer to it for a quick summary.

2. A theoretical background for gridless spike recovery

In the following $\mathcal{X}$ denotes the ambient space where the positions of the spikes live. We suppose $\mathcal{X}$ is a subset of $\mathbb{R}^d$ such that its interior $\mathring{\mathcal{X}}$ is a submanifold of dimension $d \in \mathbb{N}^*$ [13]. This setting encompasses $\mathcal{X} = \mathbb{R}^d$, the torus $\mathcal{X} = \mathbb{T}^d \overset{\text{def.}}{=} \mathbb{R}^d/\mathbb{Z}^d$, any compact with non-empty interior, *etc.* The reader is invited to take a look at the Table A1 to remind the notations.

2.1. What is a measure?

As we have stated in the section above, the Dirac measure is the proper object to describe a spike not constrained on a finite set of positions. This object is not a function, since one cannot exhibit any integrable equivalence class satisfying the properties of the Dirac (see below). Thus, one should considerate the notion of Radon measure, a formal extension of functions. From a distributional standpoint, it is a subset of the distribution space $\mathscr{D}'(\mathcal{X})$, namely the space of linear forms over the space of test functions $\mathscr{D}(\mathcal{X})$ *i.e.* smooth functions (continuous derivatives of all orders) compactly supported. This functional approach[2] consists in the definition of a measure as a linear form on some function space, namely:

Definition 2.1.1 (Evanescent continuous function on $\mathcal{X}$)**.** *We call $\mathscr{C}_0(\mathcal{X}, \mathcal{Y})$ the set of continuous functions with zero at infinity (or evanescent), namely all the continuous map $\psi : \mathcal{X} \to \mathcal{Y}$ such that :*

$$\forall \varepsilon > 0, \exists K \subset \mathcal{X} \text{ compact}, \quad \sup_{x \in \mathcal{X} \setminus K} \|\psi(x)\|_{\mathcal{Y}} \leq \varepsilon.$$

[2] One can then define equivalently the space of Radon measures, either by a set-related approach or by functional analysis approach (thanks to Riesz–Markov theorem). In the more set-related [14] insight, a *measure* is an object which takes sets as an input. A *Borel measure* is a measure defined on all open sets of $\mathcal{X}$, and a *Radon measure* is a Borel measure such that it is finite on all compact sets of $\mathcal{X}$ (by an isomorphism). The functional and the set point-of-views are different approaches to describe the same object.

When $\mathcal{Y} = \mathbb{R}$ we will simply write $\mathscr{C}_0(\mathcal{X})$. Since we dispose of a suitable test functions space, we need to precise the notion of duality at stake in this review.

Definition 2.1.2 (Topological dual space)**.** *If E is a topological vector space, we denote E^* its* topological dual i.e. *the space of all continuous linear forms $\psi : E \to \mathbb{R}$. The pairing between an element $\phi \in E$ and a map $\psi \in E^*$ is denoted by the bilinear mapping $\langle \phi, \psi \rangle_{E \times E^*} \overset{\text{def.}}{=} \psi(\phi)$ called the* duality bracket*.*

This notion allows us to define the Radon measure through duality in the following definition.

Definition 2.1.3 (Set of Radon measures)**.** *We denote $\mathcal{M}(\mathcal{X})$ the set of real signed Radon measures on $\mathcal{X}$ of finite masses. It is the topological dual of $\mathscr{C}_0(\mathcal{X})$ with* supremum *norm $\|\cdot\|_{\infty,\mathcal{X}}$ by the Riesz-Markov representation theorem[3] [15] Thus, a Radon measure m is a continuous linear form evaluated on functions $f \in \mathscr{C}_0(\mathcal{X})$, with for $m \in \mathcal{M}(\mathcal{X})$ the duality bracket denoted by $\langle f, m \rangle_{\mathscr{C}_0(\mathcal{X}) \times \mathcal{M}(\mathcal{X})} = \int_{\mathcal{X}} f \, \mathrm{d}m$.*

The term 'signed' refers to the generalisation of the concept of (positive) measure, by allowing the quantity $\langle f, m \rangle_{\mathscr{C}_0(\mathcal{X}) \times \mathcal{M}(\mathcal{X})}$ to be negative. We can define in the same way the space of real non-negative Radon measures $\mathcal{M}^+(\mathcal{X})$ dual of $\mathscr{C}_0(\mathcal{X}, \mathbb{R}^+)$ and the space of complex Radon measures $\mathcal{M}_{\mathbb{C}}(\mathcal{X})$ dual of $\mathscr{C}_0(\mathcal{X}, \mathbb{C})$. Classic examples of Radon measures are:

- the Lebesgue measure of dimension $d \in \mathbb{N}$;
- the Dirac measure δ_z centred in $z \in \mathcal{X}$, also called the δ-peak. For all $f \in \mathscr{C}_0(\mathcal{X})$ one have $\langle f, \delta_z \rangle_{\mathscr{C}_0(\mathcal{X}) \times \mathcal{M}(\mathcal{X})} = f(z)$;
- discrete measures $m_{a,x} \overset{\text{def.}}{=} \sum_{i=1}^N a_i \delta_{x_i}$ where $N \in \mathbb{N}, \mathbf{a} \in \mathbb{C}^N, \mathbf{x} \in \mathcal{X}^N$.

Since $\mathscr{C}_0(\mathcal{X})$ is a Banach space, $\mathcal{M}(\mathcal{X})$ is complete [2] by endowing it with its dual norm called the total variation (TV) norm, defined for $m \in \mathcal{M}(\mathcal{X})$ by:

$$|m|(\mathcal{X}) \overset{\text{def.}}{=} \sup\left(\int_{\mathcal{X}} f \, \mathrm{d}m, f \in \mathscr{C}_0(\mathcal{X}), \|f\|_{\infty,\mathcal{X}} \leq 1 \right).$$

The TV norm of a measure is also called its *mass*. One can note that in the case of a *discrete measure* defined as before $m_{a,x} \overset{\text{def.}}{=} \sum_{i=1}^N a_i \delta_{x_i}$, one has $|m_{a,x}|(\mathcal{X}) = \|\mathbf{a}\|_1$.

The interested reader might take a look at the appendix B.1 for more details on some functional analysis notions and results.

> **Summary:** we model a spike by a Dirac measure, an element of the Radon measure spaces $\mathcal{M}(\mathcal{X})$. This space is defined by duality, it is endowed by the TV-norm and is complete. It is however infinite dimensional and non-reflexive (see B.1), this poses additional difficulties to be taken into account in the optimisation.

2.2. Observations

Let us introduce the space where the acquired data live. We will denote by $\mathcal{H}$ this Hilbert space; for the instance of images $\mathcal{H} = \mathrm{L}^2(\mathcal{X})$. Let $m \in \mathcal{M}(\mathcal{X})$ be the source measure, we call *acquisition* $y \in \mathcal{H}$ the result of the *forward/acquisition map* $\Phi : \mathcal{M}(\mathcal{X}) \to \mathcal{H}$ evaluated on m, with measurement kernel $\varphi : \mathcal{X} \to \mathcal{H}$:

$$y \overset{\text{def.}}{=} \Phi m = \int_{\mathcal{X}} \varphi(x) \, \mathrm{d}m(x). \tag{1}$$

[3] it can also be defined as the topological dual of the space of continuous function $\mathscr{C}(\mathcal{X})$ if $\mathcal{X}$ is compact.

The latter integral ought to be not confused with the duality bracket $\langle f, m \rangle_{\mathscr{C}_0(\mathcal{X}) \times \mathcal{M}(\mathcal{X})} = \int_{\mathcal{X}} f(x) \, \mathrm{d}m(x)$ mentioned in the Definition 2.1.3 above. Indeed, while $f(x) \in \mathbb{R}$ for $x \in \mathcal{X}$, we have $\varphi(x) \in \mathcal{H}$: the integral in (1) is then a *Böchner integral* [16] *i.e.* the proper notion to deal with vector valued map. It is valid as long as φ is continuous and bounded [7,13].

Remark. *Measures are objects that generalise functions at the cost of losing some of their properties. Thus, one cannot define a product of measures (what would be the square of the Dirac?) and one ought to be aware of some caveats concerning the functions of measure: these functionals need to be at most (sub)linear in order to be well-defined [17].*

In the following, we will impose $\varphi \in \mathscr{C}^2(\mathcal{X}, \mathcal{H})$. Let us also define the adjoint operator of $\Phi : \mathcal{M}(\mathcal{X}) \to \mathcal{H}$ in the weak-$*$ topology, namely the map $\Phi^* : \mathcal{H} \to \mathscr{C}_0(\mathcal{X})$. It is defined for all $x \in \mathcal{X}$ and $p \in \mathcal{H}$ by $\Phi^*(p)(x) = \langle p, \varphi(x) \rangle_{\mathcal{H}}$. The choice of φ and $\mathcal{H}$ depends on the physical process of acquisition, indeed generic measurement kernels are:

- convolution kernel with typically $\mathcal{H} = \mathrm{L}^2(\mathcal{X})$ and $\forall x \in \mathcal{X}, \varphi(x) \overset{\text{def.}}{=} (s \mapsto \tilde{\varphi}(s - x)) \in \mathcal{H}$, for the PSF $\tilde{\varphi} \in \mathscr{C}_0^2(\mathbb{R}^d)$. One has for instance the Gaussian kernel, centred in $c \in \mathcal{X}$ with spread $\sigma > 0$, defined by $s \mapsto \tilde{\varphi}(s - c) \overset{\text{def.}}{=} 1/\sqrt[d/2]{2\pi\sigma^2}\, e^{-\|s-c\|_2^2/2\sigma^2}$;
- Fourier kernel with cut-off frequency $f_c \in \mathbb{N}$ and $\mathcal{H} = \mathbb{C}^{2f_c+1}$, for $x \in \mathcal{X} = \mathbb{T}$ in 1D:

$$\varphi(x) = \left(e^{2\mathrm{i}\pi k x}\right)_{|k| \leq f_c};$$

- Laplace kernel [8] for non-negative weighting function $\xi \in \mathscr{C}(\mathcal{X})$ specific to the physical acquisition process and $\mathcal{H} = \mathrm{L}^2(\mathbb{R}^+)$: $\forall x \in \mathcal{X}, \quad \varphi(x) \overset{\text{def.}}{=} (s \mapsto \xi(x) e^{-sx}) \in \mathcal{H}$.

These 3 kernels correspond to various physical context of imagery, hence they are encountered in multiple acquisition process, such as Nuclear Magnetic Resonance spectroscopy (Fourier), SMLM super-resolution (convolution), MA-TIRF (Laplace), *etc.*

We will now on use the following notation for the discrete forward map: let $x = (x_1, \ldots, x_N)$ and $a \in \mathbb{R}^N$: $\Phi_x(a) \overset{\text{def.}}{=} \sum_{i=1}^N a_i \varphi(x_i)$.

Shorthand: an acquisition living in the Hilbert space $\mathcal{H}$ of a measure m is the quantity Φm. Φ is the forward operator, completely defined by a kernel φ specific to the physical context of imagery.

2.3. An off-the-grid functional: the BLASSO

Let $m_{a_0,x_0} \overset{\text{def.}}{=} \sum_{i=1}^N a_{0,i} \delta_{x_{0,i}}$ be the source measure with amplitudes $\mathbf{a_0} \in \mathbb{R}^N$ and positions $\mathbf{x_0} \in \mathcal{X}^N$, the sparse spike problem is to recover this measure from the acquisition $y \overset{\text{def.}}{=} \Phi m_{a_0,x_0} + w$ where $w \in \mathcal{H}$ is an additive noise, typically white Gaussian noise. To tackle this problem, we use the following convex functional [2,18] also called the BLASSO, which stands for Beurling-LASSO:

$$\operatorname*{argmin}_{m \in \mathcal{M}(\mathcal{X})} T_\lambda(m) \overset{\text{def.}}{=} \frac{1}{2}\|y - \Phi(m)\|_{\mathcal{H}}^2 + \lambda |m|(\mathcal{X}) \qquad (\mathcal{P}_\lambda(y))$$

with regularisation parameter $\lambda > 0$ which accounts for the trade-off between fidelity and sparsity of the reconstruction. The name BLASSO was coined in the work of [18, 19] according to the link between the *Generalised Minimal Extrapolation* (GME) problem where one seeks to reconstruct a Radon measure from several observations on its Fourier

coefficients, and the work [20] of the Norwegian mathematician Beurling[4] which coincides with GME in the case of a Fourier forward operator.

The BLASSO in a noiseless setting writes down:

$$\underset{\Phi m = y_0}{\operatorname{argmin}} |m|(\mathcal{X}) \quad \text{with } y_0 = \Phi m_{a_0,x_0}. \tag{$\mathcal{P}_0(y_0)$}$$

BLASSO is genuinely linked with its discrete counterpart the (LASSO) [4]: one can formally see BLASSO as the functional limit of LASSO on a finer and finer grid. If the LASSO problem exhibits existence and uniqueness of the solution, what can one says for its off-the-grid counterpart? First of all, let us observe that:

- $m \mapsto |m|(\mathcal{X})$ is lower semi-continuous w.r.t. the weak-$*$ convergence (see Appendix B.1 for more insights);
- Φ is continuous from the weak-$*$ topology of $\mathcal{M}(\mathcal{X})$ to the weak topology of $\mathcal{H}$.

Thus, one can establish the existence of solutions to ($\mathcal{P}_\lambda(y)$) thanks to convex analysis results, as proved in [2].

Summary: the sparse spike problem is tractable thanks to the convex functional on $\mathcal{M}(\mathcal{X})$ called the BLASSO and denoted by ($\mathcal{P}_\lambda(y)$). With $m \in \mathcal{M}(\mathcal{X})$ as an input, it consists in a data term comparing observed data versus Φm, and a regularisation accounting for sparsity prior through the TV-norm of m. Existence of solutions of the BLASSO is known and proved.

The difficulties now lie in the following questions:

1. what are the conditions to recover a sparse measure, within a certain noise regime? Is the *minimum* unique?
2. under which conditions can we retrieve exactly the number of spikes, the amplitude, and the positions; when do we have support stability?
3. how can we tackle numerically the infinite dimensional and non-reflexive nature of the space $\mathcal{M}(\mathcal{X})$?

In order to address these points, we need to introduce some notions of convex analysis in the following subsection.

2.4. Dual problems and certificates

The BLASSO in the equation ($\mathcal{P}_\lambda(y)$) above is a minimisation problem with a convex functional. Then we can apply Ekeland-Temam [21, Remark 4.2] results[5] and define a dual problem which writes down for $p \in \mathcal{H}$ (see Appendix B.2 for the proof):

$$\underset{\|\phi^* p\|_{\infty,\mathcal{X}} \leq 1}{\operatorname{argmax}} \ \langle y, p \rangle_{\mathcal{H}} - \frac{\lambda}{2} \|p\|^2_{\mathcal{H}} \tag{$\mathcal{D}_\lambda(y)$}$$

which can be recast as the projection onto a closed convex [2,19] :

$$\underset{\|\phi^* p\|_{\infty,\mathcal{X}} \leq 1}{\operatorname{argmax}} \ \left\| \frac{y}{\lambda} - p \right\|^2_{\mathcal{H}} \tag{$\mathcal{D}'_\lambda(y)$}$$

[4] More precisely he studied the minimal total variation norm function among all bounded variation functions with constrained Fourier transform on a given domain.

[5] A little caveat should be raised for these results: the space of features V should be a reflexive space, which is clearly not the case here with $V = \mathcal{M}(\mathcal{X})$. However, as stated in the Appendix B.2, the reflexive hypothesis is only needed for the sake of existence proof. Since we already proved the solution's existence, this hypothesis is not relevant let alone necessary in this context.

Fenchel's duality between ($\mathcal{P}_\lambda(y)$) and ($\mathcal{D}_\lambda(y)$) is proved in [2]. Therefore, any solution m_λ of ($\mathcal{P}_\lambda(y)$) is linked [4] to the unique solution p_λ of ($\mathcal{D}_\lambda(y)$) by the extremality conditions:

$$\begin{cases} \Phi^* p_\lambda \in \partial|m_\lambda|(\mathcal{X}), \\ -p_\lambda = \dfrac{1}{\lambda}(\Phi m_\lambda - y) \end{cases} \tag{2}$$

where $\partial|\cdot|(\mathcal{X})$ is the sub-differential of the TV norm. Indeed, since the total variation is not differentiable (as the ℓ^1 norm) but lower semi-continuous w.r.t. the weak-$*$topology, we use its sub-differential which for $m \in \mathcal{M}(\mathcal{X})$ identifies to:

$$\partial|m|(\mathcal{X}) = \left\{ \eta \in \mathscr{C}_0(\mathcal{X}); \|\eta\|_{\infty,\mathcal{X}} \leq 1 \text{ and } \int_{\mathcal{X}} \eta \, \mathrm{d}m = |m|(\mathcal{X}) \right\}. \tag{3}$$

Elements of this subgradient are called *certificate*. Thanks to strong duality, one can define peculiar certificates called the *dual certificates* [1].

Definition 2.4.1. *We call* $\eta_\lambda \overset{\text{def.}}{=} \Phi^* p_\lambda$ *where* p_λ *satisfies* (2)*, a* dual certificate *of* m_λ*.*

It is a certificate since $\Phi^* p_\lambda \in \partial|m_\lambda|(\mathcal{X})$ and it is called *dual* because it verifies the second extremality (2) condition: it is thus defined by the dual solution p_λ. Loosely speaking a dual certificate η_λ is associated to a measure m_λ and it *certifies* that the measure m_λ is a *minimum* of the BLASSO. For instance, if there exists solutions of ($\mathcal{P}_\lambda(y)$) of the form $m_\lambda \overset{\text{def.}}{=} \sum_{i=1}^N a_i \delta_{x_i}$, the support satisfies [4] for all $0 \leq i \leq N : |\eta_\lambda|(x_i) = 1$.

In the same fashion, one has the link between a solution m_0 of the noiseless BLASSO ($\mathcal{P}_0(y_0)$) and its certificates η_0, which are not unique in general. Then, in the rest of the document we will refer to η_0 as the minimal norm certificate *i.e.* the dual certificate η_0 with minimal *supremum* norm $\|\eta_0\|_{\infty,\mathcal{X}}$. It is shown in [4] that this minimal norm certificate η_0 has important properties, since it somehow drives the stability of the recovered spike locations when the additive noise is small, in particular how close they are to the positions of the true measure m_{a_0,x_0}: see definition 2.5.2 in the section below.

Summary: we defined the primal problem in the former section, thanks to convexity we can define the dual problem of the BLASSO. A solution m_λ of the BLASSO and a solution p_λ of the dual problem are linked through extremality condition. The dual solution p_λ defines the dual certificate, an element of the subgradient specified by $\eta_\lambda = \Phi^* p_\lambda$: the dual certificate η_λ *certifies* that m_λ is a solution of the BLASSO. We can then establish more precise conditions on the uniqueness/support recovery.

2.5. Support recovery guarantees

We will address in this section the first two questions we have laid down, namely existence, uniqueness and support recovery conditions. A classical tool to establish some recovery properties lies in the notion of the *minimum* separation distance.

Definition 2.5.1 (Minimum separation distance)**.** *The* minimum *separation distance is a characterisation of the support of the discrete measure* m_{a_0,x_0} *by:*

$$\Delta(m_{a_0,x_0}) \overset{\text{def.}}{=} \min_{i \neq j} |x_{0,i} - x_{0,j}|.$$

The reconstruction condition is driven by this minimum separation distance, itself determined by the type of measure (complex, real, real non-negative) and the type of forward operator.

- if the operator is an acquisition of the Fourier spectrum within $[-f_c, f_c]$ with frequency cut-off f_c for $\mathcal{X} = \mathbb{T}^d$ the d-torus in the noiseless setting, it is necessary that $\Delta(m_{a_0,x_0}) \gtrsim \frac{2}{f_c}$ if the source measure is complex [1]. Upon a few conditions [22] one can weaken it to $\Delta(m_{a_0,x_0}) \gtrsim \frac{1.26}{f_c}$, and $\Delta(m_{a_0,x_0}) \gtrsim \frac{1.87}{f_c}$ if the source measure is real [1];
- regardless of the operator Φ [18,23], there is no condition on the separation for a real **positive** source measure in the noiseless setting, however stability constant explodes when $\Delta(m_{a_0,x_0}) \to 0$.

These results are important but do not provide a sharp characterisation of the recovery in the presence of noise; however, we expect to find noise in the images we deal with and therefore to be limited by this noise regime. To account for this effect we need to add some conditions on the ground-truth measure, following the work of [4] we introduce:

Definition 2.5.2 (Non-degenerate source condition)**.** *The source m_{a_0,x_0} verifies the NDSC* (Non-Degenerate Source Condition) *if:*

- *there exists $\eta \in \operatorname{Im} \Phi^*$ such that $\eta \in \partial |m_{a_0,x_0}|(\mathcal{X})$;*
- *$\forall s \in \mathcal{X} \setminus \cup_{i=1}^N \{x_{0,i}\}, |\eta_0(s)| < 1$;*
- *$\forall i \in [\![1, N]\!]$, the Hessian matrix $\boldsymbol{\nabla}^2 \eta_0(x_{0,i}) \in \mathbb{R}^{d \times d}$ is invertible.*

The first condition amounts to assuming that m_{a_0,x_0} is a solution to ($\mathcal{P}_0(y_0)$) and there exists a solution to its dual problem. If the two latter conditions are matched, we say that η_0 is *not degenerate*. This allows us to write the main result of [4] namely:

Theorem 2.5.3 (Noise robustness [4])**.** *Let Γ_{x_0} the $N \times N$ matrix defined by $\Gamma_{x_0} \overset{\text{def.}}{=} (\varphi(\cdot - x_{0,i}), \varphi'(\cdot - x_{0,i}))_{i=1}^N$. Assume that Γ_{x_0} has full column rank and that m_{a_0,x_0} verifies the NDSC. Then there exists $\alpha > 0, \lambda_0 > 0$ such that for all $0 \leq \lambda \leq \lambda_0$ and w such that $\|w\| \leq \alpha\lambda$; there exists N pairings $(a_{\lambda,i}, x_{\lambda,i})$ such that $m_\lambda \overset{\text{def.}}{=} \sum_{i=1}^N a_{\lambda,i} \delta_{x_{\lambda,i}}$ is the unique solution of ($\mathcal{P}_\lambda(y)$) composed of exactly N spikes. In particular, for $\lambda = 1/\alpha \|w\|_{\mathcal{H}}$ we have the control over the discrepancies:*

$$\forall i \in [\![1, N]\!] : \|x_{\lambda,i} - x_{0,i}\| = \mathcal{O}(\|w\|_{\mathcal{H}}) \text{ and } |a_{\lambda,i} - a_{0,i}| = \mathcal{O}(\|w\|_{\mathcal{H}}).$$

Under the Non-Degenerate Source Condition, for λ and $\|w\|_{\mathcal{H}}^2/\lambda$ small enough, one can reconstruct a measure with the same number of spikes as the ground-truth measure m_{a_0,x_0}. Furthermore, the reconstructed measure (weak-$*$)converges to the ground-truth measure when the noise level drops to 0. The authors of [4] also introduce the notion of vanishing derivatives precertificate. The η_0 certificate is indeed hard to compute from the dual problem of ($\mathcal{P}_0(y_0)$) because of the constraint $\|\eta_0\|_{\infty,\mathcal{X}} \leq 1$, the precertificate allows to leverage this computation by solving instead a linear system. The interested reader is advised to take a glance at this article among other ones [4,23] for these new concepts.

> **Shorthand:** the minimum separation distance criterion is used to assess recovery possiblities in the noiseless setting. In a low regime of noise, a theorem states that the source measure $m_{a,x}$ composed of N spikes can be recovered through BLASSO, with a control over the discrepancies (amplitudes/positions) between the reconstructed and the source measures.

We were therefore able to establish some guarantees on the reconstruction of the source measures in the presence of noise. In the next section, we propose to address the third question and to discuss strategies to compute the numerical solution of the inverse problem; a difficult task requiring to account for the difficulties of the optimisation space.

3. Numerical strategies to tackle the BLASSO

The BLASSO problem $\mathcal{P}_\lambda(y)$ is an optimisation over the set of Radon measures, an infinite dimensional and non-reflexive space. We recall that it writes down:

$$\underset{m\in\mathcal{M}(\mathcal{X})}{\operatorname{argmin}}\ T_\lambda(m) \overset{\text{def.}}{=} \frac{1}{2}\|y-\Phi(m)\|_{\mathcal{H}}^2+\lambda|m|(\mathcal{X}). \tag{$\mathcal{P}_\lambda(y)$}$$

A naive approach would be to enforce the measure m to be supported on a fine grid $(p_i)_i^L$ which is equivalent to solve the LASSO problem:

$$\min_{a\in\mathbb{R}^L}\|y-\Phi_L a\|_{\mathcal{H}}^2+\|a\|_1$$

with the discrete operator $\Phi_L a \overset{\text{def.}}{=} \sum_{i=1}^L a_i\varphi(p_i)$ and φ the kernel of the forward operator. This approach conveys numerous cons: for instance the solution of the LASSO, in small noise regime and when the step size tends to 0, contains pairs of spikes around the true one [10,11]. Furthermore, refining the step size leads to a worse conditioning of the forward operator, accounting for numerical difficulties. The following classes of algorithms better account for the infinite dimensional nature of $\mathcal{M}(\mathcal{X})$. We present in details the three methods with the most established results in the literature [13,19,24]. Before describing these methods, let us remark that there exist also some promising avenues, such as the projected gradient descent [25,26]. It relies on an over parametrised initialisation *i.e.* a discrete measure with numerous δ-peaks compared to the ground-truth, then one applies a gradient descent on the amplitudes and positions of the over parametrised measure combined at each step with a projection on a set of positions constraints to enforce the separation of the spikes. This projection can be replaced by a 'heuristic' which boils down to the merging of δ-peaks that are not enough separated [26].

3.1. Semi-definite recasting and hierarchy

Semi-definite programming was one of the first scheme solving the BLASSO in the specific case of a Fourier acquisition on the 1D torus $\mathbb{T}^1$ [1,3,18,19]. Before explaining in layman terms the SDP scheme, let us first introduce and detail the relevant quantities for this section. Let $d=1$ be the dimension of the interior of $\mathcal{X}$, let us study the case where the forward operator denoted by $\mathcal{F}_n$ (and not Φ for this section) is a Fourier coefficients measurements up to some cut-off frequency $f_c\in\mathbb{N}$, with $n=2f_c+1$ the number of measurements. We have $\mathcal{F}_n:\mathcal{M}_{\mathbb{C}}(\mathcal{X})\to\mathbb{C}^n$ and for a discrete measure $m_{a,x}\overset{\text{def.}}{=}\sum_{j=1}^N a_j\delta_{x_j}$ it writes down $\mathcal{F}_n(m_{a,x})=\left(\sum_j a_j e^{2\mathrm{i}\pi k x_j}\right)_{|k|\leq f_c}$ and its adjoint operator $\mathcal{F}_n^*:\mathbb{C}^n\to\mathscr{C}_0(\mathcal{X},\mathbb{C})$ is for $s\in\mathcal{X}$:

$$\forall c\in\mathbb{C}^n,\quad \mathcal{F}_n^*(c)(s)=\left\langle c,\left(e^{2\mathrm{i}\pi ks}\right)_{|k|\leq f_c}\right\rangle_{\mathbb{C}^n}=\sum_{|k|\leq f_c} c_{f_c+k}\,e^{2\mathrm{i}\pi ks}. \tag{4}$$

This method is based on semi-definite programming (SDP) for efficiently computing the *minima* of BLASSO. It stems from the Hilbert approach [27] when one globally decomposes the objective function into simple pieces, atoms. The solution of the dual problem of ($\mathcal{P}_\lambda(y)$), denoted here ($\mathcal{D}_\lambda^{\mathcal{F}}(y)$), is a polynomial p linked to a certificate by $\mathcal{F}_n^*p$: the idea then is the reconstruction of the dual certificate as a linear sum of trigonometric polynomials [19], which is enough to find the measure associated with this reconstructed certificate. This associated measure is a solution to the BLASSO. The dual problem, on the other hand, is tractable thanks to a semi-definite programming approach.

Since $\mathcal{F}_n^* p$ is a trigonometric polynomial for any $p \in \mathbb{C}^n$ by the definition above, one can recast the constraint $\|\mathcal{F}_n^* p\|_{\infty,\mathcal{X}} \leq 1$ (imposed by definition of a certificate, see equation (3)) and rewrite it as the intersection of the cone of positive semi-definite matrices $\{A : A \succeq 0\}$ with an affine hyperplane [1,28]. Hence, the Fenchel dual problem of $(\mathcal{P}_\lambda(y))$ for the Fourier forward operator $\mathcal{F}_n$:

$$\max_{p\in\mathbb{C}^n} \operatorname{Re}\{y,p\} - \frac{\lambda}{2}\|p\|_{\mathcal{H}}^2 \qquad \text{constrained by } \|\mathcal{F}_n^* p\|_{\infty,\mathcal{X}} \leq 1 \tag{$\mathcal{D}_\lambda^{\mathcal{F}}(y)$}$$

with hermitian product $\{\cdot,\cdot\}$, has the equivalent formulation [28]

$$\max_{p\in\mathbb{C}^n, Q\in\mathbb{C}^{n\times n}} \operatorname{Re}\{y,p\} - \frac{\lambda}{2}\|p\|_{\mathcal{H}}^2 \qquad \text{constrained by}$$
$$\begin{pmatrix} Q & p \\ p^* & 1 \end{pmatrix} \succeq 0 \text{ and } \sum_{k=1}^{n-j} Q_{k,k+j} = \delta_{0,j} \text{ for } j \in [\![1, n-1]\!] \tag{$\widetilde{\mathcal{D}}_\lambda^{\mathcal{F}}(y)$}$$

with Q a Hermitian matrix and p a vector of coefficients (accounting for the dual variable p), and $\delta_{0,j}$ the Kronecker delta equal to 1 if $j = 0$ and 0 otherwise. The choice of regulariser λ is crucial: if chosen too high it will yield a solution with fewer spikes, if chosen too low it will recover a solution with spurious spikes. This finite dimensional formulation can now be tackled with classic semi-definite programming solvers, as did the authors of [1] who proposed an algorithm of *Interior Point Method*, given in the Algorithm (1). The first step reaches a solution p, allowing the definition of the certificate $p_{2n-2}(e^{2i\pi t}) \overset{\text{def.}}{=} 1 - |\mathcal{F}_n^* p|^2(t)$, where $\mathcal{F}_n^*$ is defined in equation (4).

Algorithm 1: Interior Point Method applied to the BLASSO.

1 Solve

$$\max_{p\in\mathbb{C}^n, Q\in\mathbb{C}^{n\times n}} \operatorname{Re}\{y,p\} - \frac{\lambda}{2}\|p\|_{\mathcal{H}}^2$$

subject to $\begin{pmatrix} Q & p \\ p^* & 1 \end{pmatrix} \succeq 0$ and $\sum_{i=1}^{n-j} Q_{i,i+j} = \delta_{0,j}$ for $j = 1, \ldots, n-1$.

2 Reconstruct the support $\hat{X}$ of m by locating the roots of p_{2n-2} on the unit circle (*e.g.* by computing the eigenvalues of its companion matrix).

3 Solve $\sum_{t\in\hat{X}} a_t e^{-2i\pi kt} = y_k$ to recover the amplitudes a.

One can note the link between the dual and the primal problem, *i.e.* that p the solution of $(\widetilde{\mathcal{D}}_\lambda^{\mathcal{F}}(y))$ entails the location of the spikes: as $\mathcal{F}_n^* p$ yields its extremal points on the support of m since it is the certificate of a discrete measure, note that $p_{2n-2}(e^{2i\pi t}) = 1 - |\mathcal{F}_n^* p|^2(t)$ has all its roots on the unit circle and these roots are the support of the target measure [1]. Thus, the strategy is to solve the dual problem and then to use a root-finding algorithm on the certificate $\mathcal{F}_n^* p$ associated to the dual solution, hence reconstructing the support of the measure then the measure (after a last amplitude recover step). We present an example of the reconstruction of 3 Dirac measures on the $1D$ torus $\mathbb{T}^1$ through the observed noisy data y and the roots of the polynomial $p_{2n-2}(e^{2i\pi t})$ in Figure 2.

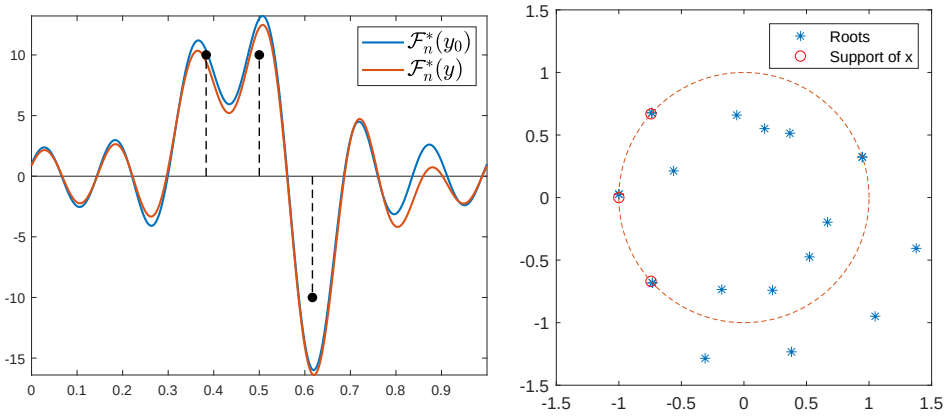

Figure 2. (**a**) Certificates associated to acquisition y and noiseless y_0, result of 3 δ-peaks (in black, plotted with 10 times their ground-truth amplitudes) through a Fourier measurement of cut-off frequency $f_c = 6$. (**b**) Localisation of the roots of the certificate associated to the dual *maximum*. All the roots (the 3 ground-truth and the spurious spike on the right) on the unit circle are interpreted as the support of the δ-peaks.

This strategy is only suitable for $d = 1$. For the multi-variate case, one needs to make use of a so-called Lasserre Hierarchy [29]. Consider the semi-definite relaxation of order m with $m > n = 2f_c + 1$:

$$\max_{p \in \mathbb{C}^n, Q \in \mathbb{C}^{n \times n}} \mathrm{Re}\{y, p\} \quad \text{constrained by}$$
$$\begin{cases} 0 \preceq \begin{pmatrix} Q & \tilde{p} \\ \tilde{p}^* & 1 \end{pmatrix} \text{ where } \tilde{p}_k = \begin{cases} c_k \text{ if } k \in [-f_c, f_c]^d \\ 0 \text{ otherwise} \end{cases} \\ \mathrm{Tr}(\Theta_k Q) = \delta_{0,k} \text{ with } k \in [\![-m, m]\!] \end{cases}$$

with $\Theta_k = \theta_{k_d} \otimes \cdots \otimes \theta_{k_1}$ where θ_{k_j} the entries of $m \times m$ elementary Toeplitz matrix are 1 on its k_j-th diagonal and 0 elsewhere, and $\otimes$ the Kronecker product. In a nutshell, Lasserre's hierarchies give a sequence of nested outer SDP approximations of the cone of moments of non-negative measure. This method has been successfully applied to super-resolution in [3]. Some reconstructions in the 1D setting with a Fourier kernel are given in the Figure 3, the interested reader may find a more in-depth tutorial in the Numerical Tours[6] on 'Sparse spikes measures' joined with the code used to compute the following figure.

[6] https://nbviewer.jupyter.org/github/gpeyre/numerical-tours/blob/master/matlab/sparsity_8_sparsespikes_measures.ipynb

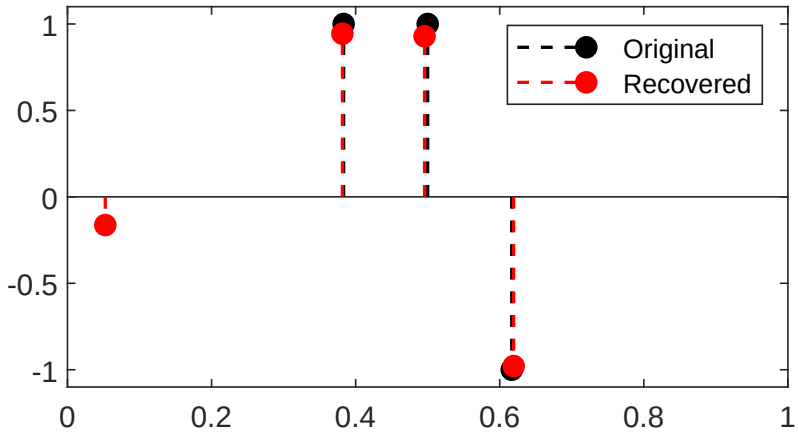

Figure 3. Reconstruction with the Interior Point Method for $\lambda = 1$. The algorithm detected a spurious spike near 0.05, otherwise amplitudes and positions of the peaks are correctly estimated.

These methods are proved to be asymptotically exact [3]. Nonetheless, it is not known if the algorithm has finite convergence in general: one does not know when to stop the hierarchy to obtain a solution of the BLASSO [7]. This stems from the fact that non-negative trigonometric polynomials in dimension $d > 1$ are not necessarily sums of square. Moreover, these SDP based approach are rather limited to a certain class of measurement map Φ, typically the Fourier forward operator or at least filters with compact Fourier supports. With the two following class of algorithm, one can better exploit the continuous setting and get rid of the discretisation drawback.

Summary (1st algorithm): the scheme boils down to the resolution of the dual problem, the reconstruction of the measure's support thanks to the certificate associated to the dual solution, and finally the solving of a linear problem to yield the corresponding estimated amplitudes. This strategy can be extended to a multivariate context but still, it is quite restrictive on the forward operator and it does not have finite convergence in general.

3.2. Greedy algorithm: the conditional gradient

The conditional gradient method also called the Frank-Wolfe (FW) algorithm [30,31] aims at solving $\min_{m \in C} f(m)$ for C a weakly compact convex set of a topological vector space and f a convex and differentiable function (the differential is then denoted by $\mathrm{d}f$). It relies on the iterative minimisation of a linearised version of f. Hence, the interest of this algorithm lies in the fact that it uses only the directional derivatives of f and that it does not require any Hilbertian structure, contrary to a classic proximal algorithm formulated in terms of Euclidean distance. We recall the definition of the conditional gradient in the pseudocode 2 for the general problem of minimising f.

Algorithm 2: Frank-Wolfe.

for $0 \leq k \leq K$ **do**
  $s^k \in \operatorname{argmin}_{s \in C} f(m^k) + \mathrm{d}f(m^k)(s - m^k)$.
  if $\mathrm{d}f(m^k)(s^k - m^k) = 0$ **then**
    m^k is a solution. Stop.
  else
    Step research: $\gamma^k \leftarrow$ either $\dfrac{2}{k+2}$ or $\operatorname{argmin}_{\gamma \in [0,1]} f(m^k + \gamma(s^k - m^k))$.
    Update: $m^{k+1} \leftarrow m^k + \gamma^k(s^k - m^k)$.
  end
end

One can make the following remarks:

- the compactness assumption on C ensures that the argmin in step 2 is non-empty;
- in line 7, we can replace m^{k+1} by any element $\hat{m} \in C$ such that $f(\hat{m}) \leq f(m^{k+1})$ without changing the convergence properties of the algorithm;

There are however two problems that prevent us from applying straightforwardly this algorithm to BLASSO: T_λ is not differentiable, and the optimisation set $\mathcal{M}(\mathcal{X})$ is not bounded. It is thus necessary to perform an *epigraphical lift* [13,32] to reach a differentiable functional that shares the same *minimum* measures as T_λ:

$$\min_{(r,m)\in C} \tilde{T}_\lambda(m,r) \overset{\text{def.}}{=} \frac{1}{2}\|y - \Phi(m)\|_{\mathcal{H}}^2 + \lambda r \qquad (\widetilde{\mathcal{P}}_\lambda(y))$$

with the bounded set $C = \{(r,m) \in \mathbb{R}^+ \times \mathcal{M}(\mathcal{X}); |m|(\mathcal{X}) \leq r \leq M\}$ and $M \overset{\text{def.}}{=} \frac{\|y\|^2}{2\lambda}$. Even though C is not weakly compact, it is compact for the weak-$*$ topology and the hypotheses for the algorithm 2 are still matched. The Frank-Wolfe algorithm is then well-defined for the energy $\tilde{T}_\lambda$, differentiable in the Fréchet sense on the Banach $\mathbb{R} \times \mathcal{M}(\mathcal{X})$. Its differential writes down :

$$\mathrm{d}\tilde{T}_\lambda : (r', m') \mapsto \int_X \Phi^*(\Phi m - y)\,\mathrm{d}m' + \lambda r'.$$

Finally, one has that m^* is a *minimum* of T_λ iff $(|m^*|(\mathcal{X}), m^*)$ minimises $(\widetilde{\mathcal{P}}_\lambda(y))$, and $T_\lambda(m^*) = \tilde{T}_\lambda(|m^*|(\mathcal{X}), m^*)$. In the rest of the document, we will omit the r-part, and we will refer to the quantity $(|m^*|(\mathcal{X}), m^*)$ by only m^*.

We note before that the update m^{k+1} in line 7 can be replaced by any value $\hat{m}$ improving the objective function, this remark is rather interesting as it can drastically improve the convergence property of the algorithm [2,33]. Hence, an interesting improvement to the Frank-Wolfe algorithm relies in the change of the final update step by a non-convex optimisation on both the amplitudes and the positions of the reconstructed δ-peaks in a simultaneous fashion. This modification is presented in the algorithm 3.

This tweak yields a theoretical convergence to the unique solution of BLASSO in a finite number of iterations, empirically a N-step convergence. This version is called the *Sliding Frank-Wolfe* algorithm [13], as the spike positions are sliding on the continuous domain $\mathcal{X}$. The authors also proved in the same paper that the generated measure sequence $m^{[k]}$ converges towards the *minimum* for the weak-$*$ topology.

A reconstruction by *Sliding Frank-Wolfe* for the same Fourier operator, ground-truth spikes and acquisition as the latter section is plotted in Figure 4. On contrary to SDP in Figure 3, no spurious spike is reconstructed. As in the SDP method, the choice of regulariser λ is crucial: if chosen too high it will yield a solution with fewer spikes than needed, if set too low it will recover a solution with spurious spikes. We set $\lambda = 1$ for the 1D Fourier example as in the former SDP section.

The line 3 in the algorithm 3 is typically solved by a grid search, the convex step in line 5 can use a FISTA solver [34] and the non-convex step in line 6 can be tackled by a modified Broyden-Fletcher-Goldfarb-Shann method (L-BFGS-B) implementation [35]. Reconstructions in the 2D setting with a convolution kernel, similar to the SMLM conditions, are presented in the Figure 5. Since luminosity is always a non-negative quantity, one can restrict [8] the SFW to build a positive measure of the cone $\mathcal{M}^+(\mathcal{X})$, by changing:

- the stopping condition to $\eta^{[k]}\left(x_*^{[k]}\right) \leq 1$;
- the LASSO step is solved for $a \in \mathbb{R}_+^{N^{[k]+1}}$;
- the non-convex step is solved on $\mathbb{R}_+^{N^{[k]+1}} \times \mathcal{X}^{N^{[k]+1}}$.

Algorithm 3: *Sliding Frank-Wolfe.*

Input: Acquisition $y \in \mathcal{H}$, number of iterations K, $\lambda > 0$.

Initialisation : $m^{[0]} = 0$, $N^{[k]} = 0$.

for k, $0 \leq k \leq K$ **do**

For $m^{[k]} = \sum_{i=1}^{N^{[k]}} a_i^{[k]} \delta_{x_i^{[k]}}$ such that $a_i^{[k]} \in \mathbb{R}$, $x_i^{[k]} \in \mathcal{X}$, find $x_*^{[k]} \in \mathcal{X}$ such that :

$$x_*^{[k]} \in \operatorname*{argmax}_{x \in \mathcal{X}} \left| \eta^{[k]}(x) \right| \qquad \text{where} \quad \eta^{[k]}(x) \stackrel{\text{def.}}{=} \frac{1}{\lambda} \Phi^*(\Phi m^{[k]} - y),$$

if $\left| \eta^{[k]}\left(x_*^{[k]}\right) \right| \leq 1$ **then**

$m^{[k]}$ is the solution of the BLASSO. Stop.

else

Compute $m^{[k+1/2]} = \sum_{i=1}^{N^{[k]}} a_i^{[k+1/2]} \delta_{x_i^{[k+1/2]}} + a_{N^{[k]}+1}^{[k+1/2]} \delta_{x_*^{[k]}}^{[k+1/2]}$ such that:

$$a_i^{[k+1/2]} \in \operatorname*{argmin}_{a \in \mathbb{R}^{N^{[k]}+1}} \frac{1}{2} \left\| y - \Phi_{x^{[k+1/2]}}(a) \right\|_{\mathcal{H}}^2 + \lambda \|a\|_1$$
$$\text{for} \quad x^{[k+1/2]} \stackrel{\text{def.}}{=} \left(x_1^{[k]}, \ldots, x_{N^{[k]}}^{[k]}, x_*^{[k]} \right).$$

Compute $m^{[k+1]} = \sum_{i=1}^{N^{[k+1]}} a_i^{[k+1]} \delta_{x_i^{[k+1]}}$ such that:

$$\left(a_i^{[k+1]}, x_i^{[k+1]} \right) \in \operatorname*{argmax}_{(a,x) \in \mathbb{R}^{N^{[k]}+1} \times \mathcal{X}^{N^{[k]}+1}} \frac{1}{2} \left\| y - \Phi_{x^{[k+1/2]}}(a) \right\|_{\mathcal{H}}^2 + \lambda \|a\|_1.$$

end

end

Output: Discrete measure $m^{[k]}$ where k is the stopping iteration.

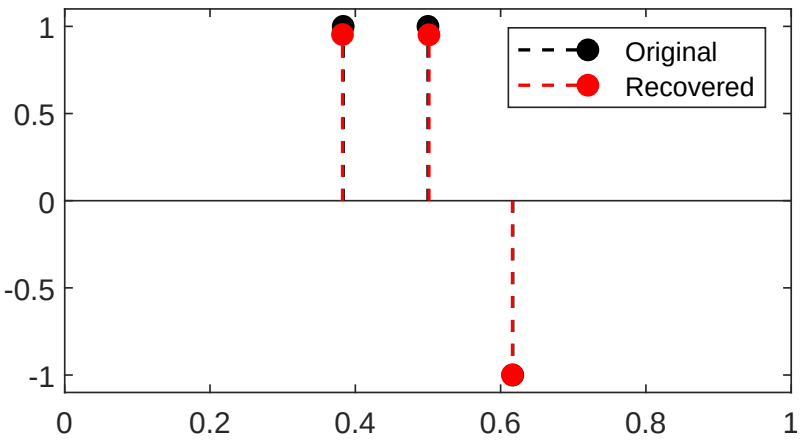

Figure 4. Reconstruction by *Sliding Frank-Wolfe* for a 1D Fourier operator, with the same settings (y, noise realisations, $\lambda = 1$) as the former section. All ground-truth spikes are reconstructed, no spurious spike is detected.

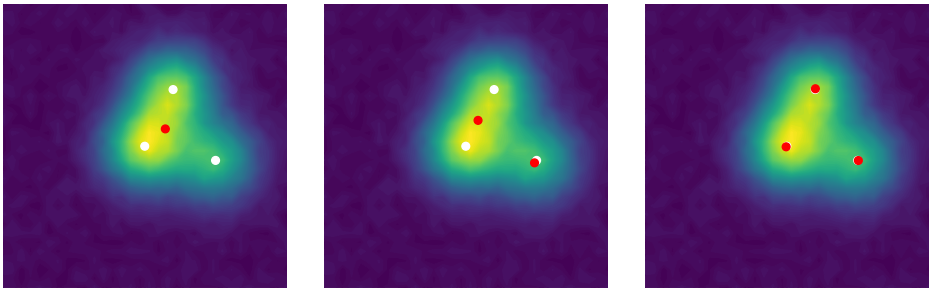

Figure 5. (**a**) First iterate $k = 0$ (**b**) Mid-computation $k = 1$ (**c**) End of the computation $k = 2$, results for SFW reconstruction on the domain $\mathcal{X} = [0, 1]^2$ for the Gaussian kernel with spread-factor $\sigma = 0.1$ and additive Gaussian noise of variance 0.1. All δ-peaks are successfully recovered only thanks to the acquisition, $\lambda = 3 \times 10^{-2}$.

Hence, this modified algorithm offers a good trade-off between precision and theoretical guarantees. However, it suffers from the high computation load for one iteration, making it slow to compute. The next section algorithm is a promising alternative with easier/cheaper iteration while still taking advantage of the continuous setting.

Shorthand (2nd algorithm): Conditional gradient method is a greedy algorithm consisting in the iterative minimisation of a linearised version of the objective convex function. This algorithm can be applied to any forward operator without restriction on the space $\mathcal{X}$. Up to a modification (SFW), the Frank-Wolfe algorithm reaches a finite convergence, empirically a N-step convergence for a source measure with N spikes. The iterations however are computationally costly, yielding long computation time.

3.3. Optimal transport based algorithm: the particle gradient descent

All the following results are proven for a domain $\mathcal{X}$ with no boundaries, *e.g.* the d-dimensional torus $\mathbb{T}^d$. The case described in the former sections – $\mathcal{X}$ is any compact of $\mathbb{R}^d$ – is included in this new setting, since any compact $\mathcal{X}$ can be periodised to yield a domain with no boundaries. The forward operator kernel $\varphi : \mathcal{X} \to \mathcal{H}$ should also be differentiable in the Fréchet sense. The least squares term in BLASSO is denoted by the more general data term $R : \mathcal{H} \to \mathbb{R}^+$, the functional T_λ of the BLASSO will now be restricted to $\mathcal{M}^+(\mathcal{X})$ and denoted J; its Fréchet differential at point $\nu \in \mathcal{M}^+(\mathcal{X})$ is denoted J'_ν:

$$J(\nu) = \|y - \Phi\nu\|^2_{\mathcal{H}} + \lambda|\nu|(\mathcal{X}), \tag{5}$$

$$J'_\nu(x) = \langle \varphi(x), \nabla R \rangle_{\mathcal{H}} + \lambda \quad \text{for all } x \in \mathcal{X} \tag{6}$$

A comprehensive guide on its computation is given in appendix B.4. In the following, we describe the setting for non-negative measures of $\mathcal{M}^+(\mathcal{X})$, but it can be extended in a straightforward fashion [24] to signed measures of $\mathcal{M}(\mathcal{X})$ by performing the method on the positive then negative part of the signed measure (see Jordan decomposition in B.1). The Figure 6 sums up the chief quantities and relations introduced in this section, the reader is advised to refer to it whenever he or she needs a global view on the optimal transport problem.

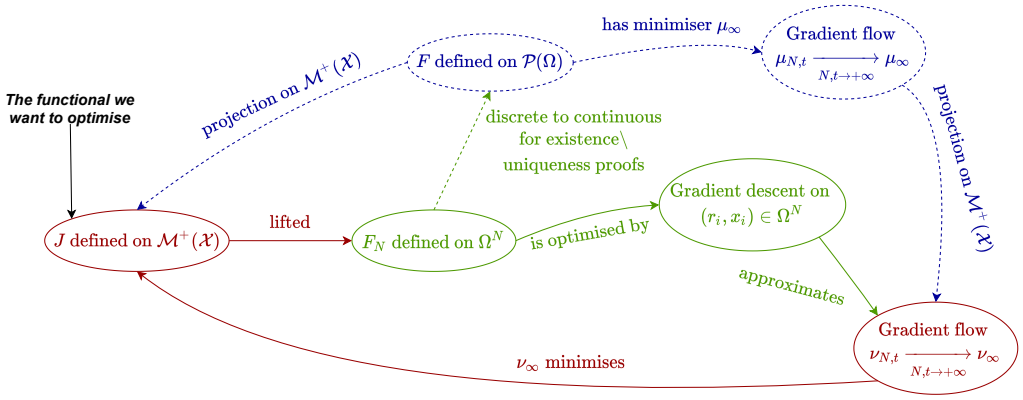

Figure 6. Digest of the important quantities mentioned in [7,24]: red refers to $\mathcal{M}^+(\mathcal{X})$ quantities, green to $\Omega^N \overset{\text{def.}}{=} (R^+ \times \mathcal{X})$ and blue to the Wasserstein space $\mathcal{P}_2(\Omega)$ and theoretical results. Dashed lines correspond to the theoretical section, and continuous lines indicate the numerical part.

Sparse optimisation on measures through optimal transport [7,24] relies on the approximation of the ground-truth positive measure m_{a_0,x_0} by a 'system of $N \in \mathbb{N}^*$ particles', *i.e.* an element of the space $\Omega^N \overset{\text{def.}}{=} (\mathbb{R}^+ \times \mathcal{X})^N$. The point is then to estimate the ground-truth measure by a gradient-based optimisation on the objective function:

$$F_N((r_1, x_1), \dots, (r_N, x_N)) \overset{\text{def.}}{=} \left\| y - \frac{1}{N} \sum_{i=1}^{N} {r_i}^2 \varphi(x_i) \right\|_{\mathcal{H}}^2 + \frac{\lambda}{N} {r_i}^2 \tag{7}$$

where (r_i, x_i) belongs to the lifted space $\Omega \overset{\text{def.}}{=} \mathbb{R}^+ \times \mathcal{X}$ endowed with a metric. Hence, the hope is that the gradient descent on F_N converges to the amplitudes and the positions of the ground-truth measure, despite the non-convexity of functional (7). Author of [24] proposes the definition of a suitable metric for the gradient of F_N, which enables separation of the variables in the gradient descent update. Let α, β two parameters such that $\alpha > 0$ and $\beta > 0$ and for any $(r, \theta) \in \Omega$, we define the Riemannian inner product of Ω called the *cone metric* endowing Ω is defined by $\forall (\delta r_1, \delta r_2) \in \mathbb{R}_+^2$, $\forall (\delta\theta_1, \delta\theta_2) \in \mathcal{X}^2$:

$$\langle (\delta r_1, \delta\theta_1), (\delta r_2, \delta\theta_2) \rangle_{(r,\theta)} \overset{\text{def.}}{=} \frac{\delta r_1 \delta r_2}{\alpha} + r^2 \frac{\langle \delta\theta_1, \delta\theta_2 \rangle_\theta}{\beta}.$$

We denote by $\langle \cdot, \cdot \rangle_\theta$ the metric on the manifold $\mathcal{X}$ at the point θ. The gradient of the functional F_N for all $i \in [\![1, N]\!]$ w.r.t. the cone metric writes down [6,24]:

$$\begin{cases} \nabla_{r_i} F_N = 2\alpha r_i J'_\nu(x_i) = -2\alpha r_i \lambda(\eta_\lambda - 1) \\ \nabla_{x_i} F_N = \beta\lambda \nabla J'_\nu(x_i) = -\beta\lambda \nabla \eta_\lambda \end{cases} \quad \text{for } \nu \overset{\text{def.}}{=} \sum_{i=1}^{N} {r_i}^2 \delta_{x_i},\ \eta_\lambda \overset{\text{def.}}{=} -J'_\nu / \lambda. \tag{8}$$

See appendix B.4 for more details on this computation. We now present the theoretical results on the particle gradient descent, which corresponding to the blue dashed lines in Figure 6. The reader is invited to refer to this figure any time he needs to get a hold on the broader picture.

3.3.1. Theoretical results

The main idea of these papers [7,24] boils down to the following observation: the minimisation of function (7) is a peculiar case of a more general problem, formulated in terms of measure of the lifted space Ω. The space is more precisely $\mathcal{P}_2(\Omega)$ subset

of $\mathcal{M}(\Omega)$, namely the space of probabilities with finite second moments endowed with the 2-Wasserstein metric *i.e.* the optimal transport distance: see Appendix B.5 for more details. Hence, the lift of the unknown $m \in \mathcal{M}^+(\mathcal{X})$ to $\mu \in \mathcal{P}_2(\Omega)$ enables to remove the asymmetry for discrete measures between position $x \in \mathcal{X}$ and amplitude $a \in \mathbb{R}^+$ by lifting $a\delta_x$ to $\delta_{(a,x)}$. The lifted functional now writes down for parameter $\lambda > 0$:

$$\forall \mu \in \mathcal{P}_2(\Omega), \quad F(\mu) \overset{\text{def.}}{=} \left\| y - \tilde{\Phi}\mu \right\|_{\mathcal{H}}^2 + \lambda \tilde{V}(\mu) \tag{9}$$

where $\tilde{\Phi}\mu \overset{\text{def.}}{=} \int_\Omega \phi(a,x) \mathrm{d}\mu(a,x)$ for $\phi(a,x) \overset{\text{def.}}{=} a\varphi(x)$ and $\tilde{V}$ is the TV-norm on the spatial component of the measure μ. The functional is non-convex, its Fréchet differential is denoted F' and for $u \in \Omega$:

$$F'(\mu)(u) \overset{\text{def.}}{=} \left\langle \tilde{R}'(\mu), \phi(u) \right\rangle_{\mathcal{H}} + \lambda$$

with $\tilde{R}' \overset{\text{def.}}{=} \left\| y - \int_\Omega \boldsymbol{\nabla}\phi(a,x) \mathrm{d}\mu(a,x) \right\|_{\mathcal{H}}^2$. Then, a discrete measure $\mu_N \overset{\text{def.}}{=} \frac{1}{N}\sum_i^N \delta_{a_i,x_i}$ of $\mathcal{P}_2(\Omega)$ can be also seen as an element of Ω^N from the standpoint of its components (a_i, x_i). It allows the authors of [7,24] to perform a precise characterisation of the source recovery conditions, through the measures and the tools of optimal transport such as gradient flow (see below).

Then one may run a gradient descent on the amplitudes and positions $(a_i, x_i) \in (\mathbb{R}^+ \times \mathcal{X})^N$ of the measure μ_N, in order to exploit the differentiability of the kernel φ. Note that the measure μ_N is over-parametrized *i.e.* its number of δ-peaks is larger compared to the number of spikes of the ground-truth measure: thus the particles, namely the δ-peaks of the space Ω are covering the domain $\mathcal{X}$ for their spatial part; see Figure 8 as an example, where μ_N is plotted in red dots.

Before giving the main results, we need to clarify the generalised notion of gradient descent to measure function called the *gradient flow* [36,37] from optimal transport theory, the main ingredient in the particle gradient descent. Let $F : \mathbb{R}^d \to \mathbb{R}$ be the objective function with certain regularity, a gradient flow describes the evolution of a curve $x(t)$ such that its starting point at $t = 0$ is $x_0 \in \mathbb{R}^d$, evolving by choosing at any time t in the direction that decreases the most the function F [37].

$$\begin{cases} x'(t) = -\boldsymbol{\nabla}F(x(t)) \quad \text{pour } t > 0 \\ x(0) = x_0. \end{cases}$$

The interest of gradient flow is its extension to spaces X with no differentiable structure. In the differentiable case one can consider the discretisation of the gradient flow *i.e.* the sequence defined for a step-size $\tau > 0, k \in \mathbb{N}^*$:

$$x_{k+1}^\tau \in \operatorname*{argmin}_{x \in X} F(x) + \frac{\left|x - x_k^\tau\right|^2}{2\tau}.$$

It is the implicit Euler scheme for the equation[7] $(x^\tau)' = -\boldsymbol{\nabla}F(x^\tau)$. The gradient flow is then the limit (under certain hypotheses) of the sequence $(x_k^\tau)_{k \geq 0}$ for $\tau \to 0$ for a starting point $x_0 \in X$. Gradient flow can be extended to metric space: indeed, for a metric space (X, d) and a map $F : X \to \mathbb{R}$ lower semi-continuous one can define the discretisation of gradient flow by the sequence

[7] or the weaker $(x^\tau)' \in \partial F(x^\tau)$ if F is convex and non-smooth.

$$x_{k+1} \in \underset{x \in X}{\operatorname{argmin}} F(x) + \frac{d(x, x_k)^2}{2\tau}. \tag{10}$$

In the case of the metric space of probability measures *i.e.* the measures with unitary mass, the limit $\tau \to 0$ of the scheme exists and converges to the unique gradient flow starting at x_0 element of the metric space. A typical case is the space of probabilities with finite second moments $\mathcal{P}_2(\Omega)$, endowed with 2-Wasserstein metric *i.e.* the optimal transport distance (see Appendix B.5): a gradient flow in this space $\mathcal{P}_2(\Omega)$ is a curve $t \mapsto \mu_t$ called a *Wasserstein gradient flow* starting at $\mu_0 \in \mathcal{P}_2(\Omega)$, for all $t \in \mathbb{R}^+$ one have $\mu_t \in \mathcal{P}_2(\Omega)$, obeying the partial differential[8] equation in the sense of distributions:

$$\partial_t \mu_t = -\operatorname{div}(\mu_t \nabla F'(\mu_t)). \tag{11}$$

This equation ensures the conservation of the mass, namely at each time $t > 0$ one have $|\mu_t|(\Omega) = |\mu_0|(\Omega)$. Hence, despite the lack of differentiability structure of $\mathcal{P}_2(\Omega)$ which forbids straightforward application of classical gradient-based algorithm, one can perform an optimisation on the space through gradient flow to reach a *minimum* of F by discretizing (11).

The interesting case of a gradient flow in $\mathcal{P}_2(\Omega)$ is the flow starting at $\mu_{N,0} \overset{\text{def.}}{=} 1/N \sum_{i=1}^N \delta_{(a_i^0, x_i^0)}$, uniquely defined by the equation (11), which writes down for all $t \in \mathbb{R}^+$: $\mu_{N,t} \overset{\text{def.}}{=} 1/N \sum_{i=1}^N \delta_{(a_i(t), x_i(t))}$ where $a_i : \mathbb{R}^+ \to \mathbb{R}^+$ and $x_i : \mathbb{R}^+ \to \mathcal{X}$ are continuous maps. This path $(\mu_{N,t})_{t \geq 0}$ is a Wasserstein gradient flow, and uses N Dirac measures over Ω to optimise the objective function F in (9). When the number of particles N goes to infinity and if $\mu_{N,0}$ converges to some $\mu_0 \in \mathcal{P}_2(\mathcal{X})$, the gradient flow $(\mu_{N,t})_{t \geq 0}$ converges to the unique Wasserstein gradient flow of F starting from μ_0, described by the time-dependent density $(\mu_t)_{t \geq 0}$ valued in $\mathcal{P}_2(\mathcal{X})$ obeying the latter partial differential equation (11).

For these non-convex gradient flow, the authors of [7] give a consistent result for gradient based optimisation methods: under certain hypothesis, the gradient flow $(\mu_{N,t})_{t \geq 0}$ converges to global *minima* in the over-parametrization limit *i.e.* for $N \to +\infty$. It relies on two important assumptions that prevent the optimisation from being blocked in non-optimal points:

- homogeneity[9] of ϕ in order to select the correct magnitude for each feature, or at least partially 1-homogeneity (*i.e.* boundedness of φ in [7]);
- diversity in the initialisation of parameters, in order to explore all combinations of features. Too few or too close particles will not reach all source peaks and will only yield local *minima*.

We can then introduce the fundamental result for the many particle limit [7], the mean-field limits of gradient flows $(\mu_{N,t})_{t \geq 0}$ despite the lack of convexity of these gradient flows:

Theorem 3.3.2 (Global convergence – informal)**.** *If the initialisation $\mu_{N,0}$ is such that $\mu_0 \overset{\text{def.}}{=} \lim_{N \to +\infty} \mu_{N,0}$ support separates[10] $\{-\infty\} \times \mathcal{X}$ from $\{+\infty\} \times \mathcal{X}$ then the gradient flow μ_t weakly-$*$ (see appendix B.1) converges in $\mathcal{P}_2(\Omega)$ to a global* minimum *of F and we also have:*

$$\lim_{N,t \to \infty} F(\mu_{N,t}) = \min_{m \in \mathcal{M}^+(\mathcal{X})} J(m).$$

[8] $\operatorname{div}(m) = \sum_{i=1}^d \frac{\partial m}{\partial x_d}$ for all $m \in \mathcal{M}(\mathcal{X})$. Derivatives ought to be understood in the distributional sense.

[9] A function f between vector spaces is positively p-homogeneous if, for $\lambda > 0$ and argument x, one have $f(\lambda x) = \lambda^d f(x)$.

[10] The support of a measure m is the complement of the largest open set on which m vanishes. In an ambient space $\mathcal{X}$, we say that a set C separates the sets A and B if any continuous path in $\mathcal{X}$ with endpoints in A and B intersects C.

Limits can be interchanged; the interested reader might take a look at [7] for precise statements and exact hypothesis (boundary conditions, 'Sard-type' regularity *e.g.* φ is d-times continuously differentiable, *etc*).

Since we have a convergence result, we can then investigate the numerical implementation. This optimisation problem is tractable thanks to the Conic Particle Gradient Descent algorithm [24] denoted CPGD: the proposed framework involves a slightly different gradient flow $(\nu_t)_{t\geq 0}$ defined through a projection of $(\mu_t)_{t\geq 0}$ onto $\mathcal{M}^+(\mathcal{X})$. This new gradient flow $(\nu_t)_{t\geq 0}$ is defined for a specific metric in $\mathcal{M}^+(\mathcal{X})$, which is now a trade-off between Wasserstein and Fisher-Rao[11] metric [24], it is then called a *Wasserstein-Fisher-Rao gradient flow*. Then the Wasserstein-Fisher-Rao gradient flow starting at $\nu_{N,0} \overset{\text{def.}}{=} \sum_{i=1}^N a_i^0 \delta_{x_i^0}$ in $\mathcal{M}^+(\mathcal{X})$ writes down $t \mapsto \nu_{N,t} \overset{\text{def.}}{=} \frac{1}{N}\sum_{i=1}^N r_i(t)^2 \delta_{x_i(t)}$ in $\mathcal{M}^+(\mathcal{X})$, rather than the Wasserstein flow $t \mapsto \mu_{N,t} \overset{\text{def.}}{=} \frac{1}{N}\sum_{i=1}^N \delta_{r_i(t),x_i(t)}$ starting at $\mu_{N,0}$ in $\mathcal{P}_2(\Omega)$. The partial differential equation of a Wasserstein-Fisher-Rao flow writes down:

$$\partial_t \nu_t = -4\alpha \nu_t T_\lambda(\nu_t) + \beta \operatorname{div}\left(\nu_t \nabla J'_\nu(\nu_t)\right) \tag{12}$$

for the two parameters $\alpha, \beta > 0$ arising from the cone metric, α tunes the Fisher-Rao metric weight while β tunes the Wasserstein metric one. All statements on convergence could be made alternatively on μ_t or ν_t, we have indeed the same theorem:

Theorem 3.3.3 (Global convergence – informal)**.** *If ν_0 has full support (its support is the whole set $\mathcal{X}$) and $(\nu_t)_{t\geq 0}$ converges for $t \to +\infty$ then the limit is a global* minimum *of J. If $\nu_{N,0} \xrightarrow[N\to+\infty]{} \nu_0$ in the weak-$*$ sense then:*

$$\lim_{N,t\to\infty} J(\nu_{N,t}) = \min_{m\in\mathcal{M}^+(\mathcal{X})} J(m).$$

Summary (3rd algorithm theoretical aspects): we introduced the proposed solution of [7,24] namely approximate the source measure by a discrete non-convex objective function of amplitudes and positions. The analytical study of the discrete function is an uphill problem and could be tackled thanks to the recast of the problem in the space of measures. Then, we exhibited the theoretical framework on gradient flows, understood in the sense of generalisation of gradient descent in the space of measures. Eventually, we presented the convergence results of the gradient flow denoted $(\nu_t)_t$ towards the *minimum* of the BLASSO, thus enabling results for the convergence. Gradient descent on the discrete objective approximates well the gradient flow dynamic and can then benefits from the convergence results exhibited before.

We now discuss the numerical results of the particle gradient descent. The reader is advised to take a look at the Figure 6, more precisely at red and green ellipses, to get a grasp on the numerical part.

3.3.4. Numerical results

We recall that a gradient flow $(\nu_{N,t})_{t\geq 0}$ starting at $\overset{\text{def.}}{=} 1/N \sum_{i=1}^N \left(r_i^{(0)}\right)^2 \delta_{x_i^{(0)}}$ can be seen as a (time continuous) generalisation of gradient descent in the space of measures, allowing precise theoretical statements on the recovery conditions. To approach this gradient flow, we use the Conic Particle Gradient Descent algorithm [24] denoted CPGD: the point is to discretise the evolution of the gradient flow $t \mapsto \nu_{N,t}$ through a numerical scheme on (12). This consists in a gradient descent on the amplitudes r and positions x through the gradient

[11] also called Hellinger metric.

of the functional F_N in equation (8), a strategy which approximates well the dynamic of the gradient flow [24].

This choice of gradient with the cone metric enables multiplicative updates in r and additive in x, the two updates being independent of each other. Then the algorithm consists in a gradient descent with the definition of $r_i'(t)$ and $x_i'(t)$ according to [6,24]:

$$\begin{cases} r_i'(t) = -2\alpha r_i \lambda(\eta_\lambda(x_i(t)) - 1) \\ x_i'(t) = -\beta\lambda \nabla \eta_\lambda(x_i(t)) \end{cases} \tag{13}$$

thanks to gradient in equation (8), for the mirror retraction[12] and $\eta_\lambda = -J_\nu'/\lambda$. The structure of the CPGD is presented in the algorithm 4. Note that the multiplicative updates in r yields an exponential of the certificate, and that the updates of the quantities r, x are separated.

Algorithm 4: Conic Particle Gradient Descent Algorithm.

Input: Gradient step sizes $\alpha, \beta > 0$ and $N \geq 1$ the number of particles.

1 Draw uniformly the initialisation discrete measure with amplitude-positions $\left(r_i^{(0)}, x_i^{(0)}\right)_{i=1}^N \in (R^+ \times \mathcal{X})^N$ such that $a_i^{(0)} = \left(r_i^{(0)}\right)^2$:

$$\nu^{(0)} \overset{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^N a_i^{(0)} \delta_{x_i^{(0)}}.$$

2 **while** stopping criterion is not met **do**

3 Mirror descent step, for all $i = 1, \ldots, N$ update:

$$\begin{aligned} r_i^{(k+1)} &= r_i^{(k)} \exp\left(2\alpha\lambda\left(\eta^{(k)}\left(x_i^{(k)}\right) - 1\right)\right) \\ x_i^{(k+1)} &= x_i^{(k)} + \beta\lambda \nabla \eta^{(k)}\left(x_i^{(k)}\right) \end{aligned}$$

where $\eta^{(k)} = -\frac{J'(\nu^{(k)})}{\lambda}$, $\nu^{(k)} = \frac{1}{N}\sum_{i=1}^N a_i^{(k)} \delta_{x_i^{(k)}}$ and $a_i^{(k)} = \left(r_i^{(k)}\right)^2$.

4 **end**

This algorithm has rather easy and cheap iterations: to reach an accuracy of ε – *i.e.* a distance such as the ∞-Wasserstein distance between the source measure m_{a_0,x_0} and the reconstructed measure m^* is below ε – the CPGD yields a typical complexity cost of $\log\left(\varepsilon^{-1}\right)$ rather than $\varepsilon^{-1/2}$ for convex program [24, Theorem 4.2]. A reconstruction from the latter 1D Fourier measurements is plotted in Figure 7, the reconstruction is obtained through two gradient flows, the former on the positive measures to recover the positive δ-peaks of the ground-truth and the latter on the negative measures to recover the negative one: the merging of the two results gives the reconstructed δ-peaks. The noiseless reconstruction[13] for 2D Gaussian convolution with the same setting as the Frank-Wolfe section is plotted in Figure 8. One can see that the spikes are well-recovered as some non-zero red and blue particles cluster around the three δ-peaks.

[12] The notion of *retraction* compatible with cone structure is central: in the Riemann context a retraction is a continuous mapping that maps a tangent vector to a point on the manifold. Formally, one could see it as a way to enforce the gradient evaluation to be mapped on the manifold. See [24] for other choices of compatible retractions and more insights on these notions.

[13] See our GitHub repository for our implementation: https://github.com/XeBasTeX

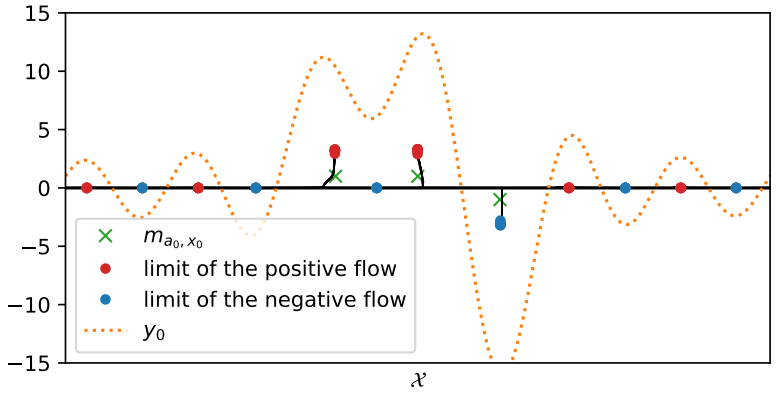

Figure 7. Reconstruction by *Conic Particle Gradient Descent* for a 1D Fourier operator in a noiseless setting, with the same ground-truth spikes as the former section. Implementation is an adaptation of [24], $\alpha = \beta = 1 \times 10^{-3}$ and $\lambda = 1$ for 1000 iterations.

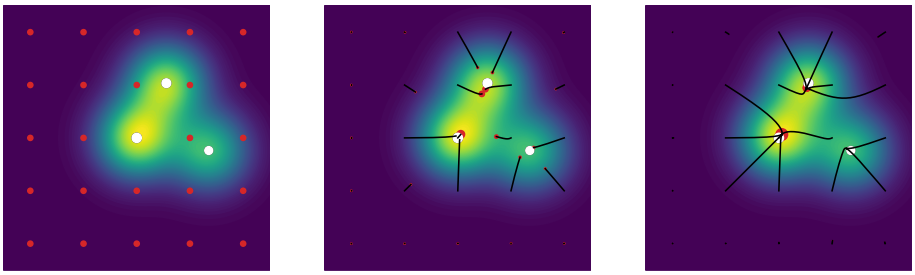

Figure 8. (**a**) Initialisation $k = 0$ (**b**) Mid-computation $k = 150$ (**c**) End of the computation $k = 1000$. *Conic Particle Gradient Descent* applied for 2D Gaussian deconvolution, the red dots are the particle measure $\nu^{(k)}$ (size of dot proportional with amplitude), the 3 white dots are the source measure, the image in the background is the noiseless acquisition y_0 and the black line are the paths of the particles $\nu^{(k)}$ — all the paths constitute the gradient flow $(\nu_t)_{t\geq 0}$. Implementation is an adaptation of [24], $\alpha = \beta = 1 \times 10^{-2}$ and $\lambda = 1$.

Summary (3rd algorithm numerical aspects): the gradient flow $(\nu_t)_t$ is computable by the Conic Particle Gradient Descent algorithm, consisting in an estimation through a gradient (w.r.t. cone metric) descent on both amplitudes and positions of an over-parametrised measure, namely a measure with a fixed number of δ-peaks exceeding the source's one. The iterations are cheaper than the SFW presented before, but the CPGD lacks guarantees in a low-noise regime.

To sum-up all the pros and cons of these algorithms, we give the table 1 for a quick digest. Since the CPGD lacks guarantees on the global optimality of its output, the following section will use the conditional gradient and more precisely the *Sliding Frank-Wolfe* in order to tackle the SMLM super-resolution problem.

Table 1. Pros and cons for the different off-the-grid algorithm strategies, Semi-definite programming (SDP) v. *Sliding Frank-Wolfe* (SFW) algorithm v. *Conic Particle Gradient Descent* (CPGD).

Algorithm	Operator	Space $\mathcal{X}$	Convergence rate	Computation time	Tuning parameters
SDP [1]	Fourier	Torus $\mathbb{T}^d$	Asymptotic	Mild	λ
SFW [2,13]	All	Any compact	Sublinear	Long	λ
CPGD [24]	All	Torus $\mathbb{T}^d$	$\log(\varepsilon)$	Quick	λ, α, β

4. Applications and results in SMLM imaging

As an illustration of off-the-grid applications' in this review, we propose to solve the super-resolution problem, aiming to retrieve biological structures at very small scales.

4.1. Metrics of quality of reconstruction

If one has access to the ground-truth *i.e.* the real position of the point sources, one is able to assess the quality of the reconstruction by:

- detection metrics, such as the Jaccard index;
- quality of reconstruction metrics, such as the L^2 norm in the discrete case.

Detection metrics can be applied to the off-the-grid output in a straightforward manner. We will rather focus in this part on the 'quality of reconstruction' metric. Any of the former algorithms returns a list of Dirac measures, which can be compared with the ground-truth measure m_{a_0,x_0}. This comparison cannot be done with discrete tools, such as the L^2 norm of the reconstructed acquisitions: we cannot compare an element of $\mathcal{M}(\mathcal{X})$ with $\mathrm{L}^2(\mathcal{X})$. Examining the L^2 norm of the x_i vector of reconstructed positions against the $x_{0,i}$ vector is not sufficient either: we need the same number of elements for x and x_0, we have to sort the vector of positions, and we have no guarantee that the matching of one position of x with another of x_0 is the right one.

Hence, a distance on the measure space is the good tool of comparison. We will use in the following the Wasserstein 1-distance $\mathcal{W}_1$ [38]: see the appendix B.5 for some recall on the useful definition and more insights on the optimal transport setting used in this section. The Wasserstein distance with measures of equal mass is defined[14] as :

Definition 4.1.1 (Balanced optimal transport)**.** *For $0 \leq p < +\infty$ and $m_1, m_2 \in \mathcal{M}(\mathcal{X})$ such that $|m_1|(\mathcal{X}) = |m_2|(\mathcal{X})$, the p-Wasserstein distance is written:*

$$\mathcal{W}_p(m_1, m_2) \overset{\text{def.}}{=} \left(\min_{\gamma \in \Gamma(m_1,m_2)} \int_{\mathcal{X}\times\mathcal{X}} |u-v|^p \,\mathrm{d}\gamma(u,v) \right)^{1/p}. \tag{14}$$

$\Gamma(m_1, m_2)$ is the set of transport plans between m_1 and m_2, one can take a look at the Appendix B.5 for more insights on this notion. However this notion is not sufficient for our application since the metric can only take measures of equal masses (*i.e.* equal TV-norm) as an input. In the case of a discrete measure, we recall that mass is simply the sum of the modulus of individual amplitudes: hence in general we cannot compare a source measure and a reconstructed measure with differing amplitudes. The classic solution is then to distribute the unit mass, divided by the number of spikes, uniformly over each δ-peak of the discrete measure. Still, it would be way more convenient to incorporate the case of differing masses in the metric. The proper metric to compare two measures of different masses is called the Kantorovtich-Rubinstein norm also referred as the Flat Metric [38–40].

Definition 4.1.2 (Unbalanced optimal transport)**.** *Let us denote $m \in \mathcal{M}(\mathcal{X})$ of finite first moment and $\tau > 0$, the following quantity is called Kantorovtich-Rubinstein norm:*

$$F_\tau(m) \overset{\text{def.}}{=} \sup_{f \in \mathscr{C}_\mathrm{b}(\mathcal{X})} \left(\int_{\mathcal{X}} f \,\mathrm{d}m, \|f\|_{\infty,\mathcal{X}} \leq \tau, f \text{ Lipschitz}, \|f\|_{\mathrm{Lip}} \leq 1 \right)$$

[14] Actually it is well-defined on the subset $X \overset{\text{def.}}{=} \left\{ m \in \mathcal{M}(\mathcal{X}), |m|(\mathcal{X}) \leq \|y\|_{\mathcal{H}}^2/2\lambda \right\}$ because only the bounded subset of $\mathcal{M}(\mathcal{X})$ are metrizable for the weak-$*$ topology, so we have to restrain the set of measures to X in order to reach a Polish space *i.e.* the convenient framework for this OT-based metric, see the Appendix B.5. Since all solutions of the BLASSO belong to X [4] we will keep this slight abuse of notation in the rest of the paper.

where $\|f\|_{\mathrm{Lip}}$ *is the Lipschitz constant of* f. *We then define the* Flat Metric d_τ *for* $m_1, m_2 \in \mathcal{M}(\mathcal{X})$ *of finite first moments:*

$$d_\tau(m_1, m_2) \overset{\text{def.}}{=} F_\tau(m_1 - m_2).$$

The parameter τ is homogeneous to a distance, and it is understood in the optimal transport sense as the cost of creating/destroying a Dirac measure. The Flat Metric coincides with the 1-Wasserstein distance, for m_1, $m_2$of equal masses, when $\tau \to +\infty$ [39]; it also coincides with the total variation norm of $m_1 - m_2$ when $\tau \to 0$. Then it may be seen as an interpolation between the total variation norm and the 1-Wasserstein norm. Moreover, when the number of δ-peaks is correctly estimated, the Flat Metric stands for the mean error in terms of localisation and is similar to the RMSE [40]. Eventually, the Flat metric can be extended to discrete reconstruction *i.e.* images on a fine grid; this metric is then a method applicable to discrete reconstruction, namely images with a finer grid.

To sum-up, there are two possibilities if one wants to compare the reconstructed measure and the ground-truth one:

- let the source measure be composed of N spikes, we set the amplitude of each δ-peak at $1/N$. We apply the same procedure to the reconstructed (with differing or not number of spikes), hence dividing uniformly the unit mass over all the δ-peaks of the considered measure. Therefore, the reconstructed luminosity is not considered as relevant and discarded: we can compute directly the 1-Wasserstein distance, since it is equal to the Flat Metric in this case;
- we want to account for the luminosity, and we use the Flat Metric to compare the reconstructed measure against the ground-truth one.

Summary: classic quality of reconstruction metrics such as the $\mathrm{L}^2(\mathcal{X})$ norm cannot be straightforwardly applied to off-the-grid reconstruction. Instead, one could use optimal transport score such as the Flat Metric: it accounts for both amplitude and position reconstructions, while it can be easily extended to discrete reconstruction (images on a fine grid).

4.2. Results for a SMLM stack

In super-resolution for biomedical imaging, one wants to retrieve some fine scale details to better study biological structures of interest. Indeed, the studied bodies are generally smaller than the Rayleigh limit at 200 nm, a length at which the phenomenon of light diffraction comes into play. This diffraction causes a blurring of the image, which can be described as a convolution of the image by the PSF mentioned above. Hence, we want to perform a *deconvolution i.e.* remove the blur of diffraction to get a super-resolved image. It is worth noticing that other imaging systems exist, for which the inverse problems to solve are a bit different from deconvolution: *e.g.* Nuclear Magnetic Resonance spectroscopy with Fourier measurements [41], MA-TIRF with Laplace [13].

In order to enhance spatial resolution over standard diffraction-limited microscopy techniques and allow imaging of biological structures below the Rayleigh criterion, one can use SMLM, which stands for 'Single Molecule Localisation Microscopy '. It is a compelling technique in fluorescence microscopy to tackle the super-resolution problem [42]. It requires photoactivable fluorophores with, roughly speaking two states, for example '*On*' and '*Off*'. These molecules are therefore only visible on the acquisitions in the '*On*' case, and the idea is then to light up some molecules in the sample to make the acquisition and to be able to locate them precisely; the fluorescent molecules are bound to the biological structure and since only a few molecules are emitting in one frame, the resulting image is rather sparse which allows accurate localisation. This process is repeated until all the molecules have been lit and imaged. All the positions of the imaged molecules frame-by-frame can then be put together to form a super-resolved image that go below the diffraction barrier, ridden of the degradation by the process of acquisition (blur, noise, *etc*.). The quality of the image

reconstruction is naturally limited by the number of acquisitions necessary to reconstruct the image, which implies a cost in time (precious insofar as the organism studied moves) and in physical memory and by the density of fluorophores lit at each stage. Indeed, there is a risk of overlap hindering the localisation of the molecules since the separation criterion is not matched.

Off-the-grid methods can be applied to any SMLM stack with only the knowledge of the forward operator, the acquisition system's PSF in this case. In this review, a gridless method based on Sliding Frank-Wolfe is tested on an 2D SMLM acquisition stack from the 2013 EPFL Challenge[15]. For this purpose, we consider the first image of the stack, locate the source points, and store the coordinates of these points. Then, we move on to the second image, we locate the source points, and so on. Note that off-the-grid method with this variational approach are not the only method taking advantage of a continuous domain like the PSF-fitting such as DAOSTORM [43], *etc*.

Deconvolution is a first challenge to solve this inverse problem, but we must also take into account the noise. One has to deal with three main types of noise on these acquisitions:

- photon noise (also known as shot noise or quantum noise) is due to the quantum nature of light. It arises from the fact that fluorophores emit photons randomly, so that between t and $t+1$ (exposure time), a variable number of photons have been emitted, and therefore a variable number of photons have been collected by the sensor. Thus the amplitude of the electrical signal generated in the sensor (at each pixel) fluctuates according to a Poisson statistic;
- the dark current is a phenomenon due to the natural agitation of electrons. This natural agitation is sufficient to occasionally eject an electron from the valence band to the conduction band without any photoelectric effect. Additional charges are therefore created which interfere with the signal. The number of electrons generated by thermal agitation follows a Poisson distribution;
- amplification and readout noise. This noise is produced by the electronic circuit that amplifies and converts the electron packets into voltage. It is generally modelled by a Gaussian noise.

Thus, we have several noises that pollute each of the observed images. To deal with this ill-posed inverse problem, we use the results on BLASSO, with the least-squares term as the data-fitting term and the TV norm as the regulariser of the inverse problem. In the Bayesian approach the least-square term is modelling the maximum of likelihood when the acquisition is polluted by Gaussian noise, hence our model is making the approximation of Gaussian noise. Measurements are discrete so at each image one have to deal with images with $N_1 \times N_2$ pixels, each of them with size (b_1, b_2). Let $(c_{i,1}, c_{i,2})$ be the centre of the ith pixel, we denote the ith camera pixels by

$$\Omega_i \overset{\text{def.}}{=} (c_{i,1}, c_{i,2}) + \left[-\frac{b_1}{2N_1}, \frac{b_1}{2N_1}\right] \times \left[-\frac{b_2}{2N_2}, \frac{b_2}{2N_2}\right].$$

We can then clarify the forward operator $\Phi : m \mapsto \mathbb{R}^{N_1 N_2}$ which encapsulates the integration over camera pixels [13], indeed with the evaluation of the discrete Gaussian kernel φ with standard deviations σ, for $i \in \{1, \ldots, N_1 N_2\}$:

$$[\varphi(x)]_i \overset{\text{def.}}{=} \frac{1}{2\pi\sigma^2} \int_{\Omega_i} e^{-\left(\frac{(x_1-s_1)^2}{2\sigma^2} + \frac{(x_2-s_2)^2}{2\sigma^2}\right)} \mathrm{d}s_1 \mathrm{d}s_2.$$

In the SMLM data set, one has the PSF standard deviation $\sigma = 149.39$nm and $N_1 = N_2 = 100$nm. The reconstruction is performed by our implementation of the Sliding Frank-

[15] https://srm.epfl.ch/DatasetPage?name=MT0.N1.HD

Wolfe in `python`[16] insofar as it is the more robust method available: indeed it works with Gaussian kernel, it has proven results in a noise regime, *etc*. The results are presented in Figure 9. The stack of 2500 images of 64×64 is qualified as high density with high SNR: the number of activated fluorophores is quite important, and the noise is not negligible[17].

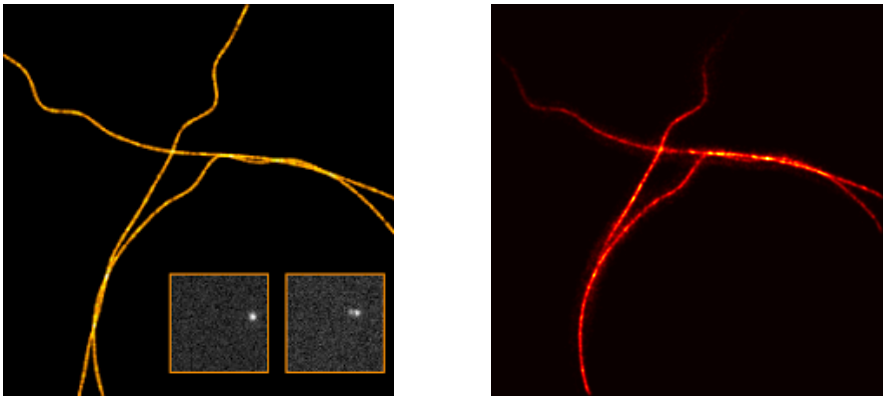

Figure 9. (**a**) Ground-truth tubulins, two excerpts of the stack in the square below: convolution + all noise described before. (**b**) Reconstructed measure by *Sliding Frank-Wolfe* visualised through Gaussian kernel with a smaller σ (see text).

Flat metric between the reconstructed measure $m_{a,x}$ and the ground-truth measure m_{a_0,x_0} is then computed, and it reaches $d_\tau(m_{a,x}, m_{a_0,x_0}) = 1.7 \times 10^{-2}$. The reconstruction is convincing and well capture the fine details of the biological structures, one can clearly see the interweaving tubulins in the right part of the image.

Note that an interesting feature of the gridless reconstruction is that once the Radon measure is computed, it is straightforward to plot it through any operator on a fine grid of one choice. Indeed, as one cannot represent a discrete measure m, we rather plot Φm where Φ is the PSF with a slightly smaller variance, in order to clearly see the deconvolution. In all of our reconstruction, we convolve the reconstruction through the PSF with variance $\sigma/6$ and plot it on a grid 32 times finer. As a matter of comparison discrete methods are performed for a fixed fine grid, and if one want a finer reconstruction one has to recompute everything.

We finally test the off-the-grid reconstruction on a real data set of tubulins with high density molecules, provided by the 2013 IEEE ISBI SMLM challenge. In this stack of 500 frames of 128×128 pixels, the FWHM (full width at half maximum) of the acquisition system is estimated at 351.8 nm. We recall that the FWHM is the width of the Gaussian curve measured between those points on the y-axis which are half the maximum amplitude, also note that it is linked to the variance σ by $\text{FWHM} = 2\sqrt{2\ln 2} \times \sigma$. We compare the reconstruction of the off-the-grid method with the output of the Deep-STORM [44] algorithm, touted as the algorithm with the most visually compelling results. The reconstructions of the gridless method and the Deep-STORM algorithm are presented in Figure 10, where one can appreciate the reconstruction by off-the-grid on fine details. The reconstruction seems a small bit blurry compared to Deep-STORM, due to the plotting through a small spread Gaussian kernel. However, it is noteworthy that both comparison perform well to retrieve the filaments, in particular in the enhancing yellow circles: the off-the-grid reconstruction seems to better preserve the structure compared to the Deep-STORM's rough output. The quality of the reconstruction is notably interesting for off-the-grid reconstruction since it does not require any test sets to yield this reconstruction, on the contrary to Deep-STORM. The only data needed is the knowledge of (an estimation of) the forward operator, the off-the-grid reconstruction can be then performed from any input without having to train the model on different type and level of noise.

[16] See our GitHub repository for our PyTorch implementation: https://github.com/XeBasTeX

[17] See https://srm.epfl.ch/DatasetPage?name=MT0.N1.HD for more insights.

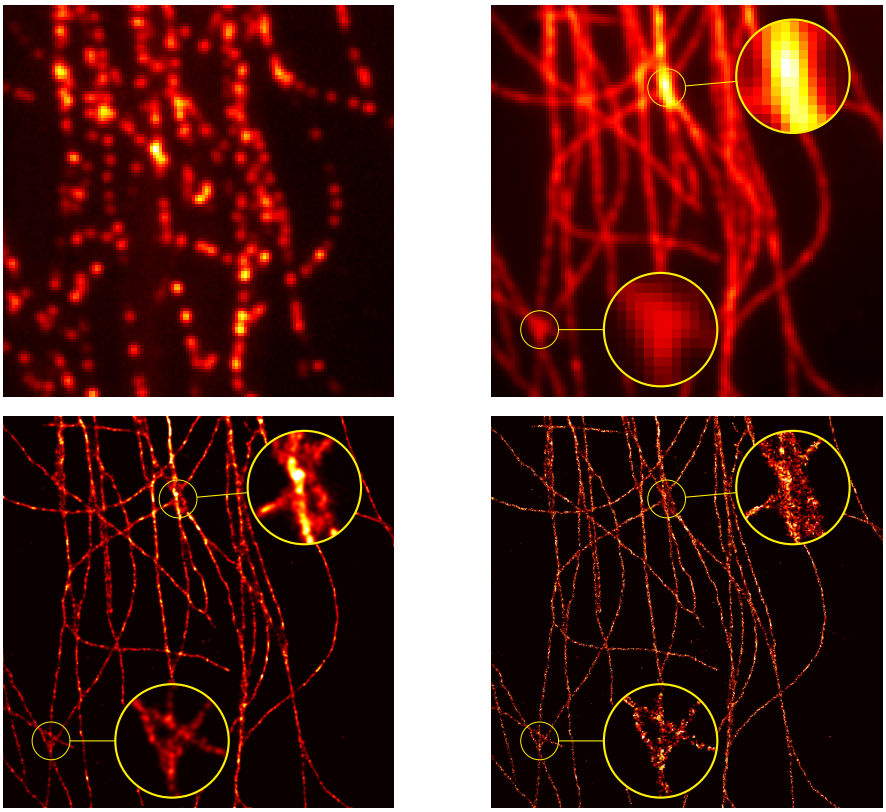

Figure 10. (**a**) Excerpt of the stack (**b**) Mean of the stack (**c**) Reconstruction by off-the-grid method (**d**) Deep-STORM.

Shorthand: we tested an off-the-grid method on both SMLM synthetic and experimental data set. The gridless problem is tractable thanks to the Sliding Frank-Wolfe algorithm, and yields compelling results. The results are all the more interesting since there is only one parameter, handy to tune and robust w.r.t. noise. Thus, it can be easily adapted to any other dataset with known acquisition operator.

5. Conclusion

We described in this review the off-the-grid variational settings for the sparse spike problem, through the definition of the space of signed measures $\mathcal{M}(\mathcal{X})$ and the functional BLASSO defined over this set. Thanks to the trade-off between the convexity of the functional and the infinite dimensional, non-reflexive space of optimisation $\mathcal{M}(\mathcal{X})$; the BLASSO can be defined to solve the sparse spike recovery problem. We review in this paper the theoretical guarantees to reach the correct *minimum* as the literature provides multiple results, in particular a sharp criterion for stable spikes recovery under a low noise regime. Numerical methods to tackle the BLASSO problem were also discussed, with insights on the SDP approach which is asymptotically exact but only suited for Fourier measurements, the Frank-Wolfe approach with known rate of convergence but a high computation load and the Conic Particle Gradient Descent with cheap iterations but lacks of guarantees. We were finally able to present the result of the off-the-grid approach with Sliding Frank-Wolfe algorithm in the case of SMLM synthetic *data* and real data from the EPFL Challenge, and to illustrate the usefulness of these methods to recover fine-scale details.

Funding: This work has been supported by the French government, through the UCA DS4H Investments in the Future project managed by the National Research Agency (ANR) with the reference number `ANR-17-EURE-0004` and through the 3IA Côte d'Azur Investments in the Future project managed by the National Research Agency with the reference number `ANR-19-P3IA-0002`.

Data Availability Statement: All the data are available at the git repository https://github.com/XeBasTeX/Journal-of-Imaging-2021 or its mirror https://gitlab.inria.fr/blaville/Journal-of-Imaging-2021.

Conflicts of Interest: the authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PSF	Point Spread Function
FWHM	Full width at half maximum
SNR	Signal-to-noise ratio
LASSO	Least Absolute Shrinkage and Selection Operator
BLASSO	Beurling-LASSO
SFW	Sliding Frank-Wolfe
OT	Optimal Transport
CPGD	Conic Particle Gradient Descent

Appendix A. Notations table

Table A1. Main notations used in the review.

d	dimension of the ambient space
$\mathcal{X}$	ambient space of the spike positions *e.g.* the torus, $[0,1]^d$, $\mathbb{R}^d$, *etc.*
$\mathscr{C}_0(\mathcal{X})$	space of evanescent continuous functions
$\mathcal{M}(\mathcal{X})$	space of signed Radon measures
$\mathcal{M}^+(\mathcal{X})$	space of non-negative Radon measures
$\mathcal{M}_{\mathbb{C}}(\mathcal{X})$	space of complex Radon measures
$\mathcal{H}$	Hilbert space, typically $\mathrm{L}^2(\mathcal{X})$
m	Radon measure
$\lvert m\rvert(\mathcal{X})$	TV norm of m
δ	Dirac measure
a, x	respectively amplitudes and positions of the spike $a\delta_x$
N	number of Dirac measures in a discrete measure
λ	regularisation parameter in ($\mathcal{P}_\lambda(y)$)
Φ	forward acquisition operator with kernel φ and adjoint Φ^*
$\mathcal{F}_n$	forward Fourier acquisition operator with n measurements
p_λ	solution of the dual problem ($\mathcal{D}_\lambda(y)$)
η_λ	dual certificate of ($\mathcal{P}_\lambda(y)$)
T_λ	BLASSO ($\mathcal{P}_\lambda(y)$) functional
Ω	Lifted space $\Omega \stackrel{\text{def.}}{=} \mathbb{R}^+ \times \mathcal{X}$
J	BLASSO functional on $\mathcal{M}^+(\mathcal{X})$
R	data-term $R : \mathcal{M}(\mathcal{X}) \to \mathcal{H}$
F_N	'discrete' functional on Ω^N
F	functional on $\mathcal{M}(\Omega)$
$(\mu_t)_t$, $(\nu_t)_t$	Gradient flows respectively in $\mathcal{P}_2(\Omega)$ and $\mathcal{M}^+(\mathcal{X})$
α, β	Cone metric/ Fisher-Rao-Wasserstein tuning parameters
$\mathcal{W}_p$	Wasserstein distance of order p
$\mathcal{P}_2(\Omega)$	space of probability measures with 2nd moment endowed with $\mathcal{W}_2$

Appendix B. Useful definitions and notions

Appendix B.1. Details on functional analysis

Definition B.1.1. *Two Radon measures μ and ν of $\mathcal{M}(\mathcal{X})$ are called* singular *if there exists two disjoints subsets A, B of the σ-algebra of $\mathcal{X}$ whose union is $\mathcal{X}$; such that μ is zero on all measurable subsets of B while ν is zero on all measurable subsets of A.*

Proposition B.1.2 (Jordan decomposition)**.** *The Jordan decomposition states for every measure $\mu \in \mathcal{M}(\mathcal{X})$ the existence of two non-negative Radon measures μ^+, $\mu^- \in \mathcal{M}^+(\mathcal{X})$ which are singular and such that $\mu = \mu^+ - \mu^-$.*

Definition B.1.3 (weak-$*$ topology)**.** *Loosely speaking, weak-$*$ convergence is convergence locally on average. We say that a sequence of Radon measures $(m_n)_{n\geq 0}$* weakly-$*$ converges *towards $m \in \mathcal{M}(\mathcal{X})$ if and only if for all $f \in \mathscr{C}_0(\mathcal{X})$:*

$$\int_{\mathcal{X}} f \, \mathrm{d}m_n \xrightarrow[n\to+\infty]{} \int_{\mathcal{X}} f \, \mathrm{d}m.$$

We note $m_n \overset{*}{\rightharpoonup} m$, it is also called the vague convergence.

Definition B.1.4. *A vector space E is said to be reflexive if the bi-dual E^{**} is identified with E.*

Remark. *Since $\mathscr{C}_0(\mathcal{X})$ is not a reflexive space for its norm* supremum*, the dual of $\mathcal{M}(\mathcal{X})$ for the topology induced by its TV norm is a complicated space, strictly larger [45] than $\mathscr{C}_0(\mathcal{X})$. However, if $\mathcal{M}(\mathcal{X})$ is endowed with the weak-$*$ topology, then $\mathcal{M}(\mathcal{X})$ is a locally convex space whose dual is $\mathscr{C}_0(\mathcal{X})$ [4].*

We also precise the notion of metrisability for the sake of the optimal transport part:

Definition B.1.5. *A topological space $(E, \mathcal{T})$ is said to be* metrisable *if there exists a distance $d : E \times E \to [0, +\infty[$ such that the topology induced by d is $\mathcal{T}$.*

$(\mathcal{M}(\mathcal{X}), *)$ is not a first-countable space, then it is not a metrisable space. To get hold on that:

Lemma B.1.6. *If E is a Banach, the weak-$*$ topology is not metrisable on E^*, except if E is of finite dimension.*

Nonetheless, all bounded subsets of $\mathcal{M}(\mathcal{X})$ are metrisable for the weak-$*$ topology. This property is of upmost importance for the definitions of classic OT metrics such as the Wasserstein distance and for the proof of Γ-convergence of the LASSO to the BLASSO [11,38]. To sum-up all the properties of the different topologies, we give the following Table A2:

Table A2. Algebraic properties of $\mathcal{M}(\mathcal{X})$ for its two main topologies.

Properties	TV Topology	Weak-$*$ Topology
Completeness	Yes	On its bounded subset
Separability	No	Yes
Reflexivity	No	Yes
Metrisable	Yes	On its bounded subset
Polish space (see B.5)	No	On its bounded subset

Appendix B.2. Proof of the Fenchel dual

Proposition B.2.1. *Let be the problem:*

$$\underset{\|\phi^* p\|_{\infty,\mathcal{X}} \leq 1}{\operatorname{argmax}} \ \langle y, p \rangle_{\mathcal{H}} - \frac{\lambda}{2} \|p\|^2_{\mathcal{H}}.$$

It is the dual problem of the BLASSO ($\mathcal{P}_\lambda(y)$).

Proof. We will apply results from [21, Remark 4.2], with a little caveat : the Banach space V should be reflexive, which is clearly not the case here with $V = \mathcal{M}(\mathcal{X})$. However, the reflexive hypothesis is only needed for the sake of existence proof. Since we already proved the solution's existence, this reflexivity hypothesis is not needed in our case. Back to the Remark 4.2, it states, for $\Lambda : V \to Y$ linear, $F : V \to \mathbb{R}$ and $V : Y \to \mathbb{R}$ convex, that the primal problem:

$$\inf_{u \in V} F(u) + G(\Lambda u)$$

has a dual problem which writes down:

$$\sup_{p^* \in Y^*} -F^*(\Lambda^* p^*) - G^*(-p^*). \tag{A1}$$

If u and p^* are respectively solutions of the primal and dual, the extremality conditions are:

$$\begin{cases} \Lambda^* p^* \in \partial F(u) \\ -p^* \in \partial G(\Lambda u). \end{cases}$$

Let use specify in our case $V \overset{\text{def.}}{=} \mathcal{M}(\mathcal{X})$, $Y \overset{\text{def.}}{=} \mathcal{H}$, $F(m) \overset{\text{def.}}{=} |m|(\mathcal{X})$ and $G(p) \overset{\text{def.}}{=} \frac{1}{2}\|y - p\|^2_{\mathcal{H}}$. One can clearly see that the adjoint of G is $G^*(p^*) = \langle y, p^* \rangle_{\mathcal{H}} + \frac{1}{2}\|p^*\|^2_{\mathcal{H}}$ for $p^* \in \mathcal{H}$; in order to determine F^* let $\psi \in V^* \overset{\text{def.}}{=} \mathscr{C}_0(\mathcal{X})$:

$$\begin{aligned} F^*(\psi) &= \sup_{m \in \mathcal{M}(\mathcal{X})} \langle \psi, m \rangle_{\mathscr{C}_0(\mathcal{X}) \times \mathcal{M}(\mathcal{X})} - |m|(\mathcal{X}) \\ &\geq \langle \psi, m \rangle_{\mathscr{C}_0(\mathcal{X}) \times \mathcal{M}(\mathcal{X})} - |m|(\mathcal{X}), \quad \forall m \in \mathcal{M}(\mathcal{X}). \end{aligned}$$

Let $x \in \mathcal{X}$, and $m = \lambda \delta_x$ with $\lambda > 0$. Then one have:

$$\sup_{m \in \mathcal{M}(\mathcal{X})} \langle \psi, m \rangle_{\mathscr{C}_0(\mathcal{X}) \times \mathcal{M}(\mathcal{X})} - |m|(\mathcal{X}) \geq \lambda(\psi(x) - 1).$$

At the limit $\lambda \to \infty$ one yields $F^*(\psi) \geq +\infty$ if $\psi(x) > 1$. A similar result for $\psi(x) < 1$ is obtained with the measure $m = -\lambda \delta_x$. One finally reach $F^*(\psi) = +\infty$ if $\|\psi\|_{\infty,\mathcal{X}} > 1$. Let us assume that $\|\psi\|_{\infty,\mathcal{X}} \leq 1$, first note that $F^*(\psi) = \sup_{m \in \mathcal{M}(\mathcal{X})} \langle \psi, m \rangle_{\mathscr{C}_0(\mathcal{X}) \times \mathcal{M}(\mathcal{X})} - |m|(\mathcal{X}) \geq 0$ (case $m = 0$). Moreover,

$$\begin{aligned} \langle \psi, m \rangle_{\mathscr{C}_0(\mathcal{X}) \times \mathcal{M}(\mathcal{X})} - |m|(\mathcal{X}) &\leq \|\psi\|_{\infty,\mathcal{X}} |m|(\mathcal{X}) - |m|(\mathcal{X}) \\ &\leq |m|(\mathcal{X})(\|\psi\|_{\infty,\mathcal{X}} - 1) \\ &\leq 0 \quad \text{since } \|\psi\|_{\infty,\mathcal{X}} \leq 1. \end{aligned}$$

By introducing the sup on both sides of the last inequality, one finally get $F^*(\psi) = 0$ if $\|\psi\|_{\infty,\mathcal{X}} \leq 1$ thus reaching the condition on the *supremum* norm.

Then from (A1) we yield the dual problem:

$$\sup_{\|\Phi^* p^*\|_{\infty,\mathcal{X}} \leq 1} \langle y, p^* \rangle_{\mathcal{H}} - \frac{1}{2}\|p^*\|^2_{\mathcal{H}}$$

and for m^* and p^* respectively solutions of the primal and dual, its extremality conditions are:

$$\begin{cases} \Phi^* p^* \in \partial |m^*|(\mathcal{X}) \\ -p^* = \Phi m^* - y. \end{cases}$$

This concludes the proof. □

Appendix B.3. Fréchet differential of J'_ν

Let $\sigma, \nu \in \mathcal{M}^+(\mathcal{X})$ and $\varepsilon > 0$. Consider the following:

$$\begin{aligned} J(\nu + \varepsilon\sigma) &= R\left(\int_{\mathcal{X}} \varphi(\theta)(\mathrm{d}\nu + \varepsilon\,\mathrm{d}\sigma)\right) + \lambda|\nu + \varepsilon\sigma|(\mathcal{X}) \\ &= R\left(\int_{\mathcal{X}} \varphi(\theta)\,\mathrm{d}\nu(\theta) + \varepsilon\int_{\mathcal{X}} \varphi(\theta)\,\mathrm{d}\sigma(\theta)\right) + \lambda|\nu|(\mathcal{X}) + \varepsilon|\sigma|(\mathcal{X}). \end{aligned}$$

The TV linearity is obtained thanks to the positivity of ν, σ. Hence, the differential J'_ν at point ν is given by:

$$\begin{aligned} \left.\frac{\mathrm{d}J(\nu + \varepsilon\sigma)}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} &= \left\langle \int_{\mathcal{X}} \varphi(\theta)\,\mathrm{d}\sigma(\theta), \boldsymbol{\nabla} R\left(\int_{\mathcal{X}} \varphi(s)\,\mathrm{d}\sigma(s)\right)\right\rangle_{\mathcal{H}} + \lambda|\sigma|(\mathcal{X}) \\ &= \int_{\mathcal{X}} \langle \varphi(\theta), \boldsymbol{\nabla} R(\varphi(s)\,\mathrm{d}\sigma(s))\rangle_{\mathcal{H}}\,\mathrm{d}\sigma(\theta) + \lambda|\sigma|(\mathcal{X}) \\ &= \langle \langle \varphi, \boldsymbol{\nabla} R\rangle_{\mathcal{H}}, \sigma\rangle_{\mathscr{C}_0(\mathcal{X})\times\mathcal{M}(\mathcal{X})} + \langle \lambda, \sigma\rangle_{\mathscr{C}_0(\mathcal{X})\times\mathcal{M}(\mathcal{X})} \\ &= \left\langle J'_\nu, \sigma\right\rangle_{\mathscr{C}_0(\mathcal{X})\times\mathcal{M}(\mathcal{X})} \end{aligned}$$

Appendix B.4. Gradient of F_N

Let R be the data fitting term *e.g.* the least squares, and h an injective function such as the map $h : r \mapsto r^2$. For $r \in \mathbb{R}_+^N$ and $\theta \in \mathcal{X}^N$ we consider:

$$F_N(r, \theta) = R\left(\frac{1}{N}\sum h(r_i)\varphi(\theta_i)\right) + \frac{1}{N}\sum h(r_i).$$

Its differential is given by:

$$\begin{aligned} \mathrm{d}F_N(x)(\delta x) &= \langle \boldsymbol{\nabla} F_N(x), \delta x\rangle \\ &= \frac{1}{N}\sum \alpha(r)^{-1}\frac{\partial F_N}{\partial r_i}\delta r_i + \sum \beta(r)^{-1}\left\langle \frac{\partial F_N}{\partial r_i}, \delta\theta_i\right\rangle_\theta. \end{aligned}$$

Moreover, we have:

$$\begin{aligned} \frac{\partial F_N}{\partial r_i} &= \frac{1}{N}h'(r_i)\varphi(\theta_i)\frac{\partial R}{\partial r_i}\left(\frac{1}{N}\sum h(r_i)\varphi(\theta_i)\right) + \frac{\lambda}{N}h'(r_i) \\ &= \frac{1}{N}h'(r_i)\left(\varphi(\theta_i)\frac{\partial R}{\partial r_i}\left(\frac{1}{N}\sum h(r_i)\varphi(\theta_i)\right) + \lambda\right) \\ &= h'(r_i)J'_\nu(\theta_i). \end{aligned}$$

Similarly,

$$\begin{aligned}\frac{\partial F_N}{\partial \theta_i} &= \frac{1}{N} h'(r_i)\varphi(\theta_i)\frac{\partial R}{\partial r_i}\left(\frac{1}{N}\sum h(r_i)\varphi(\theta_i)\right) + \frac{\lambda}{N}h'(r_i) \\ &= h(r_i)\nabla J'_\nu(\theta_i).\end{aligned}$$

which yields the gradient in [24]. One can simplify the final expression by introducing the certificate $\eta_\lambda \overset{\text{def.}}{=} J'_\nu/\lambda$.

Appendix B.5. Details on section 4

Definition B.5.1. *A space $(\mathcal{X}, d_{\mathrm{p}})$ is a* Polish space *if it is separable, metrizable, and has a topology – induced by a distance – which makes the space complete.*

$\mathcal{X}$ is a separable Hilbert space then $(\mathcal{X}, d_{\mathrm{p}})$ is a Polish metric space for d_{p} a distance on $\mathbb{R}^d$ restricted to $\mathcal{X}$. We can also introduce:

Definition B.5.2 (Transport plane)**.** *The non-negative measure $\gamma \in \mathcal{M}^+(\mathcal{X}\times\mathcal{X})$ which verifies, for all $A, B \in \mathcal{B}(\mathcal{X})$ where $\mathcal{B}(\mathcal{X})$ is the Borel σ-algebra:*

$$\gamma(A\times\mathcal{X}) = m_1(A), \quad \gamma(B\times\mathcal{X}) = m_2(B)$$

is called the transport plane *between two positive measures m_1 and m_2 of same mass.*

We call $\Gamma(m_1, m_2)$ the set of transport planes between m_1 and m_2 [38]. Metrics of optimal transport such as the Wasserstein distance use at their core these notions, and are defined only on Polish spaces: this is why we work with the measures in X from [11], restriction of $\mathcal{M}(\mathcal{X})$ with the weak-$*$ topology.

Definition B.5.3 (Wasserstein distance)**.** *Let the Polish metric space $(\mathcal{X}, d_{\mathrm{p}})$, and $p \in [1, +\infty)$. For any probability measures μ and ν of $\mathcal{X}$, the Wasserstein distance of order p between μ and ν is defined by:*

$$\mathcal{W}_p(\mu,\nu) = \left(\inf_{\gamma\in\Gamma(\mu,\nu)} \int_{\mathcal{X}} d(x,y)^p \,\mathrm{d}\gamma(x,y)\right)^{1/p}.$$

We also recall the definition of moments:

Definition B.5.4. *If $r \in \mathbb{N}$, we call* moment *of order r of a measure $m \in \mathcal{M}(\mathcal{X})$ the quantity :*

$$\int_{\mathcal{X}} x^r \,\mathrm{d}m(x).$$

We say that m is of r-finite moment if the preceding quantity is finite.

References

1. Candès, E.J.; Fernandez-Granda, C. Towards a Mathematical Theory of Super-resolution. *Communications on Pure and Applied Mathematics* **2013**, *67*, 906–956, [1203.5871]. doi:10.1002/cpa.21455.
2. Bredies, K.; Pikkarainen, H.K. Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations* **2012**, *19*, 190–218. doi:10.1051/cocv/2011205.
3. Castro, Y.D.; Gamboa, F.; Henrion, D.; Lasserre, J.B. Exact solutions to Super Resolution on semi-algebraic domains in higher dimensions. *IEEE Transactions on Information Theory* **2017**, *63*, 621–630, [1502.02436]. doi:10.1109/tit.2016.2619368.
4. Duval, V.; Peyré, G. Exact Support Recovery for Sparse Spikes Deconvolution. *Foundations of Computational Mathematics* **2014**, *15*, 1315–1355, [1306.6909]. doi:10.1007/s10208-014-9228-6.

5. Candes, E.; Romberg, J.; Tao, T. Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information. *IEEE Transactions on Information Theory* **2006**, *52*, 489–509, [math/0409186]. doi:10.1109/tit.2005.862083.
6. de Castro, Y.; Gadat, S.; Marteau, C.; Maugis, C. SuperMix: Sparse Regularization for Mixtures, [1907.10592].
7. Chizat, L.; Bach, F. On the Global Convergence of Gradient Descent for Over-parameterized Models using Optimal Transport. *Advances in Neural Information Processing Systems (NIPS)* **2018**, [1805.09545].
8. Denoyelle, Q. Theoretical and Numerical Analysis of Super-Resolution Without Grid. Theses, Université Paris sciences et lettres, 2018.
9. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **1996**, *58*, 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x.
10. Duval, V.; Peyré, G. Sparse spikes super-resolution on thin grids II: the continuous basis pursuit. *Inverse Problems* **2017**, *33*, 095008. doi:10.1088/1361-6420/aa7fce.
11. Duval, V.; Peyré, G. Sparse regularization on thin grids I: the Lasso. *Inverse Problems* **2017**, *33*, 055008. doi:10.1088/1361-6420/aa5e12.
12. Soubies, E.; Blanc-Féraud, L.; Aubert, G. A Continuous Exact l0 penalty (CEL0) for least squares regularized problem. *SIAM Journal on Imaging Sciences* **2015**, *8*, pp. 1607–1639 (33 pages). doi:10.1137/151003714.
13. Denoyelle, Q.; Duval, V.; Peyré, G.; Soubies, E. The sliding Frank–Wolfe algorithm and its application to super-resolution microscopy. *Inverse Problems* **2019**, *36*, 014001, [1811.06416]. doi:10.1088/1361-6420/ab2a29.
14. Rudin, W. *Real and Complex Analysis*; MCGRAW HILL BOOK CO, 1986. doi:10.2307/2348852.
15. Federer, H. *Geometric measure theory*; Springer: Berlin, 1996.
16. Cohn, D.L. *Measure Theory*; Springer New York, 2013. doi:10.1007/978-1-4614-6956-8.
17. Temam, R. Fonction convexe d'une mesure et applications. *Séminaire Équations aux dérivées partielles (Polytechnique) dit aussi "Séminaire Goulaouic-Schwartz"* **1982-1983**. talk:10.
18. de Castro, Y.; Gamboa, F. Exact reconstruction using Beurling minimal extrapolation. *Journal of Mathematical Analysis and Applications* **2012**, *395*, 336–354, [1103.4951]. doi:10.1016/j.jmaa.2012.05.011.
19. Azais, J.M.; Castro, Y.D.; Gamboa, F. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis* **2015**, *38*, 177–195, [1301.5873]. doi:10.1016/j.acha.2014.03.004.
20. Beurling, A. Sur les intégrales de Fourier absolument convergentes et leur application à une transformation fonctionnelle. *Ninth Scandinavian Mathematical Congress* **1938**, pp. 345–366.
21. Ekeland, I.; Témam, R. *Convex Analysis and Variational Problems*; Society for Industrial and Applied Mathematics, 1999. doi:10.1137/1.9781611971088.
22. Fernandez-Granda, C. Super-Resolution of Point Sources via Convex Programming, 2016, [arXiv:math.OC/1507.07034].
23. Denoyelle, Q.; Duval, V.; Peyré, G. Support Recovery for Sparse Super-Resolution of Positive Measures. *Journal of Fourier Analysis and Applications* **2016**. doi:10.1007/s00041-016-9502-x.
24. Chizat, L. Sparse optimization on measures with over-parameterized gradient descent. *Mathematical Programming* **2021**, [1907.10300]. doi:10.1007/s10107-021-01636-z.
25. Traonmilin, Y.; Aujol, J.F. The basins of attraction of the global minimizers of the non-convex sparse spike estimation problem. *Inverse Problems* **2020**. doi:10.1088/1361-6420/ab5aa3.
26. Traonmilin, Y.; Aujol, J.F.; Leclaire, A. Projected gradient descent for non-convex sparse spike estimation. *IEEE Signal Processing Letters* **2020**.
27. Hilbert, D. Ueber die Darstellung definiter Formen als Summe von Formenquadraten. *Mathematische Annalen* **1888**, *32*, 342–350. doi:10.1007/bf01443605.
28. Dumitrescu, B.A. *Positive Trigonometric Polynomials and Signal Processing Applications*; Springer-Verlag GmbH, 2007.
29. Lasserre, J.B. Global Optimization with Polynomials and the Problem of Moments. *SIAM Journal on Optimization* **2001**, *11*, 796–817. doi:10.1137/s1052623400366802.
30. Levitin, E.; Polyak, B. Constrained minimization methods. *USSR Computational Mathematics and Mathematical Physics* **1966**, *6*, 1–50. doi:10.1016/0041-5553(66)90114-5.
31. Frank, M.; Wolfe, P. An algorithm for quadratic programming. *Naval Research Logistics Quarterly* **1956**, *3*, 95–110. doi:10.1002/nav.3800030109.
32. Harchaoui, Z.; Juditsky, A.; Nemirovski, A. Conditional gradient algorithms for norm-regularized smooth convex optimization. *Mathematical Programming* **2014**, *152*, 75–112, [1302.2325]. doi:10.1007/s10107-014-0778-9.
33. Boyd, N.; Schiebinger, G.; Recht, B. The alternating descent conditional gradient method for sparse inverse problems. 2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). IEEE, 2015, [1507.01562]. doi:10.1109/camsap.2015.7383735.
34. Beck, A.; Teboulle, M. A fast Iterative Shrinkage-Thresholding Algorithm with application to wavelet-based image deblurring. 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2009. doi:10.1109/icassp.2009.4959678.
35. Byrd, R.H.; Lu, P.; Nocedal, J.; Zhu, C. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing* **1995**, *16*, 1190–1208. doi:10.1137/0916069.
36. Ambrosio, L.; Gigli, N.; Savare, G. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*; Springer-Verlag GmbH, 2008. doi:10.1007/978-3-7643-8722-8.

37. Santambrogio, F. Euclidean, Metric, and Wasserstein Gradient Flows: an overview, 2016, [arXiv:math.AP/1609.03890].
38. Peyré, G.; Cuturi, M. *Computational Optimal Transport*; now Publishers Inc, 2019; [1803.00567]. doi:10.1561/9781680835519.
39. Lellmann, J.; Lorenz, D.A.; Schönlieb, C.; Valkonen, T. Imaging with Kantorovich-Rubinstein discrepancy. *SIAM Journal on Imaging Sciences* **2014**, *7*, 2833–2859, [1407.0221]. doi:10.1137/140975528.
40. Denoyelle, Q.; an Pham, T.; del Aguila Pla, P.; Sage, D.; Unser, M. Optimal-transport-based metric for SMLM. 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE, 2021, [2010.13423]. doi:10.1109/isbi48211.2021.9433874.
41. Dossal, C.; Duval, V.; Poon, C. Sampling the Fourier transform along radial lines. *SIAM Journal on Numerical Analysis* **2017**, *55*, 2540–2564, [arXiv:math.NA/1612.06752]. doi:10.1137/16m1108807.
42. Sage, D.; Kirshner, H.; Pengo, T.; Stuurman, N.; Min, J.; Manley, S.; Unser, M. Quantitative evaluation of software packages for single-molecule localization microscopy. *Nature Methods* **2015**, *12*, 717–724. doi:10.1038/nmeth.3442.
43. Holden, S.J.; Uphoff, S.; Kapanidis, A.N. DAOSTORM: an algorithm for high- density super-resolution microscopy. *Nature Methods* **2011**, *8*, 279–280. doi:10.1038/nmeth0411-279.
44. Nehme, E.; Weiss, L.E.; Michaeli, T.; Shechtman, Y. Deep-STORM: super-resolution single-molecule microscopy by deep learning. *Optica* **2018**, *5*, 458–464. doi:10.1364/OPTICA.5.000458.
45. Javanshiri, H.; Nasr-Isfahani, R. The strict topology for the space of Radon measures on a locally compact Hausdorff space. *Topology and its Applications* **2013**, *160*, 887–895. doi:https://doi.org/10.1016/j.topol.2013.02.007.