Off-the-grid regularisation for Poisson inverse problems

Marta Lazzaretti, Claudio Estatico, Alejandro Melero Carrillo, Luca Calatroni

Off-the-grid regularisation for Poisson inverse problems

Marta Lazzaretti[1,2*], Claudio Estatico[1],
Alejandro Melero Carrillo[3], Luca Calatroni[2]

[1]Dipartimento di Matematica, Università di Genova, Via Dodecaneso 35, Genova, 16145, Italy.
[2]I3S Laboratory, CNRS-UniCA-Inria, 2000 Route des Lucioles, Sophia Antipolis, 06903, France.
[3]Laboratory of GTPases and Neurosignalling, Achucarro Basque Center for Neuroscience, Parque Científico UPV/EHU, Edif. Sede, Planta 3, Leioa, E-48940, Spain.

*Corresponding author(s). E-mail(s): lazzaretti@dima.unige.it;
Contributing authors: estatico@dima.unige.it;
alejandro.melero@achucarro.org; calatroni@i3s.unice.fr;

Abstract

Off-the-grid regularisation has been extensively employed over the last decade in the context of ill-posed inverse problems formulated in the continuous setting of the space of Radon measures $\boldsymbol{\mathcal{M}(\Omega)}$. These approaches enjoy convexity and counteract the discretisation biases as well the numerical instabilities typical of their discrete counterparts. In the framework of sparse reconstruction of discrete point measures (sum of weighted Diracs), a Total Variation regularisation norm in $\boldsymbol{\mathcal{M}(\Omega)}$ is typically combined with an $\boldsymbol{L^2}$ data term modelling additive Gaussian noise. To asses the framework of off-the-grid regularisation in the presence of signal-dependent Poisson noise, we consider in this work a variational model coupling the Total Variation regularisation with a Kullback-Leibler data term under a non-negativity constraint. Analytically, we study the optimality conditions of the composite functional and analyse its dual problem. Then, we consider an homotopy strategy to select an optimal regularisation parameter and use it within a Sliding Frank-Wolfe algorithm. Several numerical experiments on both 1D/2D simulated and real 3D fluorescent microscopy data are reported.

Keywords: Off-the-grid sparse regularisation, Poisson noise, Sliding Frank-Wolfe, fluorescence microscopy imaging.

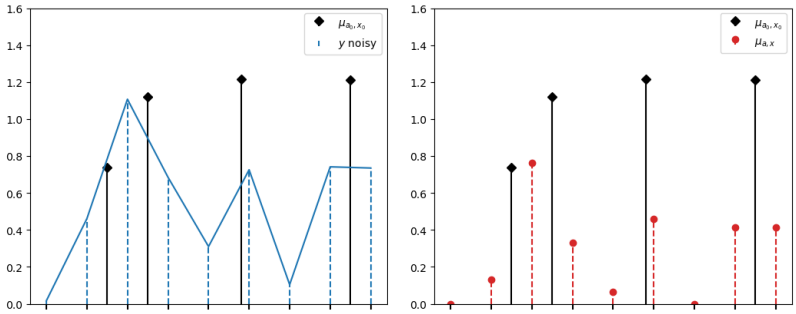
1 Introduction

Discrete (or *on-the-grid*) sparse optimisation approaches are nowadays established techniques in the field of mathematical signal and image inverse problems. In the context of linear ill-posed inverse problems, for instance, they aim at retrieving a sparse approximation of a quantity of interest x (e.g., an image) from blurred, noisy and potentially incomplete acquisitions $y \in \mathbb{R}^M$ on a regular grid of size $N \geq M$, see Fig. 1a. The discretisation parameter N determines the localisation precision, as only the centres of the discretisation intervals are candidates for reconstruction. As such, in case of rough discretisations, very approximate reconstructions may be computed (see, e.g., Fig. 1b where small values $M = N$ are used). To obtain higher precision, one typically chooses a grid-size parameter $N > M$ for reconstruction, as in Fig. 1c. Choosing a large value for N, however, may cause instabilities in the reconstructions [1] due to the higher numerical complexity.

Off-the-grid approaches aim at overcoming such difficulties. They can be thought indeed as the natural framework to deal with the case $N \to +\infty$ of *on-the-grid* formulations [2, 3]. In such framework, the spatial domain $\Omega \subseteq \mathbb{R}^d$ is not discretised by a regular grid, but, rather, the quantity of interest is modelled as an element of a suitable functional space defined on Ω. A natural framework for spike reconstruction problems, is, for instance, the space of Radon measures $\mu \in \mathcal{M}(\Omega)$ where the quantity of interest can be modelled as the measure $\mu_{a,x} = \sum_{i=1}^{n} a_i \delta_{x_i}$ with n being the number of spikes and where the question is therefore how to estimate the number of spikes n, and then how to retrieve both continuous positions $x_i \in \Omega$ and amplitudes $a_i \geq 0$ for $i = 1, \ldots, n$.

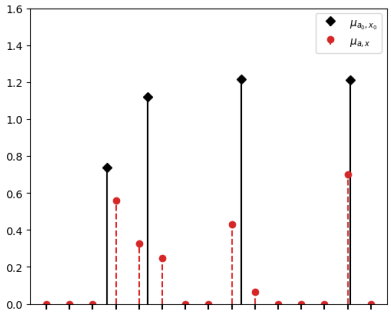
Inverse problems in the space of measures and off-the-grid optimisation methods have been first proposed in [4–7] and since then they have been a topic of intense research activity for the mathematical community, both from an analytical and numerical viewpoint, see, e.g., [8–14]. Off-the-grid methods have been proved to be particularly useful in applications where fine-scale details need to be retrieved from noisy acquisitions, such as spike detection in astronomy and microscopy [12], as well as parameter estimation in in spectroscopy [15] and density mixture estimation [16]. Standard approaches in this setting usually combine off-the-grid regularisation with an additive (Gaussian) noise modelling on the underlying signal, which is in general easier to work with, from both an analytical and a computational point of view. Such modelling corresponds to the well-studied variational formulation of the Beurling LASSO (BLASSO) model, [5, 7].

In this work, we consider an off-the-grid modelling in $\mathcal{M}(\Omega)$ under the hypothesis of signal-dependent Poisson noise in the data. This choice is motivated by some particular biological applications of interest, such as fluorescence microscopy, where, due to the photon emission nature of the light, Poisson noise is better suited than the Gaussian one to describe the process of photon counts on acquired images [17]. While in a discrete setting, a precise modelling of noise statistics if often not necessary due to the inevitable biases introduced by the choice of the regularisation employed, a natural question is whether whenever a refined off-the-grid regularisation is used, a precise noise model could indeed be relevant. As a graphical visualisation, we report in Figure

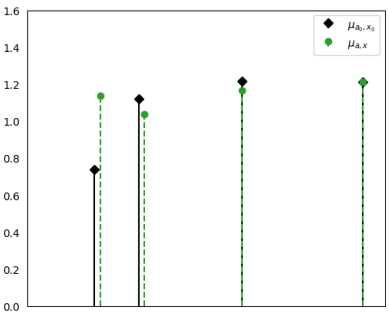


(a) Acquisition $y \in \mathbb{R}^M$

(b) Discrete reconstruction $x \in \mathbb{R}^M$



(c) Discrete reconstruction $x \in \mathbb{R}^N$



(d) Off-the-grid rec. $\mu_{a,x} \in \mathcal{M}(\Omega)$

Figure 1: Comparison between discrete (on-the-grid) and off-the-grid sparse reconstructions with a Poisson data term. In black: the ground-truth spikes to retrieve. In Fig.1a: the acquired blurred and noisy signal $y \in \mathbb{R}^M$ lying on a low-resolution grid of size M. In Fig.1b: in red, a discrete reconstruction with support on a grid with M pixels. In Fig.1c: in red, discrete reconstruction with support on a grid with $N > M$ pixels. In Fig.1d: in green, off-the-grid reconstruction.

1 a visual comparison between reconstructions obtained with on-the-grid approaches and the off-the-grid approach proposed in this work to solve a spike-deconvolution problem under the choice of the Poisson data term considered in this work.

1.1 Inverse problems in the space of Radon measures $\mathcal{M}(\Omega)$

In this section we recall some important definitions and properties of the space of Radon measures $\mathcal{M}(\Omega)$ following [3, 12, 18]. For more details on Radon measures see [19, 20].

Let $\Omega \subseteq \mathbb{R}^d$, with $d \in \mathbb{N}$, $d \geq 1$, be a compact subset of $\mathbb{R}^d$ with non-empty interior. We denote by $\mathcal{C}_0(\Omega, \mathbb{R})$ the space of real continuous functions on Ω vanishing

at infinity, namely all the continuous maps $\psi : \Omega \to \mathbb{R}$ such that:

$$\forall \varepsilon > 0, \quad \exists K \subset \Omega,\ K \text{ compact, s.t.} \qquad \sup_{x \in \Omega \setminus K} |\psi(x)| \leq \varepsilon.$$

The space of Radon measures is defined through duality.

Definition 1. *The Banach space $\mathcal{M}(\Omega)$ of real signed Radon measures on Ω is the topological dual of $\mathcal{C}_0(\Omega, \mathbb{R})$ endowed with the supremum norm $\|\cdot\|_{\infty,\Omega}$, defined by $\|\psi\|_{\infty,\Omega} := \sup_{x \in \Omega} |\psi(x)|$.*

This definition allows to characterise any measure $\mu \in \mathcal{M}(\Omega)$ as a continuous linear form evaluated on functions $\psi \in \mathcal{C}_0(\Omega, \mathbb{R})$ in terms of the duality pairing, so that

$$\forall \mu \in \mathcal{M}(\Omega) \qquad \langle \psi, \mu \rangle_{\mathcal{C}_0(\Omega,\mathbb{R}) \times \mathcal{M}(\Omega)} = \int_\Omega \psi \mathrm{d}\mu. \tag{1}$$

A Radon measure $\mu \in \mathcal{M}(\Omega)$ is a positive measure if $\langle \psi, \mu \rangle_{\mathcal{C}_0(\Omega,\mathbb{R}) \times \mathcal{M}(\Omega)}$ is positive or null for any non-negative function $\psi \in \mathcal{C}_0(\Omega, \mathbb{R})$. This specifies the meaning of the term *signed* in the above definition, as the quantity $\langle \psi, \mu \rangle_{\mathcal{C}_0(\Omega,\mathbb{R}) \times \mathcal{M}(\Omega)}$ can be also negative. Note that the space of positive (or non-negative) Radon measures $\mathcal{M}^+(\Omega)$ cannot be defined as the topological dual of $\mathcal{C}_0(\Omega, \mathbb{R}^+)$, since the latter is not a vector space, since trivially if $\psi \in \mathcal{C}_0(\Omega, \mathbb{R}^+)$ then $-\psi \notin \mathcal{C}_0(\Omega, \mathbb{R}^+)$.

It can be shown that $\mathcal{M}(\Omega)$ is a non-reflexive Banach space endowed with the Total Variation (TV) norm, here defined $\forall \mu \in \mathcal{M}(\Omega)$ as:

$$|\mu|(\Omega) = \sup \Big(\int_\Omega \psi \mathrm{d}\mu \ \Big| \ \ \psi \in \mathcal{C}_0(\Omega, \mathbb{R}),\ \|\psi\|_{\infty,\Omega} \leq 1 \Big). \tag{2}$$

The TV norm is convex and non-differentiable, hence it is possible to consider its subdifferential [5, 8]:

$$\partial |\mu|(\Omega) = \left\{ \psi \in \mathcal{C}_0(\Omega, \mathbb{R}) |\ \|\psi\|_{\infty,\Omega} \leq 1 \text{ and } \int_\Omega \psi \mathrm{d}\mu = |\mu|(\Omega) \right\}.$$

Note that for sparse discrete measures, i.e. weighted sums of Diracs:

$$\mu_{a,x} = \sum_{i=1}^N a_i \delta_{x_i} \text{ with } N \in \mathbb{N},\ a = (a_1, \ldots, a_N) \in \mathbb{R}^N,\ x = (x_1, \ldots, x_N) \in \Omega^N, \tag{3}$$

the TV norm coincides with the L^1 norm of the amplitudes vector a, that is $|\mu_{a,x}|(\Omega) = \|a\|_1$. This explains why the TV norm is considered in this setting as a generalisation of the L^1 norm. Moreover, in this special case $\langle \psi, \mu \rangle = \sum_{i=1}^N a_i \psi(x_i)$ and the subdifferential has the following expression involving the sign function

$$\partial |\mu_{a,x}|(\Omega) = \{ \psi \in \mathcal{C}_0(\Omega, \mathbb{R}) |\ \|\psi\|_{\infty,\Omega} \leq 1,\ \forall i = 1, \ldots, N\ \psi(x_i) = \mathrm{sign}(a_i) \}. \tag{4}$$

We focus now on the formulation of linear inverse problems in $\mathcal{M}(\Omega)$. As an acquisition space we will consider a Hilbert space $\mathcal{H}$. Let $\mu \in \mathcal{M}(\Omega)$ be the unknown source measure. We consider an acquisition $\bar{y} \in \mathcal{H}$ being the result of the action of the forward operator $\Phi : \mathcal{M}(\Omega) \to \mathcal{H}$ evaluated on μ, with a fixed measurement kernel $\varphi : \Omega \to \mathcal{H}$, that is:

$$\bar{y} := \Phi\mu = \int_\Omega \varphi(x) \mathrm{d}\mu(x). \tag{5}$$

Note that this integral should not be confused with the concept of duality pairing (1), which is a scalar function defined as the integral over Ω of a continuous real function with respect to a measure $\mu \in \mathcal{M}(\Omega)$. In (5), the integral is then a Böchner integral [20] for vector-valued functions $\varphi(x) \in \mathcal{H}$, that is $\varphi(x)$ is an element of $\mathcal{H}$, and it is well-defined if φ is continuous and bounded [12, 21].

The choice of the kernel φ and of the acquisition space $\mathcal{H}$ depends on the specific physical acquisition process considered. In the following, we consider a convolution kernel, which is of practical interest in fluorescence microscopy (see, e.g., [12, 22]). In this setting, a natural choice is therefore $\mathcal{H} = L^2(\Omega)$, with the convolution kernel $\varphi : \Omega \to L^2(\Omega)$ defined in terms of a Point Spread Function (PSF) $\tilde{\varphi} : \Omega \to L^2(\Omega)$ acting as:

$$\varphi(x) := (s \mapsto \tilde{\varphi}(s - x)) \in L^2(\Omega), \ \forall x \in \mathbb{R}. \tag{6}$$

Note that depending on the microscopy technique used one can have different PSFs. For instance, the Gaussian PSF, centred in $c \in \Omega$ with radius $\sigma > 0$, is defined by

$$s \mapsto \tilde{\varphi}(s - c) := (2\pi\sigma^2)^{-d/2} e^{-\|s-c\|_2^2/2\sigma^2}.$$

For other possible choices of measurement kernels, we refer the reader to [12].

We observe that the action of the forward operator Φ on finite linear combination of Diracs (3) can be explicited as follows:

$$\Phi\mu_{a,x} = \int_\Omega \varphi(x) \mathrm{d}\mu(x) = \sum_{i=1}^N a_i \varphi(x_i).$$

For simplicity, the following notation will therefore be used in the sequel $\Phi_x(a) = \sum_{i=1}^N a_i \varphi(x_i)$ to denote $\Phi\mu_{a,x}$.

In fluorescence microscopy imaging applications, the objects of interest are usually images of molecules, i.e. single point-sources emitting fluorescent light. The unknown is therefore well-described by non-negative discrete measures of the form (3). Other possible objects of interest in this field are microtubules, 1-dimensional curve structures [23, 24], and cells (2-dimensional), that can be modelled as piece-wise constant functions [25–27]. For simplicity, we will focus in the following only on inverse problems aiming at recovering (0-dimensional) measures in the form (3). The extension to more general regularisation models is left for future work.

1.2 The BLASSO problem: formulation, duality and optimality conditions

The standard sparse spike deconvolution problem, consists in recovering a (small) finite linear combination of Diracs $\mu_{a,x} = \sum_{i=1}^{N} a_i \delta_{x_i}$ from a blurred and noisy acquisition

$$y = \Phi \mu_{a,x} + \omega, \tag{7}$$

where $\Phi : \Omega \to L^2(\Omega)$ is a convolution operator, as defined in (6), and ω is an additive noise component, typically describing white Gaussian noise. The variational formulation of the problem of retrieving an estimate μ from y is

$$\operatorname*{argmin}_{\mu \in \mathcal{M}(\Omega)} T_\lambda(\mu) := \frac{1}{2}\|\Phi\mu - y\|^2 + \lambda |\mu|(\Omega), \quad \lambda > 0 \qquad (L^2 - |\cdot|)$$

which is the Beurling-LASSO (BLASSO) functional, named after the work of the mathematician Beurling [28]. It is considered the generalisation of the discrete LASSO variational problem

$$\operatorname*{argmin}_{x \in \mathbb{R}^N} \frac{1}{2}\|\tilde{\Phi}x - y\|^2 + \lambda \|x\|_1,$$

where $\tilde{\Phi} \in \mathbb{R}^{M \times N}$ is a discretisation of $\Phi : \mathcal{M}(\Omega) \to \mathcal{H}$ and $y \in \mathbb{R}^M$. In [5], the functonal $T_\lambda : \mathcal{M}(\Omega) \to \mathbb{R}$ is proved to be proper, convex and coercive, which proves the existence of solutions of $(L^2 - |\cdot|)$ and uniqueness under injectivity on Φ.

By optimality, μ_λ minimises T_λ if and only if $0 \in \partial T_\lambda(\mu_\lambda)$, i.e.

$$0 \in \partial T_\lambda(\mu_\lambda) = \Phi^*(\Phi\mu_\lambda - y) + \lambda \partial |\mu_\lambda|(\Omega) \iff \frac{1}{\lambda}\Phi^*(y - \Phi\mu_\lambda) \in \partial |\mu_\lambda|(\Omega),$$

which can be written as

$$\eta \in \partial |\mu_\lambda|(\Omega) \qquad \text{with} \qquad \eta := \frac{1}{\lambda}\Phi^*(y - \Phi\mu_\lambda), \tag{8}$$

where η is the so-called dual certificate of $(L^2 - |\cdot|)$. The dual certificate formally generalises the concept of Euler equation, playing a crucial role in the characterisation of optimality conditions for $(L^2 - |\cdot|)$ and in devising optimisation algorithms in this setting, as better specified in the following. If μ is a finite linear combination of Dirac masses (3), recalling (4), we get that optimality conditions (8) become simply

$$\eta(x_i) = \operatorname{sign}(a_i) \quad \wedge \quad \|\eta\|_\infty \leq 1. \tag{9}$$

Note that the optimality conditions (9) can be also derived by studying the dual problem of $(L^2 - |\cdot|)$:

$$\operatorname*{argmax}_{\|\Phi^* p^*\|_{\infty,\Omega} \leq 1} \langle y, p^* \rangle - \frac{\lambda}{2}\|p^*\|^2, \tag{10}$$

which provides further insights to the meaning of the dual certificate η. Given $\mu_\lambda \in \mathcal{M}(\Omega)$ solution of to $(L^2 - |\cdot|)$ and denoting by $p^*_\lambda \in L^2(\Omega)$ the solution of the dual problem (10), the following extremality conditions hold true

$$\begin{cases} \Phi^* p^*_\lambda \in \partial|\mu_\lambda|(\Omega), \\ -p^*_\lambda = \frac{1}{\lambda}\left(\Phi\mu_\lambda - y\right) \end{cases}. \tag{11}$$

From (11), we retrieve in fact the optimality conditions (9) expressed in terms of the dual certificate (8). In addition, when the dual certificate satisfies (9), by duality we have $\eta = \Phi^* p^*_\lambda$ with p^*_λ being a solution of (10). Optimality conditions thus fully characterise the solution(s) of the BLASSO problem $(L^2 - |\cdot|)$. They are indeed crucial in devising algorithms for its minimisation and, in particular, in the definition of good stopping criterions.

2 Off-the-grid Poisson inverse problems

We now present the signal-dependent Poisson modelling studied in this work, which differs from (7) as it is not additive.

Let $\Phi : \mathcal{M}(\Omega) \to L^2(\Omega)$ be the forward operator in (5). In the following, Φ is assumed to be a positive definite operator, i.e.

$$\mu \in \mathcal{M}^+(\Omega) \text{ positive measure } \Rightarrow \Phi\mu(x) \geq 0 \quad \forall x \in \Omega. \tag{12}$$

Observe that whenever the measurement kernel φ (5) is non-negative, this always holds. This is the case, for instance, in standard image deblurring problems. For simplicity, we introduce the notation $L^2(\Omega)^+ = \{f \in L^2(\Omega) \text{ such that } f(x) > 0 \text{ a .e. } x \in \Omega\}$.

The Poisson sparse spike deconvolution problem in an off-the-grid setting consists in finding a positive discrete measure $\mu \in \mathcal{M}^+(\Omega)$ such that

$$y = \mathcal{P}(\Phi\mu + b), \tag{13}$$

where $b \in L^2(\Omega)^+$ is a strictly positive background term, and $y \in L^2(\Omega)$ is a realisation of a Poisson distributed random process $\mathsf{Y} \sim \mathcal{P}(\Phi\mu + b)$. Similarly as in the standard discrete setting, we consider the Kullback-Leibler divergence as a suitable data term [29–33].

Definition 2. *The Kullback-Leibler divergence* $\mathcal{D}_{KL} : L^2(\Omega)^+ \times L^2(\Omega)^+ \longrightarrow \mathbb{R}$ *is defined by*

$$\mathcal{D}_{KL}(s,t) := \int_\Omega s(x) - t(x) + t(x)\Big(\log(t(x)) - \log(s(x))\Big)\,\mathrm{d}x. \tag{14}$$

To consider $\mathcal{D}_{KL}(\Phi\mu + b, y)$ as data fidelity term, two issues have to be considered:

- if μ is not a positive measure, then $\Phi\mu + b$ might not be positive;

- the noisy acquisition y might still vanish in a non negligible region of the domain Ω. This is due to the fact that a Poisson random variable with mean α vanishes with positive probability equal to $e^{-\alpha}$. For (14) to be well defined, it is thus required that $y > 0$ almost everywhere.

To solve the first issue, we introduce the function $\tilde{\mathcal{D}}_{KL} : L^2(\Omega) \times L^2(\Omega)^+ \longrightarrow \mathbb{R} \cup \{+\infty\}$ which extends (14) as

$$\tilde{\mathcal{D}}_{KL}(s,t) = \begin{cases} \mathcal{D}_{KL}(s,t) & s \in L^2(\Omega)^+ \\ +\infty & s \notin L^2(\Omega)^+ \end{cases}. \tag{15}$$

Moreover, we just restrict our study to positive acquisitions y, thus requiring

$$y \in L^2(\Omega)^+. \tag{16}$$

Under assumptions (12) and (16) and using (15), the quantity $\tilde{\mathcal{D}}_{KL}(\Phi\mu + b, y)$ is therefore well-defined for all $\mu \in \mathcal{M}(\Omega)$.

We can thus introduce the following variational problem for solving (13)

$$\boxed{\operatorname*{argmin}_{\mu \in \mathcal{M}(\Omega)} \tilde{\mathcal{D}}_{KL}(\Phi\mu + b, y) + \lambda |\mu|(\Omega) + \mathbb{1}_{\{\mathcal{M}^+(\Omega)\}}(\mu), \quad \lambda > 0.} \qquad (\tilde{\mathcal{D}}_{KL} - |\cdot|)$$

The Poisson fidelity (15) is there coupled with the TV norm, together with the indicator function of the positive measures $\mathcal{M}^+(\Omega)$ defined by

$$\mathbb{1}_{\{\mathcal{M}^+(\Omega)\}}(\mu) = \begin{cases} 0 & \mu \geq 0 \\ +\infty & \text{otherwise} \end{cases},$$

to ensure μ to be non-negative.

2.1 Dual problem and optimality conditions

We analyse in the following the dual problem of $(\tilde{\mathcal{D}}_{KL} - |\cdot|)$ and provide an analytical expression of the convex conjugate of the involved functions. We refer to Appendix A for the definition of convex conjugate [34] and classical concepts about duality [35].

The study of the dual problem of the problem $(\tilde{\mathcal{D}}_{KL} - |\cdot|)$ requires the computation of the convex conjugate $F^* : \mathcal{C}_0(\Omega, \mathbb{R}) \to \mathbb{R} \cup \{+\infty\}$ of the penalty term

$$F : \mathcal{M}(\Omega) \to \mathbb{R} \cup \{+\infty\}, \quad F(\cdot) := |\cdot|(\Omega) + \mathbb{1}_{\{\mathcal{M}^+(\Omega)\}}(\cdot), \tag{17}$$

and of the convex conjugate $G^* : L^2(\Omega) \to \mathbb{R} \cup \{+\infty\}$ of the fidelity term

$$G : L^2(\Omega) \to \mathbb{R} \cup \{+\infty\}, \quad G(\cdot) := \frac{1}{\lambda} \tilde{\mathcal{D}}_{KL}(\cdot, y). \tag{18}$$

To compute the convex conjugate of the Kullback-Leibler functional (18), we start considering the one-dimensional Kullback-Leibler function, defined by

$$g_t(s) = \frac{1}{\lambda}\big(s - t + t\log(t) - t\log(s)\big), \qquad s, t > 0 \text{ and } \lambda > 0.$$

Applying the definition of convex conjugate (A1) to g_t yields

$$\begin{aligned} g_t^*(s^*) &= \sup_{s>0}\ ss^* - g_t(s) = \sup_{s>0}\ ss^* - \frac{1}{\lambda}\Big(s - t + t\log(t) - t\log(s)\Big) = \\ &= \sup_{s>0}\ \underbrace{s\Big(s^* - \frac{1}{\lambda}\Big) + \frac{t}{\lambda}\log(s) + \frac{t}{\lambda} - \frac{t}{\lambda}\log(t)}_{h(s)}. \end{aligned}$$

We have two cases:

(i) If $s^* \geq \frac{1}{\lambda}$, then $\lim_{s\to+\infty} h(s) = +\infty$ implies $\sup_{s>0} h(s) = +\infty \Rightarrow g_t\big(s^*\big) = +\infty$ for all $t > 0$.

(ii) If $0 < s^* < \frac{1}{\lambda}$, then $\lim_{s\to\pm\infty} h(s) = -\infty$. Thus, being h a convex and differentiable function its supremum is attained at $\hat{s}$ such that $h'(\hat{s}) = 0$, which can be computed by

$$\begin{aligned} h'(\hat{s}) = s^* - \frac{1}{\lambda} + \frac{t}{\lambda\hat{s}} = \frac{\lambda\hat{s}s^* - \hat{s} + t}{\lambda\hat{s}} = 0 \\ \iff \lambda\hat{s}s^* - \hat{s} + t = 0 \iff \hat{s} = \frac{t}{1 - \lambda s^*}. \end{aligned}$$

Thus,

$$g_t^*(s^*) = h\Big(\frac{t}{1-\lambda s^*}\Big) = -\frac{t}{\lambda}\log(1 - \lambda s^*).$$

Observe that $g_t^*(s^*)$ is well defined since $s^* < \frac{1}{\lambda}$.

Hence, the convex conjugate g_t^* of g_t is

$$g_t^*(s^*) = \begin{cases} +\infty & s^* \geq \frac{1}{\lambda} \\ -\frac{t}{\lambda}\log(1 - \lambda s^*) & s^* < \frac{1}{\lambda} \end{cases}. \tag{19}$$

Since $\tilde{\mathcal{D}}_{KL}(\cdot, t)$ is defined also for non-positive functions, its one-dimensional counterpart is given by $\tilde{g}_t : \mathbb{R} \longrightarrow \mathbb{R} \cup \{+\infty\}$ where

$$\tilde{g}_t(s) = \begin{cases} g_t(s) & s > 0 \\ +\infty & s \leq 0 \end{cases},$$

so that its convex conjugate coincides with (19). We thus have the following lemma.

Lemma 1. *Let $G : L^2(\Omega) \to \mathbb{R} \cup \{+\infty\}$ be the function defined by*

$$G(\cdot) := \frac{1}{\lambda}\tilde{\mathcal{D}}_{KL}(\cdot, y),$$

where $\tilde{\mathcal{D}}_{KL}$ is given by (15). The convex conjugate of G is given by $G^* : L^2(\Omega) \to \mathbb{R} \cup \{+\infty\}$ defined by

$$G^*(p^*) = \begin{cases} +\infty & p^* \geq \frac{1}{\lambda} \\ -\frac{y}{\lambda}\log(1-\lambda p^*) & p^* < \frac{1}{\lambda} \end{cases} = -\frac{y}{\lambda}\log(1-\lambda p^*)\mathbb{1}_{\{z<1/\lambda\}}(p^*)\,, \tag{20}$$

where $-\frac{y}{\lambda}\log(1-\lambda p^*) = \langle -\frac{y}{\lambda}, \log(1-\lambda p^*)\rangle \in \mathbb{R}$.

We compute now the convex conjugate of F defined in (17). To simplify the notation, we denote in the following by $A, B : \mathcal{M}(\Omega) \to \mathbb{R} \cup \{+\infty\}$ the quantities

$$A(\cdot) = |\cdot|(\Omega), \qquad B(\cdot) = \mathbb{1}_{\{\mathcal{M}^+(\Omega)\}}(\cdot). \tag{21}$$

The convex conjugate of F can be obtained as infimal convolution of the convex conjugate A^* and B^* [36, 37], that is

$$F^*(\psi) = \min_{\psi_1+\psi_2=\psi} A^*(\psi_1) + B^*(\psi_2), \tag{22}$$

see Appendix A for more details and for the the computation of A^*, which is a well-known result. We report here the computation of the convex conjugate of $B(\cdot)$.

Lemma 2. *The convex conjugate of the indicator function of positive measures B of (21) is the function $B^* : \mathcal{C}_0(\Omega, \mathbb{R}) \to \mathbb{R} \cup \{+\infty\}$ given by*

$$B^*(\psi) = \begin{cases} 0 & \psi(x) \leq 0 \quad \forall x \in \Omega \\ +\infty & \exists x \in \Omega \text{ s.t. } \psi(x) > 0 \end{cases}.$$

Proof. By Definition 3, for any $\psi \in \mathcal{C}_0(\Omega, \mathbb{R})$ we get

$$\begin{aligned} B^*(\psi) &= \sup_{\mu \in \mathcal{M}(\Omega)} \langle \psi, \mu \rangle_{\mathcal{C}_0(\Omega,\mathbb{R})\times\mathcal{M}(\Omega)} - B(\mu) = \sup_{\mu \in \mathcal{M}^+(\Omega)} \langle \psi, \mu \rangle_{\mathcal{C}_0(\Omega,\mathbb{R})\times\mathcal{M}(\Omega)} \\ &\geq \langle \psi, \mu \rangle_{\mathcal{C}_0(\Omega,\mathbb{R})\times\mathcal{M}(\Omega)} \quad \forall \mu \in \mathcal{M}^+(\Omega). \end{aligned}$$

If there exists $\bar{x} \in \Omega$ such that $\psi(\bar{x}) > 0$, consider the positive measure $\bar{\mu} = \alpha\delta_{\bar{x}}$ with $\alpha > 0$. Then, $B^*(\psi) \geq \langle \psi, \bar{\mu} \rangle = \alpha\psi(\bar{x})$ and taking the limit for $\alpha \to \infty$ of the latter inequality yields $B^*(\psi) \geq +\infty$ if $\psi(x) > 1$. On the other hand, if $\psi(x) \leq 0\ \forall x \in \Omega$, then $\langle \psi, \mu \rangle = \int_\Omega \psi \mathrm{d}\mu \leq 0\ \forall \mu \in \mathcal{M}^+(\Omega)$. Moreover, $\langle \psi, 0 \rangle = 0$. Thus, $B^*(\psi) = 0$ if $\psi(x) \leq 0\ \forall x \in \Omega$, which concludes the proof. □

We thus get an expression for F^* in the following lemma.

Lemma 3. *Let $F : \mathcal{M}(\Omega) \to \mathbb{R} \cup \{+\infty\}$ be the functional defined in (17). For $\psi \in \mathcal{C}_0(\Omega, \mathbb{R})$, its convex conjugate $F^* : \mathcal{C}_0(\Omega, \mathbb{R}) \to \mathbb{R} \cup \{+\infty\}$ is defined by*

$$F^*(\psi) = \begin{cases} 0 & \text{if } \psi(x) \leq 1 \quad \forall x \in \Omega \\ +\infty & \text{otherwise} \end{cases}. \tag{23}$$

Proof. By (22), we have

$$\begin{aligned} F^*(\psi) &= \min_{\psi_1+\psi_2=\psi} A^*(\psi_1) + B^*(\psi_2) \\ &= \min_{\psi_1+\psi_2=\psi} \begin{cases} 0 & \|\psi_1\|_{\infty,\Omega} \leq 1 \\ +\infty & \|\psi_1\|_{\infty,\Omega} > 1 \end{cases} + \begin{cases} 0 & \psi_2(x) \leq 0 \ \forall x \in \Omega \\ +\infty & \exists x \in \Omega \text{ s.t. } \psi_2(x) > 0 \end{cases}. \end{aligned}$$

Notice that $F^*(\psi) \geq 0 \ \forall \psi \in \mathcal{C}_0(\Omega, \mathbb{R})$ and we have either $F^*(\psi) = 0$ or $F^*(\psi) = +\infty$. Thus, $F^*(\psi) = 0$ if and only if there exist $\psi_1, \psi_2 \in \mathcal{C}_0(\Omega, \mathbb{R})$ such that $\psi = \psi_1 + \psi_2$ with $A^*(\psi_1) = 0$ and $B^*(\psi_2) = 0$. On the contrary, $F^*(\psi) = +\infty$ if for all $\psi_1, \psi_2 \in \mathcal{C}_0(\Omega, \mathbb{R})$ such that $\psi = \psi_1 + \psi_2$ it holds $A^*(\psi_1) = +\infty$ or $B^*(\psi_2) = +\infty$. Consider now the three following cases:

i) If $\|\psi\|_{\infty,\Omega} \leq 1$, consider $\psi_1 = \psi$. This implies $A^*(\psi_1) = 0$, and $\psi_2 = 0$, which implies $B^*(\psi_2) = 0$. Thus $F^*(\psi) = 0$.

ii) If $\|\psi\|_{\infty,\Omega} > 1$ and there exists $\bar{x}$ such that $\psi(\bar{x}) > 1$, we have $F^*(\psi) = +\infty$. Indeed, assume by contradiction that $F^*(\psi) = 0$, that is there exist $\psi_1, \psi_2 \in \mathcal{C}_0(\Omega, \mathbb{R})$ such that $\psi_1 + \psi_2 = \psi$ with $\|\psi_1\|_{\infty,\Omega} \leq 1$ and $\psi_2(x) \leq 0 \ \forall x \in \Omega$. We would have

$$\psi(\bar{x}) = \underbrace{\psi_1(\bar{x})}_{\leq 1} + \underbrace{\psi_2(\bar{x})}_{\leq 0} \leq 1$$

which is absurd, since $\psi(\bar{x}) > 1$. Then, $\forall \psi_1, \psi_2$ such that $\psi_1 + \psi_2 = \psi$ either $\|\psi_1\|_{\infty,\Omega} > 1$ or $\psi_2(x) > 0$ for some $x \in \Omega$. Hence, $F^*(\psi) = +\infty$.

iii) If $\|\psi\|_{\infty,\Omega} > 1$ and $\psi(x) \leq 1 \ \forall x \in \Omega$, consider $\psi_1 = \psi^+$ and $\psi_2 = \psi^-$, with

$$\psi^+(x) = \begin{cases} \psi(x) & \psi(x) \geq 0 \\ 0 & \psi(x) < 0 \end{cases} \qquad \psi^-(x) = \begin{cases} 0 & \psi(x) > 0 \\ \psi(x) & \psi(x) \leq 0 \end{cases}.$$

It is evident that $\psi = \psi^+ + \psi^-$. Moreover, $\|\psi^+\|_{\infty,\Omega} \leq 1$, thus $A^*(\psi^+) = 0$, and $\forall\, x \in \Omega \ \psi^-(x) \leq 0$, thus $B^*(\psi^-) = 0$, so that $F^*(\psi) = 0$.

Since (i) and (iii) are equivalent to $\psi(x) \leq 1, \ \forall x \in \Omega$, and the three items cover all possible cases, the thesis is proved. □

2.2 Dual problem

The study of dual problems of sparse-regularisation models with non-negativity constraints has been carried out in [38, 39] in the discrete setting of LASSO, and in [40] for the discrete counterpart of $(\tilde{\mathcal{D}}_{KL} - |\cdot|)$, that is with the Kullback-Leibler

divergence as fidelity and the L^1 penalty. In the following we use these results to compute the dual problem of $(\tilde{\mathcal{D}}_{KL} - |\cdot|)$ in $\mathcal{M}(\Omega)$.

By Fenchel-Rockafellar duality (Lemma 6 in Appendix A), the dual problem of $(\tilde{\mathcal{D}}_{KL} - |\cdot|)$ can be obtained by plugging (20) and (23) into (A3). We thus have:

$$
\begin{aligned}
&\operatorname*{argmax}_{p^* \in L^2(\Omega)} \; -F^*(\Phi^* p^*) - G^*(-p^*) \\
&= \operatorname*{argmax}_{p^* \in L^2(\Omega)} \; -F^*(\Phi^* p^*) + \begin{cases} -\infty & -p^* \geq \frac{1}{\lambda} \\ -\frac{b}{\lambda}(1+\lambda p^*) + \frac{y}{\lambda}\log(1+\lambda p^*) & -p^* < \frac{1}{\lambda} \end{cases} \\
&= \operatorname*{argmax}_{p^* \in L^2(\Omega)} \; -F^*(\Phi^* p^*) + \begin{cases} -\infty & p^* \leq -\frac{1}{\lambda} \\ -\frac{b}{\lambda}(1+\lambda p^*) + \frac{y}{\lambda}\log(1+\lambda p^*) & p^* > -\frac{1}{\lambda} \end{cases} \\
&= \operatorname*{argmax}_{p^* \in L^2(\Omega) \text{ s.t. } p^* > -\frac{1}{\lambda}} \; -F^*(\Phi^* p^*) - \frac{b}{\lambda}(1+\lambda p^*) + \frac{y}{\lambda}\log(1+\lambda p^*) \\
&= \operatorname*{argmax}_{p^* \in L^2(\Omega) \text{ s.t. } p^* > -\frac{1}{\lambda}} \begin{cases} 0 & \forall x \in \Omega, \;\; \Phi^* p^*(x) \leq 1 \\ -\infty & \exists x \in \Omega, \;\; \Phi^* p^*(x) > 1 \end{cases} \; - \frac{b}{\lambda}(1+\lambda p^*) + \frac{y}{\lambda}\log(1+\lambda p^*) \\
&= \operatorname*{argmax}_{p^* \in \mathcal{D}} \; -\frac{b}{\lambda}(1+\lambda p^*) + \frac{y}{\lambda}\log(1+\lambda p^*)\,, \qquad (24)
\end{aligned}
$$

where $\mathcal{D} = \{p^* \in L^2(\Omega) : p^* > -\frac{1}{\lambda} \text{ and } \forall x \in \Omega, \;\; \Phi^* p^*(x) \leq 1\}$.

2.3 Extremality conditions

By Lemma 6, extremality conditions (A4) can be obtained. Given $\mu_\lambda \in \mathcal{M}(\Omega)$ solution of the primal problem $(\tilde{\mathcal{D}}_{KL} - |\cdot|)$ with regularisation parameter $\lambda > 0$ and $p^*_\lambda \in L^2(\Omega)$ solution of the dual problem (24), they read

$$
\begin{cases} \Phi^* p^*_\lambda \in \partial F(\mu_\lambda) = \partial |\mu_\lambda|(\Omega) + \partial \mathbb{1}_{\mathcal{M}^+(\Omega)}(\mu_\lambda) \\ -p^*_\lambda \in \frac{1}{\lambda}\partial_1 \tilde{\mathcal{D}}_{KL}(\Phi\mu_\lambda + b, y) = \frac{1}{\lambda}\left(I - \frac{y}{\Phi\mu_\lambda + b}\right) \end{cases}, \qquad (25)
$$

where $\partial_1 \tilde{\mathcal{D}}_{KL}(\Phi\mu_\lambda + b, y)$ denotes the subdifferential of $\tilde{\mathcal{D}}_{KL}(\cdot,\cdot)$ computed with respect to the first variable and evaluated in $(\Phi\mu_\lambda + b, y)$ and I denotes the identity operator.

Remark 1. *If $\mu_\lambda \in \mathcal{M}(\Omega)$ is solution of the primal problem $(\tilde{\mathcal{D}}_{KL} - |\cdot|)$ and $p^*_\lambda \in L^2(\Omega)$ is solution of the dual problem (24), then from (25) we have*

$$
-p^*_\lambda = \frac{1}{\lambda}\left(I - \frac{y}{\phi\mu_\lambda + b}\right) \quad \Rightarrow \quad p^*_\lambda = \frac{y - \Phi\mu_\lambda - b}{\lambda(\Phi\mu_\lambda + b)}. \qquad (26)
$$

It follows that $p^*_\lambda > -\frac{1}{\lambda} \iff y > 0$, which holds by hypothesis (16).

In the extremality conditions (25), the set-sum $\partial|\mu_\lambda|(\Omega) + \partial\mathbb{1}_{\mathcal{M}^+(\Omega)}(\mu_\lambda)$ requires attention. Recall that, in general, for two proper convex functions $f_1, f_2 : \mathcal{X} \to \mathbb{R} \cup$

$\{+\infty\}$ there holds $\partial(f_1 + f_2) \subseteq \partial f_1 + \partial f_2$, with the equality holding if and only if $\text{int}(\text{dom}(f_1)) \cap \text{int}(\text{dom}(f_2)) \neq \emptyset$, see, e.g., [41].

In the following we show that $\partial F(\cdot) = \partial |\cdot|(\Omega) + \partial \mathbb{1}_{\mathcal{M}^+(\Omega)}(\cdot)$. Let $\mu \in \mathcal{M}^+(\Omega)$ and consider B as in (21). By definition, the subdifferential of B at $\mu \in \mathcal{M}^+(\Omega)$ is

$$\begin{aligned}\partial B(\mu) &= \{\eta \in \mathcal{C}_0(\Omega, \mathbb{R})|\ B(\tilde{\mu}) \geq B(\mu) + \langle \eta, \tilde{\mu} - \mu \rangle\ \forall \tilde{\mu} \in \mathcal{M}(\Omega)\} \\ &= \{\eta \in \mathcal{C}_0(\Omega, \mathbb{R})|\ \langle \eta, \tilde{\mu} - \mu \rangle \leq 0\ \forall \tilde{\mu} \in \mathcal{M}^+(\Omega)\}, \end{aligned} \tag{27}$$

that is the normal cone to $\mathcal{M}^+(\Omega)$ at μ. We now consider (27) for some scenarios of interest:

- If $\mu = 0$, we have $\eta \in \partial B(0)$ if and only if $\forall \tilde{\mu} \in \mathcal{M}^+(\Omega)$ there holds $\langle \eta, \tilde{\mu} \rangle \leq 0$. Since $\tilde{\mu}$ is a positive measure, we have that $\langle \eta, \tilde{\mu} \rangle = \int_\Omega \eta \mathrm{d}\tilde{\mu} \leq 0$ for any η such that $\eta(x) \leq 0$ for all $x \in \Omega$. Thus,

$$\partial B(0) = \{\eta \in \mathcal{C}_0(\Omega, \mathbb{R})| \quad \eta(x) \leq 0 \quad \forall x \in \Omega\}.$$

- If $\mu = \bar{a}\delta_{\bar{x}}$ with $\bar{a} > 0$, $\bar{x} \in \Omega$, since $\eta \in \partial B(\mu)$ if and only if $\langle \eta, \tilde{\mu} - \mu \rangle \leq 0\ \forall \tilde{\mu} \in \mathcal{M}^+(\Omega)$, then $\eta \in \partial B(\mu)$ if and only if

$$\int_\Omega \eta \mathrm{d}\tilde{\mu} = \langle \eta, \tilde{\mu} \rangle \leq \langle \eta, \mu \rangle = \int_\Omega \eta \mathrm{d}\mu = \bar{a}\eta(\bar{x}).$$

  Consider $\tilde{\mu} = 2\mu$. Hence, $\int_\Omega \eta \mathrm{d}\tilde{\mu} = \int_\Omega \eta 2\mathrm{d}\mu = 2\bar{a}\eta(\bar{x})$ and $2\bar{a}\eta(\bar{x}) \leq \bar{a}\eta(\bar{x})$ if and only if $\eta(\bar{x}) = 0$. Thus, $\eta \in \partial B(\mu)$ implies $\eta(\bar{x}) = 0$.
  Let $\eta \in \partial B(\mu)$, which entails $\eta(\bar{x}) = 0$, and suppose $\exists \tilde{x} \in \Omega$ s.t. $\eta(\tilde{x}) > 0$. Then,

$$\forall \tilde{\mu} \in \mathcal{M}^+(\Omega) \qquad \int_\Omega \eta \mathrm{d}\tilde{\mu} \leq \int_\Omega \eta \mathrm{d}\mu = \bar{a}\eta(\bar{x}) = 0.$$

  The above inequality with $\tilde{\mu} = \tilde{a}\delta_{\tilde{x}}$, $\tilde{a} > 0$, yields $0 \geq \int_\Omega \eta \mathrm{d}\tilde{\mu} = \tilde{a}\eta(\tilde{x}) > 0$, which is absurd. Thus, $\eta(x) \leq 0\ \forall x \in \Omega$ and

$$\partial B(\mu) = \{\eta \in \mathcal{C}_0(\Omega, \mathbb{R})|\ \eta(x) \leq 0\ \forall x \in \Omega \ \text{ and } \eta(\bar{x}) = 0\}.$$

- Similarly, if $\mu = \sum_{i=1}^N a_i \delta_{x_i}$ with $a_i > 0$ and $x_i \in \Omega$ for $i = 1, \ldots, N$, then we have:

$$\partial \mathbb{1}_{\mathcal{M}^+(\Omega)}(\mu) = \partial B(\mu) = \left\{\eta \in \mathcal{C}_0(\Omega, \mathbb{R}) \middle|\ \eta(x) \leq 0\ \forall x \in \Omega \ \text{ and } \eta(x_i) = 0,\ i = 1, \ldots, N\right\}. \tag{28}$$

We now have an analytical expression for both $\partial |\cdot|(\Omega)$ and $\partial \mathbb{1}_{\mathcal{M}^+(\Omega)}(\cdot)$, whenever μ is a finite linear combination of Diracs. We thus state the following proposition.

Proposition 1. *If $\mu = \sum_{i=1}^N a_i \delta_{x_i}$ with $a_i > 0$, $x_i \in \Omega$, $i = 1, \ldots, N$, we have*

$$\partial |\mu|(\Omega) + \partial \mathbb{1}_{\mathcal{M}^+(\Omega)}(\mu) = \left\{\eta \in \mathcal{C}_0(\Omega, \mathbb{R} \middle|\ \eta(x) \leq 1\ \forall x \in \Omega \text{ and } \eta(x_i) = 1,\ i = 1, \ldots, N\right\}. \tag{29}$$

Proof. If $\mu = \sum_{i=1}^{N} a_i \delta_{x_i}$ with $a_i > 0$, $x_i \in \Omega$, by (4) and (28) we have

$$\partial|\mu|(\Omega) = \{\eta_1 \in \mathcal{C}_0(\Omega, \mathbb{R}) : \ \|\eta_1\|_\infty \leq 1 \text{ and } \eta_1(x_i) = 1, \ i = 1, \dots, N\},$$

$$\partial \mathbb{1}_{\mathcal{M}^+(\Omega)}(\mu) = \{\eta_2 \in \mathcal{C}_0(\Omega, \mathbb{R}) : \ \forall x \in \Omega \ \eta_2(x) \leq 0 \text{ and } \eta_2(x_i) = 0, \ i = 1, \dots, N\}.$$

Thus, we have the following inclusion

$$\partial|\mu|(\Omega) + \partial \mathbb{1}_{\mathcal{M}^+(\Omega)}(\mu) \subseteq \{\eta \in \mathcal{C}_0(\Omega, \mathbb{R}) \big| \ \eta(x) \leq 1 \ \forall x \in \Omega \text{ and } \eta(x_i) = 1, \ i = 1, \dots, N\}.$$

To show that the latter is indeed an equality, we verify now the opposite inclusion. Let $\eta \in \mathcal{C}_0(\Omega, \mathbb{R})$ such that $\eta(x) \leq 1$ $x \in \Omega$ and $\eta(x_i) = 1$, $i = 1, \dots, N$. Our aim is to prove that there exist $\eta_1 \in \partial|\mu|(\Omega)$ and $\eta_2 \in \partial \mathbb{1}_{\mathcal{M}^+(\Omega)}(\mu)$ such that $\eta = \eta_1 + \eta_2$. Let's consider the two cases $\|\eta\|_\infty \leq 1$ and $\|\eta\|_\infty > 1$.

i) If $\|\eta\|_\infty \leq 1$, we can take $\eta_1 = \eta$, $\eta_2 = 0$.
ii) If $\|\eta\|_\infty > 1$, we can take

$$\eta_1(x) = \begin{cases} \eta(x) & -1 \leq \eta(x) \leq 1 \\ -1 & \eta(x) \leq -1 \end{cases}, \qquad \eta_2(x) = \begin{cases} 0 & -1 \leq \eta(x) \leq 1 \\ \eta(x) + 1 & \eta(x) \leq -1 \end{cases}.$$

In both cases, we have found $\eta_1 \in \partial|\mu|(\Omega)$ and $\eta_2 \in \partial \mathbb{1}_{\mathcal{M}^+(\Omega)}(\mu)$ such that $\eta = \eta_1 + \eta_2$, so that the opposite inclusion also holds. This concludes the proof. □

Thanks to (29) of the latter proposition, it is now possible to better characterise the extremality conditions (25), under the assumption that μ_λ, solution of $(\tilde{\mathcal{D}}_{KL} - |\cdot|)$, is a discrete measure. Indeed, if $\mu_\lambda = \sum_{i=1}^{N_\lambda} (a_\lambda)_i \delta_{(x_\lambda)_i}$ we have that (25) is equivalent to

$$\Phi^* p_\lambda^*(x) \leq 1 \quad \forall x \in \Omega \quad \text{and} \quad \Phi^* p_\lambda^*\big((x_\lambda)_i\big) = 1, \ i = 1, \dots, N_\lambda, \tag{30}$$

with p_λ^* solution of the dual problem (24), explicitly given by (26).

It is possible to obtain a result similar to (30) by analysing the optimality conditions of $(\tilde{\mathcal{D}}_{KL} - |\cdot|)$. The functional

$$T_\lambda^{KL}(\mu) := \tilde{\mathcal{D}}_{KL}(\Phi\mu + b, y) + \lambda|\mu|(\Omega) + \mathbb{1}_{\{\mathcal{M}^+(\Omega)\}}(\mu)$$

is convex. Hence μ_λ minimises T_λ^{KL} if and only if $0 \in \partial T_\lambda^{KL}(\mu_\lambda)$, i.e.

$$\begin{aligned} 0 &\in \Phi^* \partial_1 \tilde{\mathcal{D}}_{KL}(\Phi\mu_\lambda, y) + \lambda \partial|\mu_\lambda|(\Omega) + \partial \mathbb{1}_{\mathcal{M}^+(\Omega)}(\mu_\lambda) \\ &= \Phi^*\Big(I - \frac{y}{\Phi\mu_\lambda + b}\Big) + \lambda \partial|\mu_\lambda|(\Omega) + \partial \mathbb{1}_{\mathcal{M}^+(\Omega)}(\mu_\lambda)\,, \end{aligned}$$

or equivalently

$$\frac{1}{\lambda}\Phi^*\Big(\frac{y}{\Phi\mu_\lambda + b} - I\Big) \in \partial|\mu_\lambda|(\Omega) + \frac{1}{\lambda}\partial \mathbb{1}_{\mathcal{M}^+(\Omega)}(\mu_\lambda)\,,$$

which can be written as

$$\eta \in \partial|\mu_\lambda|(\Omega) + \frac{1}{\lambda}\partial \mathbb{1}_{\mathcal{M}^+(\Omega)}(\mu_\lambda)\,, \tag{31}$$

by defining the dual certificate of $(\tilde{\mathcal{D}}_{KL} - |\cdot|)$

$$\eta := \frac{1}{\lambda}\Phi^*\Big(\frac{y}{\Phi\mu_\lambda + b} - I\Big). \tag{32}$$

If μ is a finite linear combination of Diracs, Proposition 1 yields the explicit expression (29) for the sum of the two subdifferentials of (31). Hence, inclusion (31) simply becomes

$$\eta(x) \leq 1 \;\; \forall x \in \Omega \quad \text{and} \quad \eta(x_i) = 1,\; i = 1, \dots, N. \tag{33}$$

Note the similarity between (33) and (30), and of the dual certificate (32) and the quantity $\Phi^* p_\lambda^*$ in the extremality conditions (25).

3 The Sliding Frank-Wolfe algorithm

Both problems $(\tilde{\mathcal{D}}_{KL} - |\cdot|)$ and $(L^2 - |\cdot|)$ are defined over the space $\mathcal{M}(\Omega)$, an infinite dimensional non-reflexive Banach space. Due to non-reflexivity, it is not simple to define therein proximal-based algorithms, see [42] for some recent attempt. Any solver for such problems shall take into account the infinite dimensional nature of $\mathcal{M}(\Omega)$: popularly used algorithms in this setting are semi-definite programming approaches (for Fourier measurements) [7], conditional gradient algorithms [12, 43, 44] and particle gradient descent [21, 45], which is an optimal-transport based algorithm. In this work, we will only focus on conditional gradient strategies, namely the Frank-Wolfe and Sliding Frank-Wolfe algorithms, see [46] for a survey.

In [12], the authors detail how the conditional gradient algorithm, also known as Frank-Wolfe algorithm [43], can be used to minimise the BLASSO $(L^2 - |\cdot|)$ functional using an epigraphic lift, and then propose the Sliding version, which significantly improves the reconstruction quality by adding an extra step where both positions and amplitudes are re-optimised. Similar strategies are here used to minimise $(\tilde{\mathcal{D}}_{KL} - |\cdot|)$. We report here the statement of the result without a proof as it follows from [12, Lemma 4].

Lemma 4. *The solution $\bar{\mu} \in \mathcal{M}(\Omega)$ to $(\tilde{\mathcal{D}}_{KL} - |\cdot|)$ is equivalent to the solution $\bar{\mu} \in \mathcal{M}(\Omega)$ to the problem*

$$\operatorname*{argmin}_{(t,\mu)\in C} \; \tilde{T}_\lambda^{KL}(\mu, t) := \tilde{\mathcal{D}}_{KL}(\Phi\mu + b, y) + \lambda t + \mathbb{1}_{\{\mathcal{M}^+(\Omega)\}}(\mu),$$

where $C := \{(t,\mu) \in \mathbb{R}^+ \times \mathcal{M}(\Omega); |\mu|(\Omega) \leq t \leq M\}$ with $M := \frac{\tilde{\mathcal{D}}_{KL}(b,y)}{2\lambda}$.

Given a convex and smooth fidelity functional $f_{y^\delta,b} : \mathcal{M}(\Omega) \to \mathbb{R}$ defined in terms of the observed data $y^\delta \in L^2(\Omega)$ and, potentially, a background term $b^\delta \in L^2(\Omega)+$,

Algorithm 1 Sliding Frank-Wolfe (SFW) algorithm [12]

Initialisation: $\mu^0 = 0$.

repeat for $k = 1, 2, \dots, K_{\max}$

$\mu^k = \sum_{i=1}^{N^k} a_i^k \delta_{x_i^k}, a_i^k \in \mathbb{R}, x_i^k \in \Omega$, find $x_*^k \in \Omega$ s.t.:

$$x_*^k \in \operatorname*{argmax}_{x \in \Omega} \left| \eta(\lambda, \mu^k)(x) \right| \quad \text{where} \quad \eta(\lambda, \mu^k) \quad \text{is defined by (34) with } \mu = \mu^k$$

if $\left| \eta(\lambda, \mu^k) \left(x_*^k \right) \right| \leq 1$

μ^k is a solution of $(\mathcal{P}(\lambda)) \Rightarrow$ **stop**

else

– **insertion step**: add support for the new spike and amplitudes' estimation

$$\begin{aligned} x^{k+1/2} &= \left(x_1^k, \dots, x_{Nk}^k, x_*^k \right) \\ a^{k+1/2} &\in \operatorname*{argmin}_{a \in \mathbb{R}^{N^k+1}} f_{y^\delta, b} \left(\Phi_{x^{k+1/2}} a \right) + \lambda \|a\|_1 + \alpha \mathbb{1}_{\geq 0}(a) \\ \text{Update:} \qquad \mu^{k+1/2} &= \sum_{i=1}^{N^k} a_i^{k+1/2} \delta_{x_i^k} + a_{N^k+1}^{k+1/2} \delta_{x_*^k} \end{aligned}$$

– **sliding step**: using a non-convex solver initialised with $\left(a^{k+1/2}, x^{k+1/2} \right)$

$$\begin{aligned} \left(a^{k+1}, x^{k+1} \right) &\in \operatorname*{arg\,min}_{(a,x) \in \mathbb{R}^{N^k+1} \times \Omega^{N^k+1}} f_{y^\delta, b} \left(\Phi_x a \right) + \lambda \|a\|_1 + \alpha \mathbb{1}_{\geq 0}(a) \\ \text{Update:} \qquad \mu^{k+1} &= \sum_{i=1}^{N^k+1} a_i^{k+1} \delta_{x_i^{k+1}} \end{aligned}$$

– **pruning**: eventually remove zero amplitudes Dirac masses from μ^{k+1}

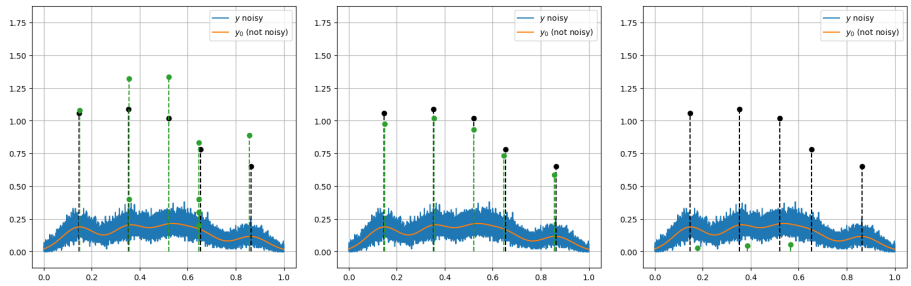
until convergence

we consider the following general optimisation problem:

$$\operatorname*{argmin}_{\mu \in \mathcal{M}(\Omega)} f_{y^\delta, b}(\Phi \mu) + \lambda |\mu|(\Omega) + \alpha \mathbb{1}_{\mathcal{M}^+(\Omega)}(\mu), \qquad \lambda > 0, \quad \alpha \in \{0, 1\}, \qquad (\mathcal{P}(\lambda))$$

which encompasses the BLASSO as well as problem $(\mathcal{P}(\lambda))$ and where $\alpha \in \{0, 1\}$ may enforce non-negativity constraints.

The pseudocode corresponding to the optimisation of the general problem $(\mathcal{P}(\lambda))$ is reported in Algorithm 1. Observe that the stopping criterion is expressed in terms of the dual certificate $\eta(\lambda, \mu) \in L^2(\Omega)$ of the general problem $(\mathcal{P}(\lambda))$, which is defined



(a) $\lambda \approx 0$ (b) $\lambda = 1$ (c) $\lambda = 3000$

Figure 2: Reconstructions obtained using SFW for a 1D sparse deconvolution problem with Poisson noise. Ground truth spikes (black) and reconstructed ones (green) using Algorithm 1 for some choices of λ are shown. When $\lambda \approx 0$, the number and intensities of spikes are overestimated, while for $\lambda \gg 0$ they are underestimated.

in terms of the subgradient of the fidelity and reads

$$\eta(\lambda, \mu) = \frac{1}{\lambda}\tilde{\eta}(\mu) \quad \text{with} \quad \tilde{\eta}(\mu) := \begin{cases} -\Phi^* \frac{\partial}{\partial \mu} f_{y^\delta, b}(\Phi\mu) & \alpha = 0 \\ \left(-\Phi^* \frac{\partial}{\partial \mu} f_{y^\delta, b}(\Phi\mu)\right)_+ & \alpha = 1 \end{cases}, \tag{34}$$

depending on the parameter α so that the dual certificate (8) of ($L^2 - |\cdot|$) and the dual certificate (32) of ($\tilde{\mathcal{D}}_{KL} - |\cdot|$) can be obtained by choosing α and $f_{y^\delta, b}$, respectively. The generalised optimality condition for ($\mathcal{P}(\lambda)$) reads

$$\|\eta(\lambda, \mu)\|_\infty \leq 1, \tag{35}$$

and, under the hypothesis that the solution is a finite linear combination of Diracs as (3), the dual certificate of the solution satisfies $\eta(\lambda, \mu)(x_i) = \text{sign}(a_i)$ where the points $x_i \in \Omega$ are the support of μ.

4 Parameter selection via algorithmic homotopy

The performance of Algorithm 1 for solving ($\mathcal{P}(\lambda)$) depends on the choice of the regularisation parameter $\lambda > 0$: it plays indeed a fundamental role in the sparsity pattern of the solution and in the enforcement of the stopping rule, $\max_{x \in \Omega} \left|\eta(\lambda, \mu^k)(x)\right| \leq 1$. Namely, a high value of λ forces only few iterations of the algorithm to be performed. Such choice impacts also both the estimation and the sliding steps of Algorithm 1, being it associated with the sparsity-promoting L^1 penalty used to compute the amplitude vector. On the other hand, smaller values of λ provide a better data fit, with more spikes with higher intensities, see Figure 2.

In [47], the authors propose a method based on algorithmic homotopy [48, 49] to choose an optimal regularisation parameter for the resolution of BLASSO ($L^2 - |\cdot|$).

The idea behind homotopy algorithms is to avoid the exploration of the whole Pareto frontier by grid search, while providing an iterative procedure computing only few parameters up to a target value. More precisely, starting from an initial overestimated value $\lambda_1 > 0$, a solution μ_{λ_1} to $(\mathcal{P}(\lambda))$ is computed by, e.g., Algorithm 1. At each homotopy iteration, if the solution does not fit well the data up to some tolerance σ_{target} depending on the noise magnitude δ, then λ is decreased. A new solution μ_{λ_2} to $(\mathcal{P}(\lambda))$ is thus computed in the next homotopy step and so on. Homotopy algorithms thus explore the Pareto frontier for a small set of values λ and select its biggest value for which the solution to $(\mathcal{P}(\lambda))$ meets a convergence criterion depending on δ. In Figure 3, one can see in red the discrete values produced by the homotopy strategy we are going to describe, which stops when the fidelity term goes under the value of $\sigma_{\text{target}}(\delta)$, in grey.

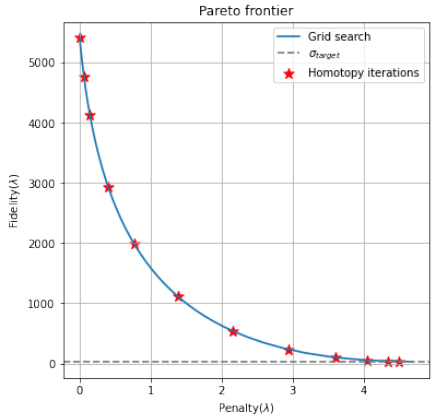


Figure 3: Homotopy algorithms explore the Pareto frontier iteratively for a strictly decreasing sequence of regularisation parameters λ. In blue: fine discretisation of the Pareto frontier with grid search. In red: homotopy iterations, corresponding to different values of λ. In gray: target value for the fidelity.

In th following we follow [47], where Sliding Frank-Wolfe with homotopy is proposed for the BLASSO problem $(L^2 - |\cdot|)$. We extend the strategy in Algorithm 2 for a general fidelity term within an off-the-grid setting. Note that each homotopy iteration $t \in \mathbb{N}$ performs the following steps:

- Compute $\mu_{\lambda_t} := \hat{\mu}_t$, solution of $(\mathcal{P}(\lambda))$ with $\lambda = \lambda_t$.
- Check if $\sigma_t := f_{y^\delta,b}(\Phi\mu_{\lambda_t}) \leq \sigma_{\text{target}}(\delta)$, where the target value $\sigma_{\text{target}}(\delta)$ depends on the noise level $\delta > 0$.
- If the condition above is not met, decrease λ: $\lambda_{t+1} < \lambda_t$.

In the following we discuss the choice of the starting value for the sequence of λ's, its updating rule and the choice of a suitable value σ_{target} for the noise scenario considered. In Table 1, we outline the different choices made for Alg.2 to both $(L^2 - |\cdot|)$ and $(\tilde{\mathcal{D}}_{KL} - |\cdot|)$.

Algorithm 2 Homotopy algorithm in $\mathcal{M}(\Omega)$

Input: $y \in L^2(\Omega)$, $b \in L^2(\Omega)$, $b \geq 0$, $\Phi \in \mathcal{L}(\mathcal{M}(\Omega), L^2(\Omega))$
Parameters: $\gamma \in (0,1)$, $c > 0$, $\sigma_{\text{target}} > 0$
Output: optimal $\hat{\mu} \in \mathcal{M}(\Omega)$
Initialisation: $\hat{\mu}_0 \in \mathcal{M}(\Omega)$ and $\lambda_1 = \gamma \left\| \eta(1, \hat{\mu}_0) \right\|_\infty$
repeat for $t = 1, 2, \ldots, T_{\max}$

1. Compute $\hat{\mu}_t$ solution of $(\mathcal{P}(\lambda))$ with $\lambda = \lambda_t$ with warm start $\mu_t^{[0]} = \hat{\mu}_{t-1}$.
2. Compute σ_t from the residual:

$$\sigma_t = f_{y^\delta, b}(\Phi \hat{\mu}_t)$$

3. **if** $\sigma_t < \sigma_{\text{target}}$

   $\hat{\mu}_t$ is a solution $\Rightarrow$ **stop**

4. **else if** $\sigma_t \geq \sigma_{\text{target}}$

$$\text{Update} \quad \lambda_{t+1} = \frac{\lambda_t \left\| \eta(\lambda_t, \hat{\mu}_t) \right\|_\infty}{c+1} \tag{36}$$

until $\sigma_t < \sigma_{\text{target}}$

		$(L^2 - \lvert\cdot\rvert)$	$(\tilde{\mathcal{D}}_{KL} - \lvert\cdot\rvert)$
λ_1	$:=$	$\gamma \lVert \Phi^* y^\delta \rVert_\infty$	$\gamma \left\| \left(\Phi^* \left(\frac{y-b}{b} \right) \right)_+ \right\|_\infty$
σ_t	$:=$	$\frac{1}{2} \lVert \Phi \hat{\mu}_t - y^\delta \rVert^2$	$\tilde{\mathcal{D}}_{KL}(\Phi \hat{\mu}_t, y^\delta)$
$\sigma_{\text{target}}(\delta)$	$:=$	$\frac{1}{2} \lVert y - y^\delta \rVert^2 = \frac{\delta^2}{2}$	$\tilde{\mathcal{D}}_{KL}(y, y^\delta)$

Table 1: Homotopy algorithmic choices for problems $(L^2 - \lvert\cdot\rvert)$ and $(\tilde{\mathcal{D}}_{KL} - \lvert\cdot\rvert)$.

Starting value

The element $\eta(\lambda, \mu)$ is crucial for the definition of Algorithm 2 as it allows to define a good starting value λ_1. We propose to initialise as follows:

$$\lambda_1 := \gamma \left\| \eta(1, \hat{\mu}_0) \right\|_\infty = \gamma \left\| \tilde{\eta}(\hat{\mu}_0) \right\|_\infty, \qquad \gamma \in (0,1) \tag{37}$$

where $\hat{\mu}_0 \in \mathcal{M}(\Omega)$ is the initialisation of the solution and η is the dual certificate (34). The parameter $\gamma \in (0,1)$ is a relaxation parameter usually chosen close to 1. The choice (37) is motivated by optimality conditions (35). If one takes at the first iteration $\lambda_1 \geq \left\| \eta(1, \hat{\mu}_0) \right\|_\infty$, then $\hat{\mu}_0$ is an optimal solution for $(\mathcal{P}(\lambda))$ with $\lambda = \lambda_1$, since

$$\left\| \eta(\lambda_1, \hat{\mu}_0) \right\|_\infty = \left\| \frac{1}{\lambda_1} \tilde{\eta}(\hat{\mu}_0) \right\|_\infty = \frac{1}{\lambda_1} \left\| \eta(1, \hat{\mu}_0) \right\|_\infty \leq 1 \quad \Longleftrightarrow \quad \lambda_1 \geq \left\| \eta(1, \hat{\mu}_0) \right\|_\infty .$$

In this case, the algorithm does not improve upon the initialisation $\hat{\mu}_0$ since it does not perform any iteration. On the contrary, choosing $\lambda_1 < \|\eta(1, \hat{\mu}_0)\|_\infty$ ensures that the initial measure $\hat{\mu}_0$ is updated since the dual certificate computed with respect to the initialisation $\hat{\mu}_0$ and $\lambda_1 > 0$ is such that (37) has supremum norm that satisfies

$$\|\eta(\lambda_1, \hat{\mu}_0)\|_\infty = \frac{1}{\lambda_1} \|\eta(1, \hat{\mu}_0)\|_\infty = \frac{1}{\gamma \|\eta(1, \hat{\mu}_0)\|_\infty} \|\eta(1, \hat{\mu}_0)\|_\infty = \frac{1}{\gamma} > 1.$$

Updating rule

The updating rule (36) for λ together with the choice of a strictly positive parameter $c > 0$ ensures that the measure $\hat{\mu}_t$, which is used to initialise $(\mathcal{P}_{\lambda_{t+1}})$ as $\mu_{t+1}^{[0]} = \hat{\mu}_t$, is not an optimal solution for the problem. Indeed, the dual certificate computed in correspondence with λ_{t+1} and $\hat{\mu}_t$ reads

$$\eta(\lambda_{t+1}, \hat{\mu}_t) = \frac{\lambda_t}{\lambda_{t+1}} \eta(\lambda_t, \hat{\mu}_t) = \frac{1+c}{\|\eta(\lambda_t, \hat{\mu}_t)\|_\infty} \eta(\lambda_t, \hat{\mu}_t) \implies \|\eta(\lambda_{t+1}, \hat{\mu}_t)\|_\infty = 1 + c > 1.$$

Thus, $\mu_{t+1}^{[0]} = \hat{\mu}_t$ does not satisfy (35) for $(\mathcal{P}(\lambda))$ with $\lambda = \lambda_{t+1}$, so the homotopy step $t+1$ computes a new candidate solution $\hat{\mu}_{t+1}$. This is indeed consistent with the choice of rejecting $\hat{\mu}_t$ at the previous step t.

4.1 Descent property

In this section, we show that the homotopy algorithm 2 produces a strictly decreasing sequence of residual errors $(\sigma_t)_t$. This properties gives insight on the good convergence of the algorithm. We remark that the proof follows [47, Proposition 3.1] but is here adapted to more general choices of the data term $f_{y^\delta, b}$.

Proposition 2. *The homotopy algorithm (Alg. 2) for the resolution of $(\mathcal{P}(\lambda))$ produces a strictly decreasing sequence of residual distances $(\sigma_t)_{t \in \mathbb{N}}$ provided that $f_{y^\delta, b}(\cdot)$ is twice differentiable and strictly convex.*

Proof. We consider

$$\mu^*(\lambda) \in \operatorname*{argmin}_\mu \; C(\mu; y^\delta, \lambda) := f_{y^\delta, b}(\Phi\mu) + \lambda |\mu|(\Omega),$$

where the argmin is computed over $\mathcal{M}(\Omega)$ if $\alpha = 0$ in $(\mathcal{P}(\lambda))$ and over $\mathcal{M}^+(\Omega)$ if $\alpha = 1$. Denoting by $\partial/\partial\mu$ the Radon-Nikodym derivative w.r.t. μ, we have that the minimiser $\mu^*(\lambda)$ satisfies

$$\frac{\partial C(\mu; y^\delta, \lambda)}{\partial \mu}\Big|_{\mu=\mu^*} = \Phi^* \partial f_{y^\delta, b}(\Phi\mu^*) + \lambda = 0, \tag{38}$$

Deriving the expression above with respect to λ and applying the chain rule yields:

$$0 = \frac{\partial}{\partial \lambda} \Big(\frac{\partial C(y, \lambda, \mu)}{\partial \mu}\Big|_{\mu=\mu^*} \Big) = \frac{\partial}{\partial \lambda} \Big(\Phi^* \partial f_{y^\delta, b}(\Phi\mu^*) + \lambda \Big)$$

$$0 = \frac{\partial}{\partial \mu}\Big(\Phi^* \partial f_{y^\delta,b}(\Phi\mu)\Big)\Big|_{\mu=\mu^*} \cdot \frac{\partial \mu^*(\lambda)}{\partial \lambda} + 1$$
$$0 = \Phi^*\Phi \cdot \partial^2 f_{y^\delta,b}(\Phi\mu^*) \cdot \frac{\partial \mu^*(\lambda)}{\partial \lambda} + 1.$$

Since Φ is positive definite and $f_{y^\delta,b}(\cdot)$ is stricly convex, from the computations above it follows that $\frac{\partial \mu^*(\lambda)}{\partial \lambda} < 0$. Consider now the following:

$$\frac{\partial}{\partial \lambda} f_{y^\delta,b}\big(\Phi\mu^*(\lambda)\big) = \Phi^* \partial f_{y^\delta,b}(\Phi\mu^*) \cdot \frac{\partial \mu^*(\lambda)}{\partial \lambda}.$$

Recalling (38), we notice that $\frac{\partial}{\partial \mu} f_{y^\delta,b}(\Phi\mu^*) = -\lambda$. Moreover, since $\frac{\partial \mu^*(\lambda)}{\partial \lambda} < 0$, then $\frac{\partial}{\partial \lambda} f_{y^\delta,b}\big(\Phi\mu^*(\lambda)\big) > 0$. Since by homotopy $\lambda_{t+1} < \lambda_t$, and denoting by $\hat{\mu}_t := \mu^*(\lambda_t)$, we have

$$\sigma_t = f_{y^\delta,b}(\Phi\hat{\mu}_t) < \sigma_{t+1} = f_{y^\delta,b}(\Phi\hat{\mu}_{t+1}),$$

which concludes the proof. □

Remark 2. *For BLASSO ($L^2 - |\cdot|$), the fidelity $f_{y^\delta,b}(\Phi\mu) = \frac{1}{2}\|\Phi\mu - y\|^2$ satisfies the hypothesis required to prove Proposition 2. Indeed,*

$$\partial^2 f_{y^\delta,b} = \partial^2 \left(\frac{1}{2}\|\cdot - y^\delta\|^2\right) = 1,$$

that was explicitly used in [47, Proposition 3.1].

For problem ($\tilde{\mathcal{D}}_{KL} - |\cdot|$) the fidelity $f_{y^\delta,b}(\Phi\mu) = \tilde{\mathcal{D}}_{KL}(\Phi\mu + b, y)$ is twice differentiable with

$$\partial^2 f_{y^\delta,b} = \partial \tilde{\mathcal{D}}_{KL}(\cdot + b, y) = \frac{y^\delta}{\|\cdot + b\|^2}.$$

Remark 3. *The proof of Proposition 2 holds for any strictly decreasing updating rule. However, using an updating rule different from (36) there is no guarantee that at iteration $t+1$ of Algorithm 2 the measure $\mu_{t+1}^{[0]} = \hat{\mu}_t$ is not already optimal for $\mathcal{P}(\lambda_{t+1})$, thus requiring an immediate update of λ.*

5 Numerical tests

In this section, we report numerical results to ($\tilde{\mathcal{D}}_{KL} - |\cdot|$) computed by means of the Sliding Frank-Wolfe (SFW) Algorithm 1 with homotopy (Algorithm 2) for several off-the-grid sparse deconvolution problems in simulated 1D/2D and real 3D fluorescence microscopy data.

5.1 1D comparisons between Gaussian and Poisson modelling

The aim of this first set of experiments on simulated 1D blurred signals corrupted with Poisson noise is to validate our model ($\tilde{\mathcal{D}}_{KL} - |\cdot|$) and to compare its performance with BLASSO ($L^2 - |\cdot|$). In a discrete setting, several works (e.g., [17, 50, 51]

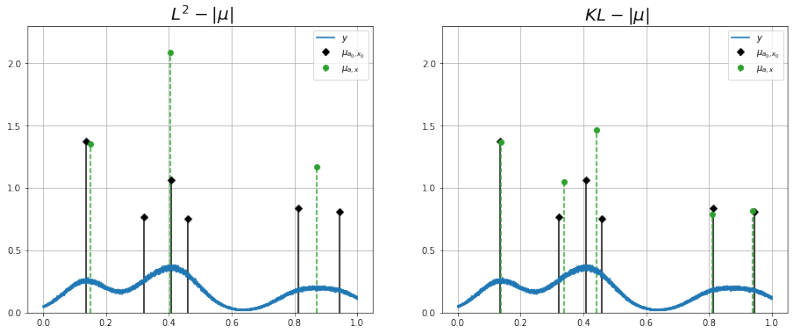


Figure 4: 1D comparison between Gaussian (left) and Poisson (right) models. In black: ground truth spikes. In green: reconstructed spikes. For both models, $\lambda = 8.82$.

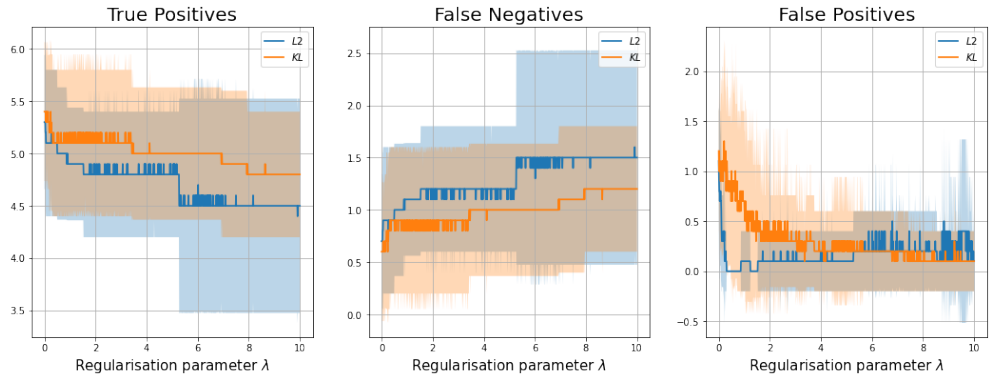
and [52] for examples in microscopy) have proposed both analytical and numerical approaches for precisely modelling signal-dependent Poisson noise using discrete version of the KL divergence (14). Generally speaking, such choices improves performance (in terms of, e.g., localisation/reconstruction quality, especially in low-photon count regimes) in comparison to simpler L^2 Gaussian models, but only slightly. This is due to the biases introduced by the handcrafted regularisation employed, such as, e.g., ℓ_1 or TV-type. Aiming at reconstructing weighted sums of Diracs, for which the regulariser (2) is tailored, we wonder in the following experiments whether in such infinite-dimensional setting the improvement is more evident.

For that, we test 10 different ground truth signals with 6 randomly located spikes in the 1D domain $\Omega = [0, 1]$. For each ground truth signal, the position of each spike is sampled from a uniform distribution over Ω, as well as their amplitudes from a uniform distribution over $[1-d, 1+d]$ with $d = 0.4$. The corresponding acquisitions are blurred by a Gaussian 1D PSF with $\sigma = 0.07$, a spatially constant background $b = 0.01$ is considered, and then several Poisson noise realisations are generated as acquired data. In Figure 4, one simulated ground truth signal μ_{gt} is shown (black spikes) together the corresponding Poisson noisy and blurred data (blue signal). All 1D signals are then reconstructed by using both the Poisson ($\widetilde{\mathcal{D}}_{KL} - |\cdot|$) and the Gaussian noise model BLASSO ($L^2 - |\cdot|$) with $\lambda \in (0, 10]$ using Algorithm 1. Figure 4 shows an example between the two reconstructions (BLASSO, left, Poisson model, right) μ_{rec} (green spikes) for $\lambda = 8.82$. For such illustrative example, the Poisson model provides a better estimate (in terms both of amplitudes and localisation) than the Gaussian model.

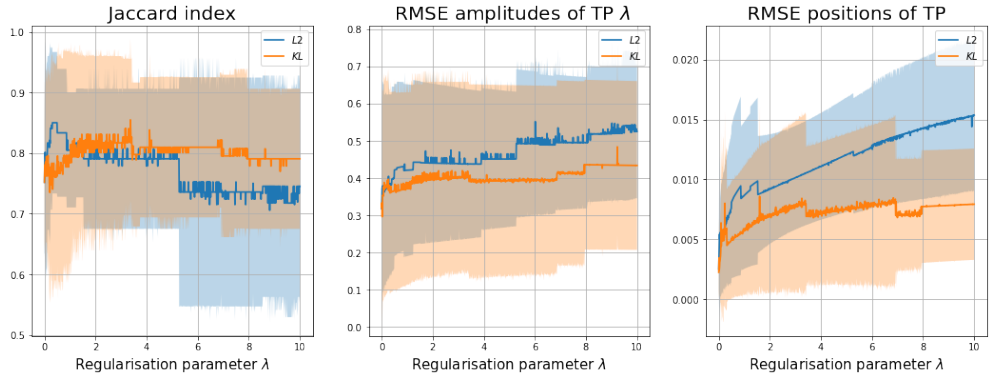
To assess such observation over different choices of regularisation parameters (to which the models could be sensitive), we then performed a statistical test comparing localisation/reconstruction performance for all the generated signals. To evaluate the goodness of the reconstructions, we consider the Jaccard index defined in terms of the number of True Positive (TP), False Positive (FP) and False Negative (FN) spikes as follows

$$\text{Jac}_\delta(\mu_{\text{gt}}, \mu_{\text{rec}}) = \frac{\#\text{TP}}{\#\text{TP} + \#\text{FP} + \#\text{FN}} \in [0, 1]$$

with tolerance radius $\delta > 0$. TP are reconstructed spikes that are at a distance less than δ from a ground truth spile, while reconstructed spikes that are more than δ distant from each ground truth spike are denoted by FP. FN are spikes in the ground truth which have not been associated to any TP.



(a) TP, FN and TP



(b) Jaccard index and RMSE of amplitudes and positions

Figure 5: Mean values over 100 different randomly generated ground truth signals with 6 spikes and their corresponding reconstructions. Shaded area corresponds to standard deviation. Maximum number of iterations of SFW: $2N_{\text{spikes}}$. Tolerance radius for computation of the Jaccard Index is $\delta = 0.05$.

Similarly as in [12], we used also the RMSE of the amplitudes a and positions x of the TP spikes as a different quality metric:

$$\text{RMSE}_x(\mu_{\text{gt}}, \mu_{\text{rec}}) = \sqrt{\frac{1}{\#\text{TP}} \sum_{i \in \text{TP}} \left((x_{\text{rec}})_i - (x_{gt})_i\right)^2}$$

Parameters	$L^2 - \|\mu\|$	$\tilde{\mathcal{D}}_{KL} - \|\mu\|$
Max. number of homotopy iterations	$2N_{\text{molecules}}$	$2N_{\text{molecules}}$
Max. number of inner SFW iterations	1	1
Homotopy parameter c	15	40
Homotopy parameter γ	0.9	0.9
Choice of σ_{target}	$1.5 \times \frac{1}{2}\|\Phi\mu_{\text{gt}} + b - y\|^2$	$1.5 \times \mathcal{D}_{KL}(\Phi\mu_{\text{gt}} + b, y)$

Table 2: Parameters used by the homotopy algorithm (Alg.2) in the 1D simulated comparison tests.

$$\text{RMSE}_a(\mu_{\text{gt}}, \mu_{\text{rec}}) = \sqrt{\frac{1}{\#\text{TP}} \sum_{i \in \text{TP}} \big((a_{\text{rec}})_i - (a_{gt})_i\big)^2}.$$

In Figure 5(a), we plot the number of TP, FN and FP reconstructed by models ($L^2 - |\cdot|$) and ($\tilde{\mathcal{D}}_{KL} - |\cdot|$) for a large finely discretised range of λ. In Figure 5(b) the Jaccard index (computed with $\delta = 0.05$) and the Root Mean Square Error (RMSE) of amplitudes and positions are also reported. The proposed Poisson model ($\tilde{\mathcal{D}}_{KL} - |\cdot|$) has a better performance in terms of TP and FN and, overall, in terms of the Jaccard index and RMSE of amplitudes and positions. Only for the number of FP, BLASSO ($L^2 - |\cdot|$) results slightly better than ($\tilde{\mathcal{D}}_{KL} - |\cdot|$) for small values of λ. This is due to the fact that ($\tilde{\mathcal{D}}_{KL} - |\cdot|$) usually requires more iterations of SFW before reaching convergence. This results in a better estimation of the number of molecules, with TP being closer to the actual number of spikes, which may cause an overestimation of the number of spikes with a consequently higher value of FP.

Using the same dataset, we then compare the results obtained running the homotopy Algorithm 2 for the automatic selection of the regularisation parameter λ for both problems ($L^2 - |\cdot|$) and ($\tilde{\mathcal{D}}_{KL} - |\cdot|$) solved by Algorithm 1 as inner solver, with the algorithmic parameters specified in Table 2. To reduce the computational burdens, we observed that only one iteration of SFW was enough, as the estimated measures is anyway updated in the next homotopy step. For the same reason, we set the maximum number of homotopy outer iterations to be equal to twice the number of peaks in the ground truth. As far as σ_{target} is concerned, being in a simulated environment, we computed the exact value of the residual σ_{exact} as $f_{y^\delta,b}(\Phi\mu_{\text{gt}})$, to be compared with the residual σ_t at current iteration t of Algorithm 2. Since σ_{exact} is unknown in real situations, a possible strategy for its estimation will be discussed in the next section.

In Table 3, we report the values of TP,FN,FP, Jaccard index and RMSE of the reconstructed signals for both models. The final estimated λ is also reported in Table 3, together with the number of performed homotopy iterations and of the value σ_{target}. Using homotopy, we observe that we retrieve values which are comparable with the best ones obtained using SFW with grid search. This shows the effectiveness of the homotopy strategy. Overall, the algorithm applied to solve ($\tilde{\mathcal{D}}_{KL} - |\cdot|$) yields better results than ($L^2 - |\cdot|$) in the presence of Poisson noise, with a reduction of the number of FN and an improvement of the accuracy in terms of Jaccard index.

	$L^2 - \|\mu\|$	$\tilde{\mathcal{D}}_{KL} - \|\mu\|$
Jaccard index	0.74	**0.76**
Number of TP	4.50	**4.80**
Number of FN	1.50	**1.20**
Number of FP	**0.10**	0.40
RMSE on amplitudes of TP	0.41	0.44
RMSE on positions of TP	0.014	0.015
Final estimated λ	6.09	40.21
Number of homotopy iterations	4.55	3.93
Value of σ_{target}	4.09	77.16

Table 3: Homotopy algorithm: comparison between BLASSO and the Poisson off-the-grid modelling. Mean values over 100 different randomly generated ground truths with 6 spikes.

5.2 2D examples: choice of σ_{target}

To avoid the choice of σ_{target} to depend on the ground truth image, we propose in this section an heuristic strategy to estimate a reasonable value σ_{target} relying on the sole acquisition y^δ and on the assumption that the signal is sparse. To better illustrate the proposed strategy we consider a 2D example of simulated blurred and noisy microscopy acquisitions on the domain $\Omega = [0,1]^2$ [1], see Fig. 6a.

Under a suitable sparsity level of the ground truth image, it is safe to assume that its corresponding noisy and blurred acquisition y^δ presents regions containing only background noise, which we denote by $\Omega_{\text{bg}} \subset \Omega$. In Figure 6a we show y^δ and highlight $y^\delta|_{\Omega_{\text{bg}}}$ in transparency, i.e. the acquisition "masked" to the area of background noise in the external square-ring. We propose to estimate the value of σ_{target} as follows

$$\sigma_{\text{target}} = f_{y^\delta,b}(0)\big|_{\Omega_{\text{bg}}} \frac{|\Omega|}{|\Omega_{\text{bg}}|}, \tag{39}$$

where the restriction of the fidelity term to $\mu = 0$ is due to the fact we assume the desired image μ to be null in Ω_{bg}, i.e. $\mu|_{\Omega_{\text{bg}}} = 0$. Note that considering $\sigma_{\text{target}} = f_{y^\delta,b}(0)\big|_{\Omega_{\text{bg}}}$ would be equivalent to assume that the noise is null on $\Omega \setminus \Omega_{\text{bg}}$, which is obviously not true. The formula (39) is thus adjusted to account for noise not only on Ω_{bg} but on the whole domain Ω. Note that it is also possible to approximately estimate the (constant) background $b \in L^2(\Omega)$ in Ω_{bg} by taking

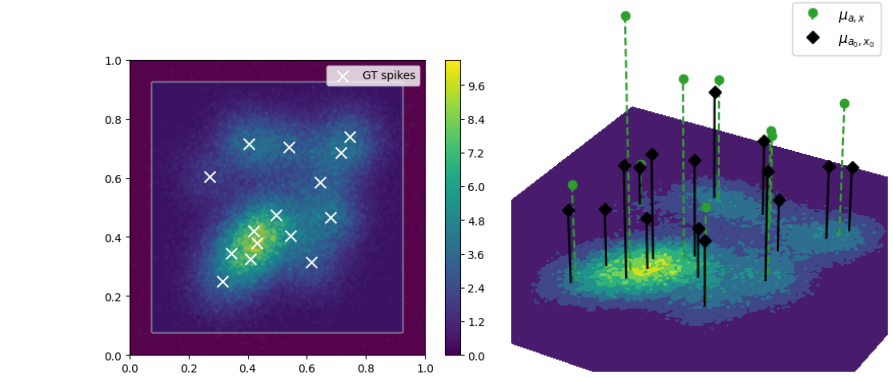
$$b = \frac{1}{|\Omega_{\text{bg}}|} \int_{\Omega_{\text{bg}}} y^\delta(t) \mathrm{d}t. \tag{40}$$

For the image shown in Figure 6a, we compare in Table 4 different choices of σ_{target} for $(\tilde{\mathcal{D}}_{KL} - |\cdot|)$ with $f_{y^\delta,b}$ being the Kullback-Leibler fidelity term. In particular, we consider in the second row the estimate proposed in [30], where a discrepancy principle

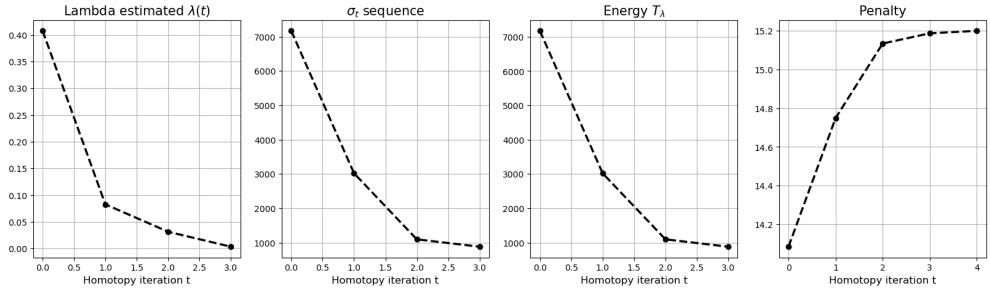
[1] Blurring operator: 2D Gaussian PSF with $\sigma = 0.07$. Spatially constant background $b = 0.05$.

Theoretical value (based on the ground truth)	$\tilde{\mathcal{D}}_{KL}(\Phi\mu_{gt} + b, y^\delta)$	842.3
Poisson discrepancy principle (41) [30]	$\frac{\lvert\Omega\rvert}{2}$	8192
Estimation based only on y^δ and Ω_{bg} (39)	$\tilde{\mathcal{D}}_{KL}(b, y_{bg})\frac{\lvert\Omega\rvert}{\lvert\Omega_{\text{bg}}\rvert}$	846.4

Table 4: Different estimates of σ_{target}.



(a) Data (b) Reconstruction



(c) Values of λ_t, σ_{target} and cost functional along the homotopy iterations.

Figure 6: 2D fluorescence microscopy simulated image. (a) 2D sparse image (white crosses denote off-the-grid locations) and its corresponding noisy blurred acquisition y^δ with $y^\delta|_{\Omega_{\text{bg}}}$ visualised in trasparency. (b) 2D reconstruction (green spikes) obtained with homotopy algorithm (Alg. 2) compared with the ground truth spikes (black).

for Poisson data is studied under the following approximation

$$\tilde{\mathcal{D}}_{KL}(\Phi\mu_{gt} + b, y^\delta) \approx \frac{|\Omega|}{2}. \tag{41}$$

This value is obtained by computing the expected value for Kullback-Leibler fidelity and by approximating it with a first order Taylor expansion. As observed also in [30],

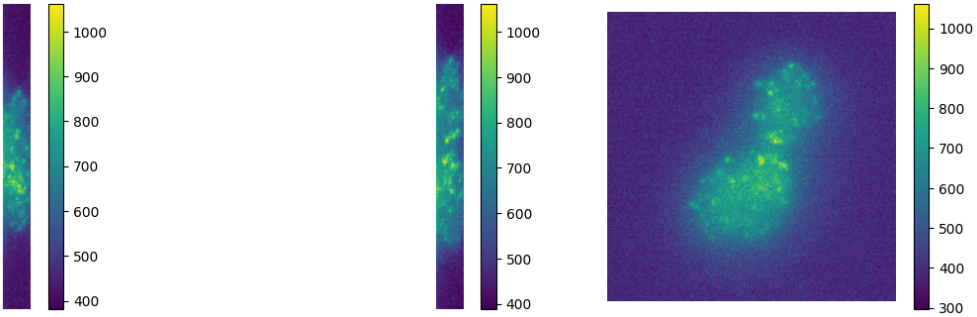


Figure 7: ERES 3D real data (I and II columns: lateral views, III column: top view).

the estimate (41) might not be optimal and, indeed, one should consider

$$\tilde{\mathcal{D}}_{KL}(\Phi\mu_{gt} + b, y^{\delta}) \approx \frac{1-\epsilon}{2}|\Omega|, \tag{42}$$

where $\epsilon \in \mathbb{R}$ is small. When Ω is big, (42) might lead to bad estimates even if ϵ is very small. As showed in Table 4, the approximation (41) does not lead indeed to an accurate estimation of σ_{target}. On the contrary, the estimation (39) proposed (last row), is close to the real value (first row) which is known given the simulated setting. We remark that in recent work [53] similar masking strategies are used to define parameter selection strategies for variational noise in Poisson scenarios, with a detailed description of the modifications arising to the underlying statistical laws when performing such choice. The application of analogous strategies to the problem considered is an interesting venue for future work.

By using the homotopy algorithm (Alg.2) with SFW (Alg. 1) [2] for the reconstruction of the data in Figure 6a with background and σ_{target} estimated by (40) and (39), respectively, we obtain the results shown in Figure 6.

5.3 Real 3D wide-field dataset

We finally consider a real 3D fluorescence microscopy setting with blurred and noisy volume acquisitions. The image was acquired in wide-field microscopy with a 100x/1.49 objective, a Hamamatsu Flash 4 camera with 6.5um pixels, by Alejandro Melero Carrillo at the MRC-LMB and was used in [54]. It is an acquisition of yeasts expressing fluorescent proteins (*SEC16-sfGFP* and a *SEC24-sfGFP*) localised at the Endoplasmic Reticulum exit sites (ERES). The acquired volume y^{δ} is shown in Figure 7 with maximum intensities projections over the xz, yz, yx planes. The 3D volume blurred and noisy acquisition has $190 \times 190 \times 17$ voxels with voxel size of 65nm in yz and 250nm in z. Signal dependency of the noise is observed.

To reconstruct a sparse volume from the 3D acquisition y^{δ}, we use the homotopy algorithm (Alg. 2) with an accelerated version of the SFW, called Boosted SFW

[2] Max. number of (outer) homotopy it. 20, max. number of (inner) SFW it. 1, $c = 30$, $\gamma = 0.9$.

proposed in [47], as an inner solver to minimise $(\tilde{\mathcal{D}}_{KL} - |\cdot|)$. Boosted SFW reduces computational costs by limiting the number of sliding steps.

Parameters	$(\tilde{\mathcal{D}}_{KL} - \|\cdot\|)$
Max. number of homotopy iterations	10
Max. number of inner SFW iterations	50
Homotopy parameter c	0.5
σ_{target} given by (39)	1102067.75
σ_{target} given by (41)	306850
Constant background estimate (40)	b = 337.77

Table 5: Parameters used in Algorithm 2 for the reconstruction of the 3D volume

We consider a 3D convolution kernel (6) estimated as a 3D Gaussian PSF,

$$\varphi\big((x,y,z)\big) = \frac{1}{\sqrt{(2\pi)^3}\sigma_x\sigma_y\sigma_z}\exp\left[-\frac{x^2}{2\sigma_x^2}\right]\exp\left[-\frac{y^2}{2\sigma_y^2}\right]\exp\left[-\frac{z^2}{2\sigma_z^2}\right].$$

The standard deviation σ_x, σ_y of the 3D PSF can be estimated from the Full Width Half Maximum, which is given by $\text{FWHM} = 0.61\lambda_{\text{wavelength}}/\text{NA}$, where $\lambda_{\text{wavelength}}$ is the emission wavelength of the fluorescent proteins and NA is the numerical aperture of the microscope. Note that if the FWHM is known, then it is possible to retrieve information about the variance parameters of the PSF since $\text{FWHM} = 2.355 \cdot \sigma_x$ and $\text{FWHM} = 2.355 \cdot \sigma_y$. For the standard deviation in the z-axis, we assume $\sigma_z = 2 \cdot \sigma_x$. Since the value of the numerical aperture is known, $\text{NA} = 1.49$, and $\lambda_{\text{wavelength}} = 508\text{nm}$ for the green fluorescent proteins under test, we obtain a PSF estimation with $\sigma_x = \sigma_y = 89\text{nm}$ and $\sigma_z = 178\text{nm}$, which appears to be a good approximation of the underlying blur model.

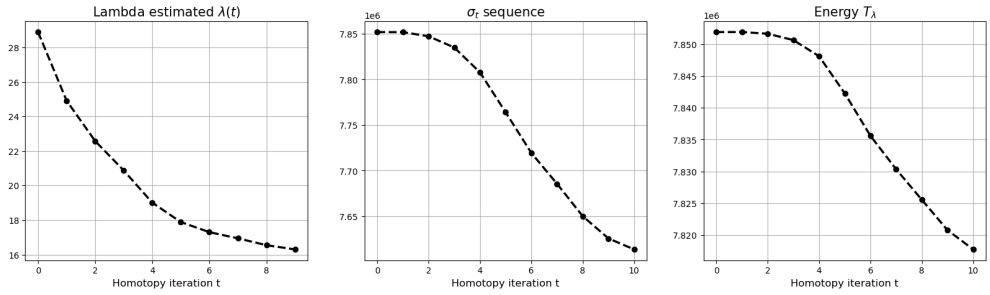


Figure 8: ERES 3D data. Values of λ_t, σ_{target} and cost functional along the homotopy iterations.

The parameters used to run the homotopy algorithm 2 with BSFW as inner solver are reported in Table 5. Note that as far as the estimate of σ_{target} is concerned, the

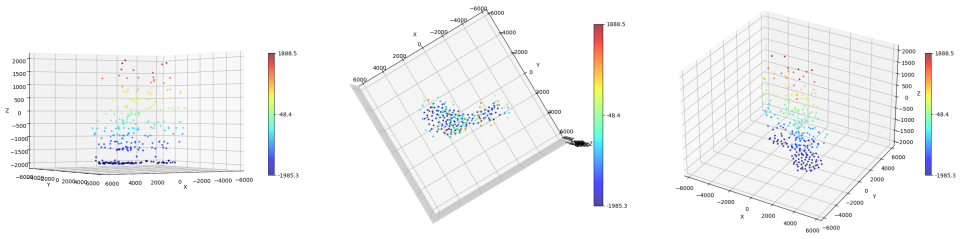


Figure 9: Sparse reconstruction of the 3D real ERES data

formulas (41) and (39) give very different results, so that the estimate given by (39) was considered as more accurate as showed in the previous section. The background is estimated by (40) as $b = 337.77$. We computed this reconstruction using 50 homotopy iterations fixing to 10 the maximum number of inner loops of the BSFW algorithm, using Google Colab CPUs for about 10 hours. The reconstructed volume μ_{rec} counted 274 spikes. While probably underestimating the exact result this first result is promising, since the use of Algorithm 2 yields very precise localisation of spikes automatically, with no need of estimating the regularisation parameter, and no a-priori information about the data. A better visualisation of the reconstruction under different views is given in Figure 9 using the visualisation codes used in [12] and provided by the authors at their GitHub page[3].

6 Conclusions

In this work, we considered sparse inverse problems in the off-the-grid setting of the space of radon measures $\mathcal{M}(\Omega)$ under the assumption of signal-dependent Poisson noise in the measurement. Such choice is motivated by fluorescence microscopy applications, where the noise observed is modelled as Poisson to account for photon-counting processes. We proposed a variational approach where the Total Variation norm is coupled with the Kullback-Leibler fidelity term and a non-negativity constraint. For the resulting composite problem we derive analytically the optimality and extremality conditions. Using the Sliding Frank-Wolfe algorithm as numerical solver, we then discussed the choice of a good regularisation parameter by means of an algorithmic homotopy strategy. Finally, we presented several numerical experiments on simulated 1D/2D data to validate the proposed Poisson model, to compare its performance with the analog Gaussian model and to confirm the good performance of the homotopy strategy proposed. To conclude, we tested the proposed framework on a 3D real fluorescence microscopy dataset.

Acknowledgements. We thank B. Laville and L. Blanc-Féraud (Morpheme, Inria), and R. Petit (MaLGa, University of Genoa) for thoughtful discussions and for providing useful codes of the Sliding Frank-Wolfe algorithm for the resolution of 1D and 2D BLASSO results. We thank J. Boulanger (MRC-LMB, Cambrisge) for making

[3] https://github.com/qdenoyelle/sfw4blasso

this work possible by having put ML, LC and CE in touch with AM, who acquired the 3D data and for thoughtful discussions.

Funding. LC and ML acknowledge the support received by the projects ANR MICROBLIND (ANR-21-CE48-0008) and ANR JCJC TASKABILE (ANR-22-CE48-0010). CE acknowledges the support received by the MUR Excellence Department Project awarded to Dipartimento di Matematica, Università di Genova, CUP D33C23001110001. CE and ML acknowledge the support of the "PRIN 2022ANC8HL - Inverse Problems in the Imaging Sciences (IPIS)" project, granted by the Italian Ministero dell'Università e della Ricerca within the framework of the European Union - Next Generation EU program and of the Italian INdAM group on scientific calculus GNCS.

Data availability. The codes used for implementing the models and algorithms described in this work are available at https://github.com/martalazzaretti/KL-TV-off-the-grid.

Declarations

Conflict of interest. The authors declare that they have no conflict of interest.

Appendix A Technical Results

We report here the well-known definition of convex conjugate and some of its properties.

Definition 3. *Given $f : \mathcal{X} \longrightarrow \mathbb{R}$, with $\mathcal{X}$ real Banach space, the convex conjugate of f is the function $f^* : \mathcal{X}^* \longrightarrow \mathbb{R} \cup \{+\infty\}$ defined by*

$$f^*(x^*) := \sup_{x \in \mathcal{X}} \langle x^*, x \rangle_{\mathcal{X}^* \times \mathcal{X}} - f(x), \quad x^* \in \mathcal{X}^*. \tag{A1}$$

It is possible to obtain the convex conjugate of the sum of functions as infimal convolution of the convex conjugate of each singular function [36, 37].

Proposition 3. *Let $f_1, \dots, f_n : \mathcal{X} \to \mathbb{R}$ with $\mathcal{X}$ real Banach space. Suppose there exists at least a point $x \in \mathcal{X}$ such that $f_1, \dots, f_n$ are continuous in x. Then,*

$$(f_1 + \ldots + f_n)^* (x^*) = \min_{x_1^* + \ldots + x_n^* = x^*} f_1^*(x_1^*) + \ldots + f_n^*(x_n^*).$$

The convex conjugate of any norm is a well-known result. We report here for the sake of completeness, the convex conjugate of the TV norm.

Lemma 5. *Let $A : \mathcal{M}(\Omega) \to \mathbb{R}$, $A(\cdot) = |\cdot|(\Omega)$ be the TV-norm. The convex conjugate of A is the function $A^* : \mathcal{C}_0(\Omega, \mathbb{R}) \to \mathbb{R} \cup \{+\infty\}$ given by*

$$A^*(\psi) = \begin{cases} 0 & \|\psi\|_{\infty,\Omega} \leq 1 \\ +\infty & \|\psi\|_{\infty,\Omega} > 1 \end{cases}.$$

Proof. By Definition 3 of convex conjugate, for any $\psi \in \mathcal{C}_0(\Omega, \mathbb{R})$ we write

$$A^*(\psi) = \sup_{\mu \in \mathcal{M}(\Omega)} \langle \psi, \mu \rangle_{\mathcal{C}_0(\Omega,\mathbb{R})\times\mathcal{M}(\Omega)} - |\mu|(\Omega) \geq \langle \psi, \mu \rangle_{\mathcal{C}_0(\Omega,\mathbb{R})\times\mathcal{M}(\Omega)} - |\mu|(\Omega), \quad \forall \mu \in \mathcal{M}(\Omega).$$

Taking $\mu = \alpha\delta_x$ with $x \in \Omega$ and $\alpha > 0$, we have

$$\sup_{\mu \in \mathcal{M}(\Omega)} \langle \psi, \mu \rangle_{\mathcal{C}_0(\Omega,\mathbb{R})\times\mathcal{M}(\Omega)} - |\mu|(\Omega) \geq \alpha(\psi(x) - 1),$$

and the limit for $\alpha \to \infty$ of the latter inequality yields $A^*(\psi) \geq +\infty$ if $\psi(x) > 1$. A similar result for $\psi(x) < 1$ is obtained with the measure $\mu = -\alpha\delta_x$, with $\alpha > 0$. Thus, we have $A^*(\psi) = +\infty$ if $\|\psi\|_{\infty,\Omega} > 1$.

Assume now that $\|\psi\|_{\infty,\Omega} \leq 1$. First notice that

$$A^*(\psi) = \sup_{\mu \in \mathcal{M}(\Omega)} \langle \psi, \mu \rangle_{\mathcal{C}_0(\Omega,\mathbb{R})\times\mathcal{M}(\Omega)} - |\mu|(\Omega) \geq 0$$

by considering $\mu = 0$. Moreover,

$$\langle \psi, \mu \rangle_{\mathcal{C}_0(\Omega,\mathbb{R})\times\mathcal{M}(\Omega)} - |\mu|(\Omega) \leq \|\psi\|_{\infty,\Omega}|\mu|(\Omega) - |\mu|(\Omega) \leq |\mu|(\Omega)\left(\|\psi\|_{\infty,\Omega} - 1\right) \leq 0$$

since $\|\psi\|_{\infty,\Omega} \leq 1$. By evaluating the sup on both sides of the last inequality, since $A(\psi) \geq 0$ one finally gets $A^*(\psi) = 0$ if $\|\psi\|_{\infty,\Omega} \leq 1$, which concludes the proof. □

Lemma 6. (Fenchel-Rockafellar duality.) *Let V, Y be two Banach spaces. For $\Lambda : V \to Y$ linear operator, $F : V \to \mathbb{R}$ and $G : Y \to \mathbb{R}$ convex functionals, we consider the following primal problem:*

$$\operatorname*{argmin}_{u \in V} \; F(u) + G(\Lambda u). \tag{A2}$$

The corresponding dual problem reads

$$\operatorname*{argmax}_{p^* \in Y^*} -F^*(\Lambda^* p^*) - G^*(-p^*), \tag{A3}$$

where $\Lambda^* : Y^* \to V^*$ is the adjoint operator of Λ and $F^* : V^* \to \mathbb{R} \cup \{+\infty\}$, $G^* : Y^* \to \mathbb{R} \cup \{+\infty\}$ are the convex conjugate of F and G.

Moreover, if $u \in V$ and $p^* \in Y^*$ are respectively solutions of the primal (A2) and dual (A3) problems, the following extremality conditions hold:

$$\begin{cases} \Lambda^* p^* \in \partial F(u) \\ -p^* \in \partial G(\Lambda u) \end{cases}. \tag{A4}$$

References

[1] Duval, V., Peyré, G.: Sparse spikes super-resolution on thin grids II: the continuous basis pursuit. Inverse Probl. **33**(9), 095008 (2017)

[2] Duval, V., Peyré, G.: Sparse regularization on thin grids I: the Lasso. Inverse Probl. **33**(5), 055008 (2017)

[3] Laville, B., Blanc-Féraud, L., Aubert, G.: Off-the-grid variational sparse spike recovery: Methods and algorithms. J. Imaging. **7**(12) (2021)

[4] Scherzer, O., Walch, B.: Sparsity regularization for radon measures. In: Proc.-SSVM 2009, pp. 452–463. Springer, Berlin Heidelberg (2009)

[5] Bredies, K., Pikkarainen, H.K.: Inverse problems in spaces of measures. ESAIM: COCV. (2013)

[6] De Castro, Y., Gamboa, F.: Exact reconstruction using Beurling minimal extrapolation. J. Math. Anal. Appl. (2012)

[7] Fernandez-Granda, C.: Support detection in super-resolution. arXiv preprint 1302.3921 (2013)

[8] Duval, V., Peyré, G.: Exact support recovery for sparse spikes deconvolution. Found. Comput. Math. **15**, 1315–1355 (2013)

[9] Denoyelle, Q., Duval, V., Peyré, G.: Support recovery for sparse super-resolution of positive measures. J. Fourier Anal. Appl. (2017)

[10] Boyer, C., De Castro, Y., Salmon, J.: Adapting to unknown noise level in sparse deconvolution. Inf. Inference J. IMA (2017)

[11] Poon, C., Peyré, G.: Multi-dimensional sparse super-resolution. SIAM J. Math. Anal. (2019)

[12] Denoyelle, Q., Duval, V., Peyré, G., Soubies, E.: The sliding frank–wolfe algorithm and its application to super-resolution microscopy. Inverse Probl. **36**(1), 014001 (2019)

[13] Flinth, A., Gournay, F., Weiss, P.: On the linear convergence rates of exchange and continuous methods for total variation minimization. Mathematical Programming **190**(1), 221–257 (2021)

[14] Bénard, P.-J., Traonmilin, Y., Aujol, J.-F., Soubies, E.: Estimation of off-the grid sparse spikes with over-parametrized projected gradient descent: theory and application. Inverse Problems (2024)

[15] Dossal, C., Duval, V., Poon, C.: Sampling the fourier transform along radial lines.

SIAM J. Numer. Anal. **55**(6), 2540–2564 (2017)

[16] De Castro, Y., Gadat, S., C., M., Maugis-Rabusseau, C.: SuperMix: Sparse regularization for mixtures. Ann. Statist. **49**(3), 1779–1809 (2021)

[17] Bertero, M., Boccacci, P., Ruggiero, V.: Inverse Imaging with Poisson Data. 2053-2563. IOP Publishing, UK (2018)

[18] Poon, C.: An Introduction to Sparse Spikes Recovery via the BLASSO. Lecture notes (2019)

[19] Rudin, W.: Real and Complex Analysis, 3rd Ed. McGraw-Hill, Inc., USA (1987)

[20] Cohn, D.L.: Measure Theory. Springer, New York (2013)

[21] Chizat, L., Bach, F.: On the global convergence of gradient descent for over-parameterized models using optimal transport. In: Proc. 32nd Int. Conf. Neural Inf. Process. Syst., NIPS'18, pp. 3040–3050 (2018)

[22] Koulouri, A., Heins, P., Burger, M.: Adaptive superresolution in deconvolution of sparse peaks. IEEE Trans. Signal Process. **69**, 165–178 (2021)

[23] Laville, B., Blanc-Féraud, L., Aubert, G.: Off-the-grid curve reconstruction through divergence regularization: An extreme point result. SIAM J. Imaging Sci. **16**(2), 867–885 (2023)

[24] Laville, B., Blanc-Féraud, L., Aubert, G.: Off-the-grid charge algorithm for curve reconstruction in inverse problems. In: Proc.-SSVM 2023, pp. 393–405. Springer, Cham (2023)

[25] De Castro, Y., Duval, V., Petit, R.: Towards off-the-grid algorithms for total variation regularized inverse problems. J. Math. Imaging Vis. **65**(1), 53–81 (2023)

[26] De Castro, Y., Duval, V., Petit, R.: Towards off-the-grid algorithms for total variation regularized inverse problems. In: Proceedings SSVM 2021, pp. 553–564. Springer, Cham (2021)

[27] De Castro, Y., Duval, V., Petit, R.: Exact recovery of the support of piecewise constant images via total variation regularization. arXiv preprint: 2307.03709 (2023)

[28] Beurling, A.: Sur les intégrales de fourier absolument convergentes et leur application à une transformation fonctionnelle. In: Proc. Ninth Scand. Math. Congr., Helsinki, Finland, pp. 345–366 (1938)

[29] Bertero, M., Boccacci, P., Desidera, G., Vicidomini, G.: Image deblurring with Poisson data: from cells to galaxies. Inverse Probl. **25**, 123006 (2009)

[30] Bertero, M., Boccacci, P., Talenti, G., Zanella, R., Zanni, L.: A discrepancy principle for Poisson data. Inverse Probl. **26**, 105004 (2010)

[31] Li, J., Shen, Z., Yin, R., Zhang, X.: A reweighted l^2 method for image restoration with Poisson and mixed Poisson-Gaussian noise. Inverse Probl. Imaging. **9**(3), 875–894 (2015)

[32] Sawatzky, A., Brune, C., Kösters, T., Wübbeling, F., Burger, M.: EM-TV Methods for Inverse Problems with Poisson Noise, pp. 71–142. Springer, Cham (2013)

[33] Lanza, A., Morigi, S., Sgallari, F., Wen, Y.-W.: Image restoration with Poisson–Gaussian mixed noise. "Comput. Methods Biomech. Biomed. Eng. Imaging Vis. **2**(1), 12–24 (2014)

[34] Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton, NJ, USA (1970)

[35] Ekeland, I., Témam, R.: Convex Analysis and Variational Problems. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA (1999)

[36] Hiriart-Urruty, J.-B.: A note on the legendre-fenchel transform of convex composite functions. In: Alart, P., Maisonneuve, O., Rockafellar, R.T. (eds.) Nonsmooth Mechanics and Analysis, pp. 35–46. Springer, Boston, MA (2006)

[37] Fajardo, M.D., Vicente-Pérez, J., Rodríguez, M.M.L.: Infimal convolution, c-subdifferentiability, and Fenchel duality in evenly convex optimization. Inverse Probl. **20**(2), 375–396 (2012)

[38] Ndiaye, E., Fercoq, O., Gramfort, A., Salmon, J.: Gap safe screening rules for sparsity enforcing penalties. J. Mach. Learn. Res. **18**(1), 4671–4703 (2017)

[39] Wang, J., Ye, J.: Two-layer feature reduction for sparse-group lasso via decomposition of convex sets. In: Proc. 27th Int. Conf. Neural Inf. Process. Syst. - Vol. 2, NIPS'14, pp. 2132–2140. MIT Press, Cambridge, MA, USA (2014)

[40] Dantas, C.F., Soubies, E., Févotte, C.: Safe screening for sparse regression with the Kullback-Leibler divergence. In: Proc. - ICASSP IEEE Int. Conf. Acoust., Toronto, Canada (2021)

[41] Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, (2011)

[42] Valkonen, T.: Proximal methods for point source localisation. J. Nonlinear Anal. Optim. **Volume 4** (2023)

[43] Frank, M., Wolfe, P.: An algorithm for quadratic programming. Nav. Res. Logist. Q. **3**(1-2), 95–110 (1956)

[44] Bredies, K., Lorenz, D.A., Maass, P.: A generalized conditional gradient method and its connection to an iterative shrinkage method. Computational Optimization and Applications **42**(2), 173–193 (2009)

[45] Chizat, L.: Sparse optimization on measures with over-parameterized gradient descent. Math. Program. **194**(1–2), 487–532 (2022)

[46] Pokutta, S.: The Frank-Wolfe Algorithm: A Short Introduction. Jahresbericht der Deutschen Mathematiker-Vereinigung **126**(1), 3–35 (2024)

[47] Courbot, J.-B., Colicchio, B.: A fast homotopy algorithm for gridless sparse recovery. Inverse Probl. **37**(2), 025002 (2021)

[48] Osborne, M., Presnell, B., Turlach, B.: A new approach to variable selection in least squares problems. IMA J. Numer. Anal. **20**(3), 389–403 (2000)

[49] Osborne, M.R., Presnell, B., Turlach, B.A.: On the LASSO and Its Dual. J. Comput. Graph. Stat. **9**(2), 319–337 (2000)

[50] Harmany, Z.T., Marcia, R.F., Willett, R.M.: This is SPIRAL-TAP: Sparse Poisson Intensity Reconstruction ALgorithms—Theory and Practice. IEEE Trans. Image Process. **21**(3), 1084–1096 (2012)

[51] Calatroni, L., De Los Reyes, J.C., Schönlieb, C.-B.: Infimal convolution of data discrepancies for mixed noise removal. SIAM J. Imaging Sci. **10**(3), 1196–1233 (2017)

[52] Reiser, N., Guo, M., Shroff, H., Riviere, P.J.L.: Phase Diverse Phase Retrieval for Microscopy: Comparison of Gaussian and Poisson Approaches. arXiv preprint 2308.00734 (2023)

[53] Bevilacqua, F., Lanza, A., Pragliola, M., Sgallari, F.: Masked unbiased principles for parameter selection in variational image restoration under poisson noise. Inverse Probl. **39**(3), 034002 (2023)

[54] Gomez-Navarro, N., Melero, A., Li, X.-H., Boulanger, J., Kukulski, W., Miller, E.A.: Cargo crowding contributes to sorting stringency in COPII vesicles. J. Cell. Biol. **219**(7), 201806038 (2020)