Off-the-grid regularisation for Poisson inverse problems

Marta Lazzaretti[1,2], Luca Calatroni[2], Claudio Estatico[1]

[1]DIMA, Università di Genova
[2]I3S Lab, CNRS, Université Côte d'Azur

Off-the-grid and continuous methods for optimization and inverse problems in imaging workshop

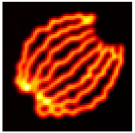
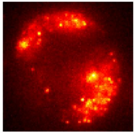
22 November 2023

Outline

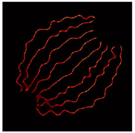
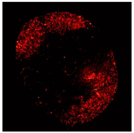
Introduction

Motivation: image reconstruction in biological imaging

GOAL: imaging of biological structures < 200 nm at fine scale and high precision

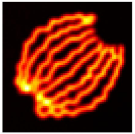



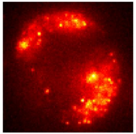
Acquisition: $\mathbf{y}$

Unknown signal: $\mathbf{x}$

Inverse problem formulation: discrete approach

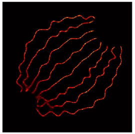


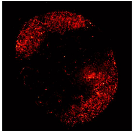


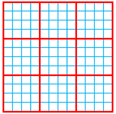
$$\mathbf{A} = \mathbf{R_L H} \in \mathbb{R}^{M \times N}$$

$$\mathbf{H} \in \mathbb{R}^{N \times N}, \ \mathbf{R_L} \in \mathbb{R}^{M \times N}$$

Acquisition: $\mathbf{y}$







Coarse and fine grid

Unknown signal: $\mathbf{x}$

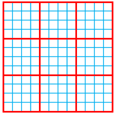
Inverse problem formulation: discrete approach

Acquisition: $\mathbf{y}$

$$\mathbf{A} = \mathbf{R_L H} \in \mathbb{R}^{M \times N}$$

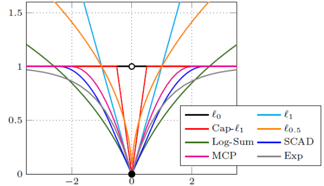
$$\mathbf{H} \in \mathbb{R}^{N \times N}, \ \mathbf{R_L} \in \mathbb{R}^{M \times N}$$

Unknown signal: $\mathbf{x}$



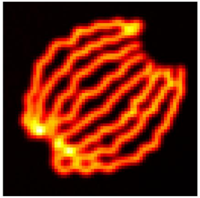
Coarse and fine grid

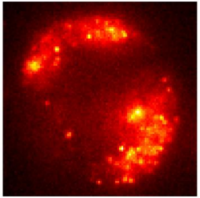
$$\underset{\mathbf{x} \in \mathbb{R}^N}{\arg\min} \ \frac{1}{2}\|\mathbf{Ax} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_0$$

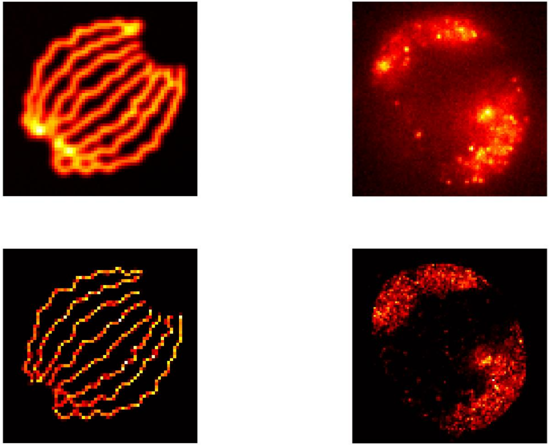


$$\underset{\mathbf{x} \in \mathbb{R}^N}{\arg\min} \ \frac{1}{2}\|\mathbf{Ax} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

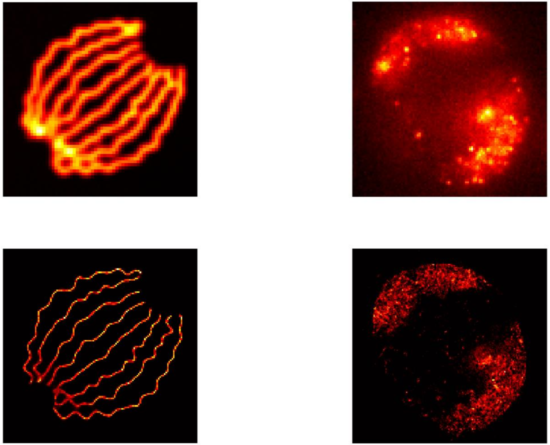
$$(\ell_2 - \ell_1)$$



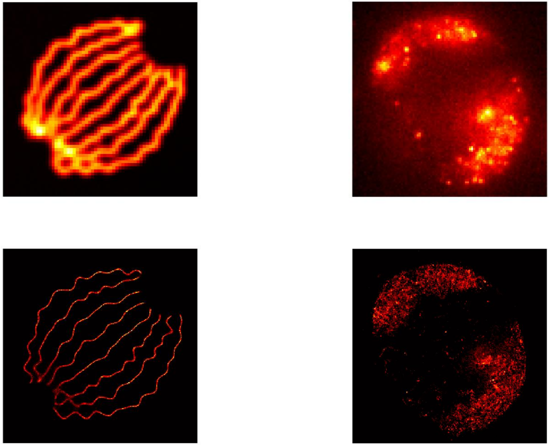




$L = 1$

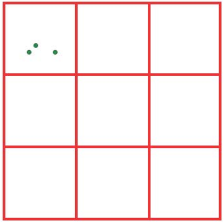


$L = 2$

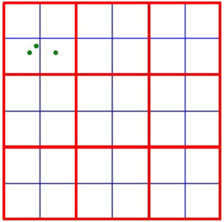


$$L = 4$$

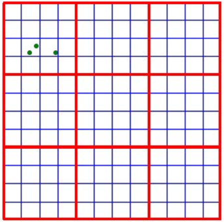
Going *off-the-grid* for spike deconvolution



L=1



L=2



L=4

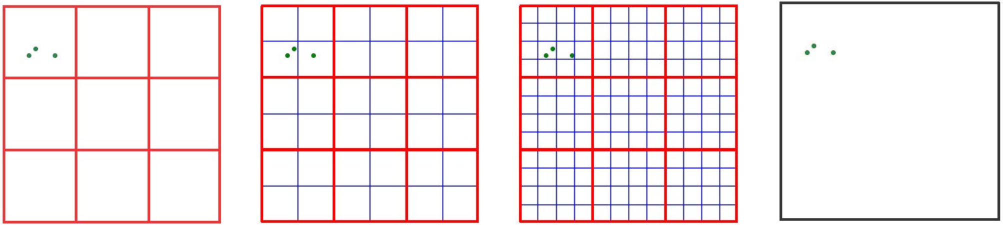
Coarser grid

Discretization errors
if N is small

Finer grid

Instabilities
if N is large

Going *off-the-grid* for spike deconvolution



L=1 | L=2 | L=4 | $L \longrightarrow +\infty$

Coarser grid | Finer grid | Grid-less

Discretization errors if N is small

Instabilities if N is large

Deconvolution of **point-like objects**, not on a fine grid, but on a **continuous domain** $\mathcal{X} \subseteq \mathbb{R}^d$

Off-the-grid approach

The space of Radon measures

- $\mathcal{X} \subseteq \mathbb{R}^d$ continuous and compact domain
- $\mathcal{M}(\mathcal{X})$ = **space of Radon measures** on $\mathcal{X}$

Point-like objects → **spikes**, modelled as **Dirac delta** $a\delta_x \in \mathcal{M}(\mathcal{X})$

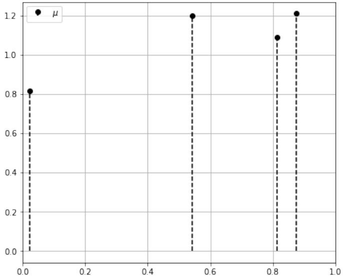
- $\mathcal{M}(\mathcal{X})$ = topological dual of $\mathscr{C}(\mathcal{X})$ with the supremum norm $\|\cdot\|_{\infty,\mathcal{X}}$.
- $\mu \in \mathcal{M}(\mathcal{X})$ is a continuous linear form evaluated on functions $f \in \mathscr{C}(\mathcal{X})$: $\langle f, m\rangle_{\mathscr{C}(\mathcal{X})\times\mathcal{M}(\mathcal{X})} = \int_{\mathcal{X}} f \,\mathrm{d}m$
- $\mathcal{M}(\mathcal{X})$ is a Banach space endowed with the TV-norm

$$|\mu|(\mathcal{X}) := \sup\left(\int_{\mathcal{X}} f \mathrm{d}\mu, f \in \mathscr{C}(\mathcal{X}), \|f\|_{\infty,\mathcal{X}} \leq 1\right)$$

$\mathcal{M}(\mathcal{X})$ **generalisation of** $\mathrm{L}^1(\mathcal{X})$: $\quad \mathrm{L}^1(\mathcal{X}) \hookrightarrow \mathcal{M}(\mathcal{X})$

If $\mu = \sum_{i=1}^N a_i\delta_{x_i}$ is a discrete measure then $|\mu|(\mathcal{X}) = \sum_{i=1}^N |a_i| = \|a\|_1$.

Off-the-grid framework

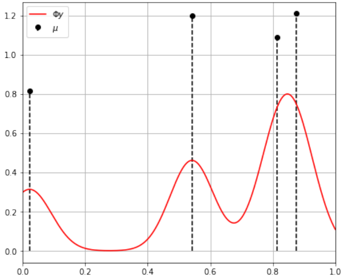


Point-like objects → **spikes**, modelled as **Dirac delta** $\delta_x \in \mathcal{M}(\mathcal{X})$ in the **space of Radon measures**

Unknown signal to reconstruct: finite linear combinations of Dirac deltas $\mu_{a,x} = \sum_{i=1}^{K} a_i \delta_{x_i}$

- $a_i > 0$ is the **amplitude** of the i-th spike
- $x_i \in \mathcal{X}$ represents its **position**
- K number of spikes

Off-the-grid framework

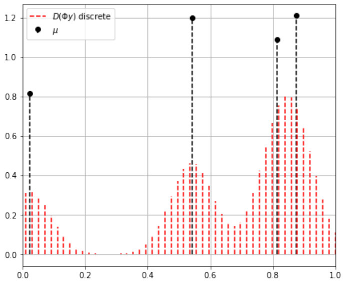


Forward operator: $\tilde{\Phi} : \mathcal{M}(\mathcal{X}) \longrightarrow L^2(\mathcal{X})$ given by

$$\tilde{\Phi}\mu = \int_{\mathcal{X}} \varphi(s) \mathrm{d}\mu(s)$$

with e.g. Gaussian PSF $\varphi(s) = \prod_{i=1}^{d} \frac{1}{\sqrt{(2\pi)}\sigma_i} \exp\left[-\frac{s_i^2}{2\sigma_i^2}\right]$

Off-the-grid framework



Blurring operator: $\tilde{\Phi} : \mathcal{M}(\mathcal{X}) \longrightarrow L^2(\mathcal{X})$ given by

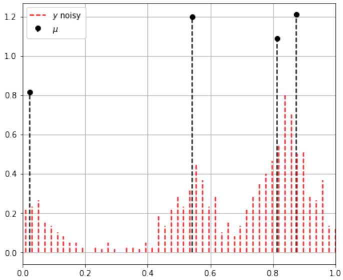
$$\tilde{\Phi}\mu = \int_{\mathcal{X}} \varphi(s)\mathrm{d}\mu(s)$$

with Gaussian PSF $\varphi(s) = \prod_{i=1}^{d} \frac{1}{\sqrt{(2\pi)}\sigma_i} \exp\left[-\frac{s_i^2}{2\sigma_i^2}\right]$

Downsampling operator $D : L^2(X) \longrightarrow \mathbb{R}^M$

$$\Phi = D\tilde{\Phi}$$

BLASSO PROBLEM



Gaussian noise:

$$y = \Phi\mu + \eta$$

$$\Downarrow$$

$$\underset{\mu \in \mathcal{M}(\mathcal{X})}{\arg\min} \frac{1}{2}\|\Phi\mu - y\|_2^2 + \lambda|\mu|(\mathcal{X}) \qquad (\ell_2 - |\mu|)$$

- Bredies-Pikkarainen 2012
- De Castro-Gamboa 2012
- Fernandez-Granda 2013
- Duval-Peyré 2014
- Denoyelle-Duval-Peyré 2017
- Boyer-De Castro-Salmon 2017
- Poon-Peyré 2019
- ...
- and more

Optimality conditions for BLASSO $\ell_2 - |\mu|$

$$\arg\min_{\mu \in \mathcal{M}(\mathcal{X})} \frac{1}{2}\|\Phi\mu - y\|_2^2 + \lambda|\mu|(\mathcal{X}) \qquad (\mathcal{P}_\lambda(y))$$

$$\Downarrow$$

$$\arg\max_{p \text{ s.t. } \|\Phi^* p\|_\infty \leq 1} \left\|\frac{y}{\lambda} - p\right\|^2 \qquad (\mathcal{D}_\lambda(y))$$

Extremality conditions:

μ_λ solution of $(\mathcal{P}_\lambda(y))$

p_λ solution of $(\mathcal{D}_\lambda(y))$

$$\begin{cases} -p_\lambda = -\frac{1}{\lambda}(\Phi\mu_\lambda - y) \\ \Phi^* p_\lambda \in \partial|\mu_\lambda|(\mathcal{X}) \end{cases}$$

Dual certificate:

$$\eta_\lambda(\mu) = \frac{1}{\lambda}\Phi^*\Big((y - \Phi\mu)\Big)$$

Optimality conditions for $(\mathcal{P}_\lambda(y))$:

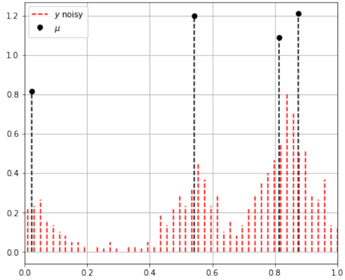
$\mu = \sum_{i=1}^{N} a_i \delta_{x_i}$ solution of $(\mathcal{P}_\lambda(y))$:

$$\Downarrow$$

$$\|\eta_\lambda(\mu)\|_\infty \leq 1$$

Off-the-grid for Poisson noise scenarios

$\mathcal{D}_{KL} - |\mu|$ off-the-grid variational problem for Poisson data



- $b \in \mathbb{R}^M_+$ strictly positive **background**
- **Poisson noise:**

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

- Hp on y: $y \in \mathbb{R}^M_+$ strictly positive
- Φ positive definite

$$\Downarrow$$

$$\underset{\mu \in \mathcal{M}(\mathcal{X})}{\arg\min}\, \mathcal{D}_{KL}(\Phi\mu + b, y) + \lambda|\mu|(\mathcal{X}) + \mathbb{1}_{\mathcal{M}(\mathcal{X})^+}(\mu) \qquad (\mathcal{D}_{KL} - |\mu|)$$

Dual problem of $\mathcal{D}_{KL} - |\mu|$

$$\underset{\mu \in \mathcal{M}(\mathcal{X})}{\arg\min} \underbrace{\frac{1}{\lambda}\mathcal{D}_{KL}(\Phi\mu + b, y)}_{G(\Phi\mu)} + \underbrace{|\mu|(\mathcal{X}) + \mathbb{1}_{\mathcal{M}(\mathcal{X})^+}(\mu)}_{F(\mu)} \qquad (\mathcal{D}_{KL} - |\mu|)$$

- $G : L^2(\mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$

$$G(s) := \begin{cases} \frac{1}{\lambda}\int_{\mathcal{X}} \Big(s + b\Big)(x) - y(x) + y(x)\log(y(x)) - y(x)\log\Big[\Big(s+b\Big)(x)\Big] \,\mathrm{d}x & s + b \in L^2(\mathcal{X})^+ \\ +\infty & s + b \notin L^2(\mathcal{X})^+ \end{cases}$$

- $F : \mathcal{M}(\mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$, $F(\cdot) = |\cdot|(\mathcal{X}) + \mathbb{1}_{\mathcal{M}(\mathcal{X})^+}(\cdot)$
- $\Phi : \mathcal{M}(\mathcal{X}) \to L^2(\mathcal{X})$

$$\underset{p^* \in \left(L^2(\mathcal{X})\right)^*}{\arg\max} \; -F^*(\Phi^* p^*) - G^*(-p^*) \qquad (\text{Dual of } \mathcal{D}_{KL} - |\mu|)$$

Convex conjugate of $\mathcal{D}_{KL}$

Given $t, b, \lambda > 0$, consider the one-dimensional Kullback-Leibler function, defined by

$$f_{t,b}(s) = \begin{cases} \frac{1}{\lambda}\left(s + b - t + t\log(t) - t\log(s+b)\right) & s + b > 0 \\ +\infty & s + b \leq 0 \end{cases}$$

$$f^*_{t,b}(s^*) = \sup_{s\in\mathbb{R}} ss^* - f_{t,b}(s) = \sup_{s+b>0} ss^* - \frac{1}{\lambda}(s + b - t + t\log(t) - t\log(s+b)) =$$
$$= \sup_{s+b>0} s\left(s^* - \frac{1}{\lambda}\right) - \frac{b}{\lambda} + \frac{t}{\lambda}\log(s+b) + \frac{t}{\lambda} - \frac{t}{\lambda}\log(t)$$

$$G^*(p^*) = \begin{cases} +\infty & p^* \geq \frac{1}{\lambda} \\ \frac{\mathbf{b}}{\lambda}(1 - \lambda p^*) - \frac{\mathbf{y}}{\lambda}\log(1 - \lambda p^*) & p^* < \frac{1}{\lambda} \end{cases}$$

Convex conjugate of the penalty

$$F(\cdot) = A(\cdot) + B(\cdot) \qquad F, A, B : \mathcal{M}(\mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$$

$$F(\cdot) = \underbrace{|\cdot|(\mathcal{X})}_{A(\cdot)} + \underbrace{\mathbb{1}_{\mathcal{M}(\mathcal{X})^+}(\cdot)}_{B(\cdot)}$$

Convex conjugate of F can be obtained with the infimal convolution, see (Urruty 2006), (Fajardo 2012):

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

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

Convex conjugate of the indicator function $\mathbb{1}_{\mathcal{M}(\mathcal{X})^+}(\cdot)$

$$
\begin{aligned}
B^*(\psi) &= \sup_{m \in \mathcal{M}(\mathcal{X})} \langle \psi, m \rangle_{\mathcal{C}(\mathcal{X}) \times \mathcal{M}(\mathcal{X})} - B(m) \\
&= sup_{m \in \mathcal{M}(\mathcal{X})^+} \langle \psi, m \rangle_{\mathcal{C}(\mathcal{X}) \times \mathcal{M}(\mathcal{X})} \\
&\geq \langle \psi, m \rangle_{\mathcal{C}(\mathcal{X}) \times \mathcal{M}(\mathcal{X})} \forall m \in \mathcal{M}(\mathcal{X})^+
\end{aligned}
$$

If $\exists \bar{x} \in \mathcal{X}$ such that $\psi(\bar{x}) > 0$, consider $\bar{m} = \alpha \delta_{\bar{x}}$ with $\alpha > 0$.

$$
B^*(\psi) \geq \langle \psi, \bar{m} \rangle = \alpha \psi(\bar{x}) \xrightarrow{\alpha \to +\infty} +\infty \Rightarrow B^*(\psi) = +\infty.
$$

Notice that, if $\psi(x) \leq 0 \ \forall x \in \mathcal{X}$ $\langle \psi, m \rangle = \int_{\mathcal{X}} \psi \mathrm{d}m \leq 0 \ \forall m \in \mathcal{M}(\mathcal{X})^+$. Moreover, $\langle \psi, 0 \rangle = 0$. Thus, $B^*(\psi) = 0$ if $\psi(x) \leq 0 \ \forall x \in \mathcal{X}$.

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

Convex conjugate of the penalty

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

$$
F^*(\psi) = \begin{cases} 0 & \forall x \in \mathcal{X} \ \psi(x) \leq 1 \\ +\infty & \exists x \in \mathcal{X} \ \psi(x) > 1 \end{cases}
$$

Dual problem of $\mathcal{D}_{KL} - |\mu|$

$$\underset{\mu \in \mathcal{M}(\mathcal{X})}{\arg\min} \underbrace{\frac{1}{\lambda}\mathcal{D}_{KL}(\Phi\mu + b, y)}_{G(\Phi\mu)} + \underbrace{|\mu|(\mathcal{X}) + \mathbb{1}_{\mathcal{M}(\mathcal{X})^+}(\mu)}_{F(\mu)} \qquad (\mathcal{D}_{KL} - |\mu|)$$

$$\underset{p \in \mathcal{S}}{\arg\max} \underbrace{-\frac{b}{\lambda}(1 + \lambda p) + \frac{y}{\lambda}\log(1 + \lambda p)}_{-G^*(-p^*)} \qquad (\text{Dual of } \mathcal{D}_{KL} - |\mu|)$$

$$\mathcal{S} = \{p \in L^2(\mathcal{X}) : \underbrace{p > -\frac{1}{\lambda}}_{-G^*(-p^*)} \text{ and } \underbrace{\forall x \in \mathcal{X}\ \Phi^* p(x) \leq 1}_{-F^*(\Phi^* p^*)}\}$$

Extremality conditions

$$\begin{cases} -p_\lambda = \frac{1}{\lambda}\left(\frac{y}{\Phi\mu_\lambda + b} - I\right) \\ \Phi^* p_\lambda \in \partial|\mu_\lambda|(\mathcal{X}) + \partial\mathbb{1}_{\mathcal{M}(\mathcal{X})^+}(\mu_\lambda) \end{cases}$$

Optimality conditions for $\mathcal{D}_{KL} - |\mu|$

$$\mu = \sum_{i=1}^{N} a_i \delta_{x_i} \text{ is a solution of } (\mathcal{D}_{KL} - |\mu|)$$

$$\Downarrow$$

$$\left\| \left(\eta_\lambda(\mu) \right)_+ \right\|_\infty \leq 1,$$

i.e. $\eta_\lambda(\mu) \in \partial|\mu|(\mathcal{X}) + \partial \mathbb{1}_{\mathcal{M}(\mathcal{X})^+}(\mu)$

Dual certificate:

$$\eta_\lambda(\mu) = \frac{1}{\lambda} \Phi^* \left(I - \frac{y}{\Phi\mu + b} \right)$$

If $\mu = \sum_{i=1}^{N} a_i \delta_{x_i}$ with $a_i > 0$, $x_i \in \mathcal{X}$,

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

Sliding Frank Wolfe algorithm (Denoyelle-Duval-Peyré-Soubies 2019)

- Add **one spike per iteration**: new **position** computed from the **dual certificate**

$$\begin{cases} \eta^{[k]}(x) = \frac{1}{\lambda}\Phi^*\left(I - \frac{y}{\Phi\mu_{[k]}+b}\right)(x) \\ x_*^{[k]} \in \operatorname{argmax}_{x\in\mathcal{X}} \left(\eta^{[k]}(x)\right)_+ \\ x^{\text{new}} = (x_1^{[k]}, \cdots, x_k^{[k]}, x_*^{[k]}) \end{cases}$$

- If $\eta^{[k]}(x_*^{[k]}) < 1$, STOP.
- Estimate **amplitudes** of the reconstructed spikes, solving:

$$a^{\text{new}} \in \underset{a\in\mathbb{R}^{k+1}}{\operatorname{argmin}} \mathcal{D}_{KL}(\Phi_{x^{\text{new}}}(a) + b, y) + \lambda\|a\|_1 + \mathbb{1}_{\mathbb{R}_+^{k+1}}(a)$$

- **Sliding step**: non-convex step to adjust amplitudes and positions

$$(a^{[k+1]}, x^{[k+1]}) \in \underset{(a,x)\in\mathbb{R}^{k+1}\times\mathcal{X}^{k+1}}{\operatorname{argmin}} \mathcal{D}_{KL}(\Phi_x(a) + b, y) + \lambda\|a\|_1 + \mathbb{1}_{\mathbb{R}_+^{k+1}}(a)$$

solved starting from $\left(a^{\text{new}}, x^{\text{new}}\right)$

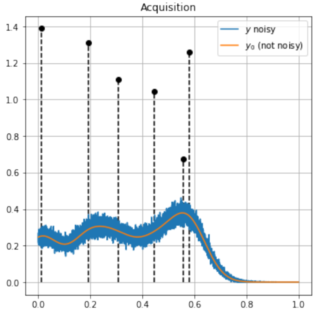
1D comparison between $\ell_2 - |\mu|$ and $\mathcal{D}_{KL} - |\mu|$

- 100 randomly simulated ground truths with 6 spikes
- Corresponding acquisitions simulated with Gaussian blur + Poisson noise
- Reconstructions with $\ell_2 - |\mu|$ for the following regularisation parameters:

$$10^{-x} * \eta_0, \quad x = 0, 1, \ldots, 10$$

with

$$\eta_\lambda(\mu) = \frac{1}{\lambda}\Phi^* (y - \Phi\mu)$$
$$\eta_1(0) = \Phi^* y$$
$$\eta_0 = \|\Phi^* y\|_\infty$$



If I initialise SFW with $\mu_{[0]} = 0$ and $\lambda = \eta_0$,

$$\eta_\lambda(\mu_{[0]}) = \frac{1}{\eta_0}\Phi^* y, \qquad \|\eta_\lambda(\mu_{[0]})\|_\infty = 1$$

1D comparison between $\ell_2 - |\mu|$ and $\mathcal{D}_{KL} - |\mu|$

- 100 randomly simulated ground truths with 6 spikes
- Corresponding acquisitions simulated with Gaussian blur + Poisson noise
- Reconstructions with $\mathcal{D}_{KL} - |\mu|$ for the following regularisation parameters:

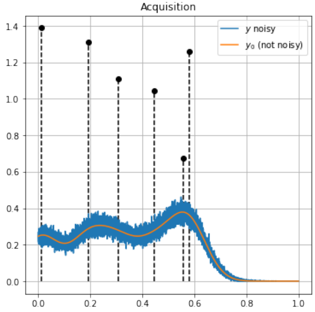
$$10^{-x} * \eta_0, \quad x = 0, 1, \dots, 10$$

with

$$\eta_\lambda(\mu) = \frac{1}{\lambda}\Phi^*\left(I - \frac{y}{\Phi\mu + b}\right)$$

$$\eta_1(0) = \Phi^*\left(I - \frac{y}{b}\right)$$

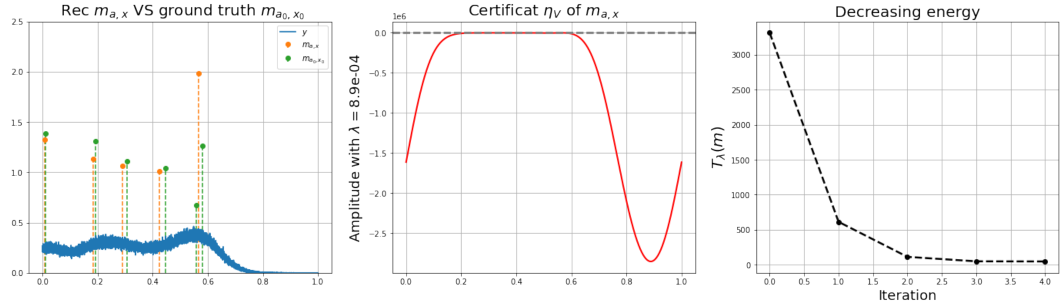
$$\eta_0 = \left\| \left(\Phi^*\left(\frac{b-y}{b}\right)\right)_+ \right\|_\infty$$



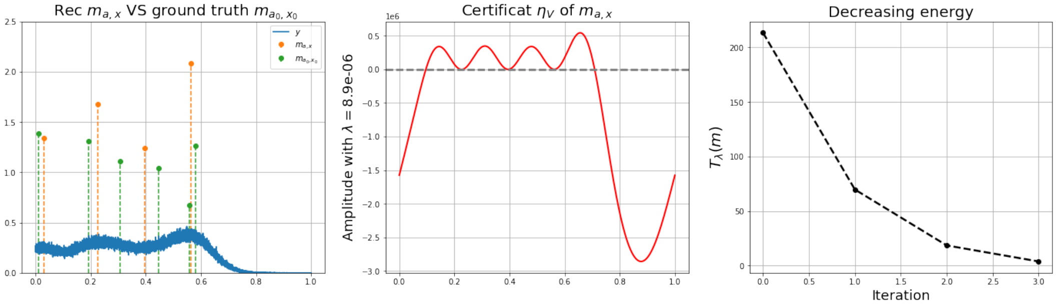
If I initialise SFW with $\mu_{[0]} = 0$ and $\lambda = \eta_0$,

$$\eta_\lambda(\mu_{[0]}) = \frac{1}{\eta_0}\Phi^*\left(\frac{b-y}{b}\right), \qquad \left\| \left(\eta_\lambda(\mu_{[0]})\right)_+ \right\|_\infty = 1$$

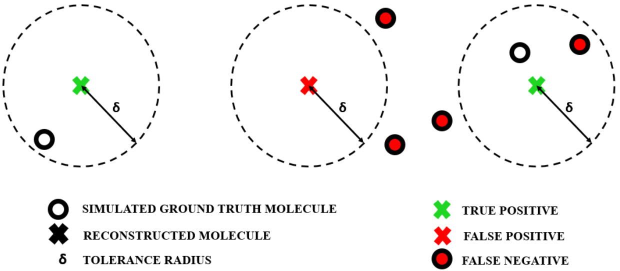
1D results: $\mathcal{D}_{KL} - |\mu|$ off-the-grid reconstruction



1D results: $\ell_2 - |\mu|$ off-the-grid reconstruction



Quality metrics considered



$$\mathrm{Jac}_\delta(\mu_{GT}, \mu_{REC}) = \frac{\#\mathrm{TP}}{\#\mathrm{TP} + \#\mathrm{FP} + \#\mathrm{FN}} \in [0, 1]$$

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

1D comparison between $\ell_2 - |\mu|$ and $\mathcal{D}_{KL} - |\mu|$

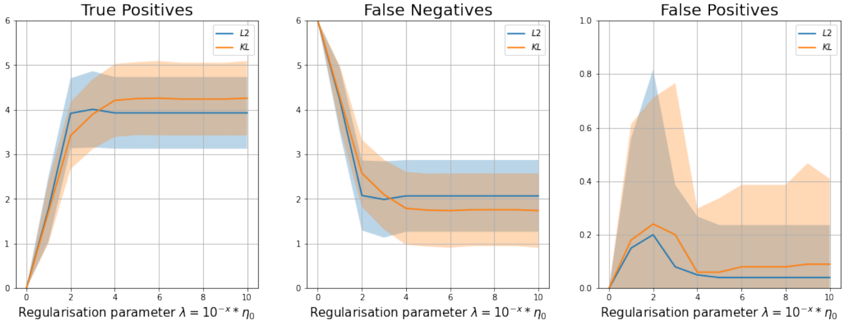


Figure: Mean values over 100 different randomly generated ground truths with 6 spikes and their corresponding reconstructions. Shaded area corresponds to standard deviation.

Parameters	$\ell_2 - \lvert\mu\rvert$	$\mathcal{D}_{KL} - \lvert\mu\rvert$
η_0	$\lVert\Phi^* y\rVert_\infty$	$\left\lVert \left(\Phi^*\left(\frac{b-y}{b}\right)\right)_+ \right\rVert_\infty$
λ	$10^{-x}\eta_0,\ x = 0, 1, \dots, 10$	$10^{-x}\eta_0,\ x = 0, 1, \dots, 10$
Max. iter.	$2 * N_{molecules}$	$2 * N_{molecules}$
Tolerance radius δ	0.05	0.05

1D comparison between $\ell_2 - |\mu|$ and $\mathcal{D}_{KL} - |\mu|$

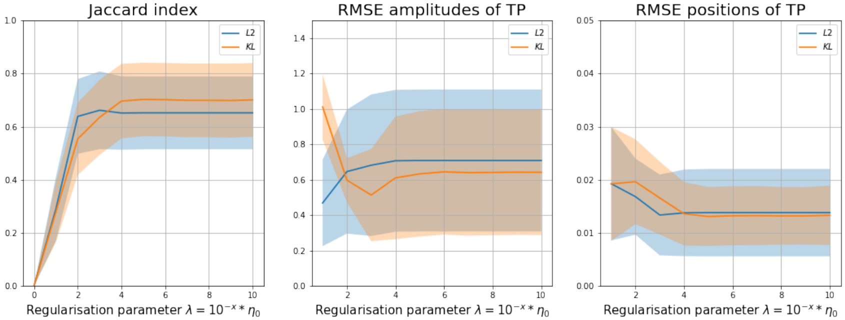
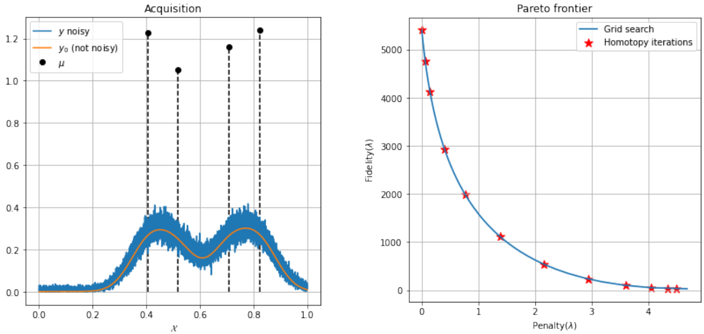


Figure: Mean values over 100 different randomly generated ground truths with 6 spikes and their corresponding reconstructions. Shaded area corresponds to standard deviation.

Homotopy

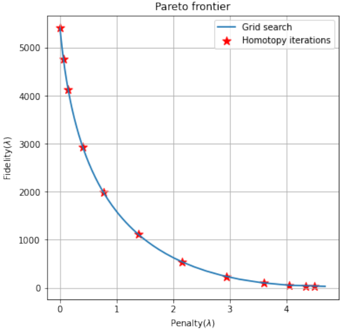
How to choose λ for SFW?



Homotopy strategy (Malioutov-Cetin-Willsky 2005) (Donoho-Tsaig 2006)

Explore the Pareto frontier by solving the $(\mathcal{D}_{KL} - |\mu|)(\lambda)$ problem for a **decreasing sequence of regularisation parameters** λ until the solution $\hat{\mu}_\lambda$ satisfies:

$$\mathcal{D}_{KL}(\Phi\hat{\mu}_\lambda + b, y) \leq \sigma_{\text{target}}$$



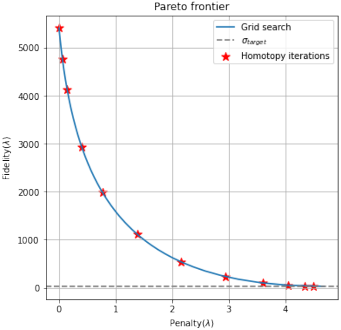
Design of the homotopy algorithm:

- Choice of starting value for λ
- How to account for past knowledge
- How λ will evolve
- Choice of σ_{target} to stop the algorithm

Homotopy strategy (Malioutov-Cetin-Willsky 2005) (Donoho-Tsaig 2006)

Explore the Pareto frontier by solving the $(\mathcal{D}_{KL} - |\mu|)(\lambda)$ problem for a **decreasing sequence of regularisation parameters** λ until the solution $\hat{\mu}_\lambda$ satisfies:

$$\mathcal{D}_{KL}(\Phi\hat{\mu}_\lambda + b, y) \leq \sigma_{\text{target}}$$



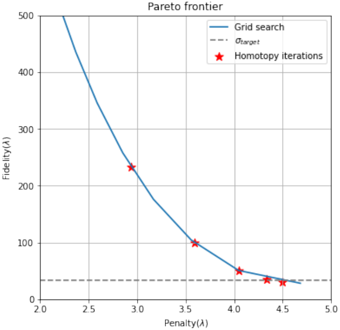
Design of the homotopy algorithm:

- Choice of starting value for λ
- How to account for past knowledge
- How λ will evolve
- Choice of σ_{target} to stop the algorithm

Homotopy strategy (Malioutov-Cetin-Willsky 2005) (Donoho-Tsaig 2006)

Explore the Pareto frontier by solving the $(\mathcal{D}_{KL} - |\mu|)(\lambda)$ problem for a **decreasing sequence of regularisation parameters** λ until the solution $\hat{\mu}_\lambda$ satisfies:

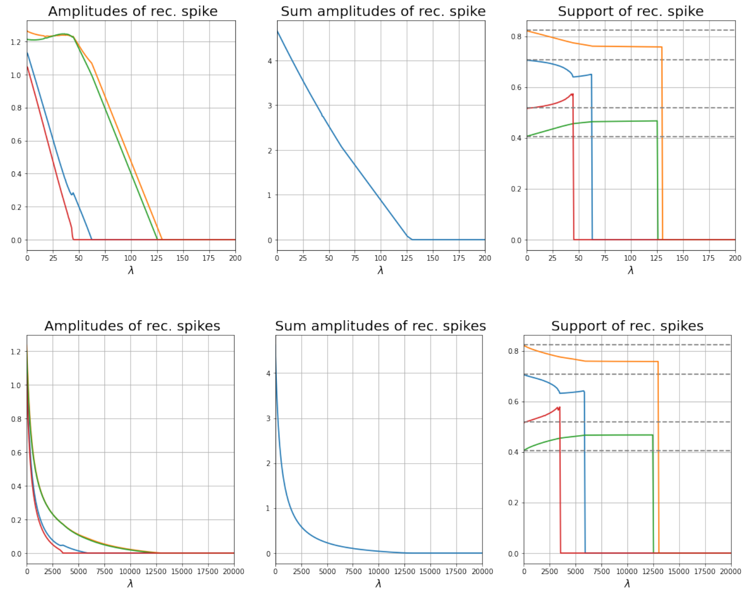
$$\mathcal{D}_{KL}(\Phi\hat{\mu}_\lambda + b, y) \leq \sigma_{\text{target}}$$



Design of the homotopy algorithm:

- Choice of starting value for λ
- How λ will evolve
- How to account for past knowledge
- Choice of σ_{target} to stop the algorithm

Regularisation path



$$\lambda \longmapsto a_i(\lambda)$$

Discrete setting:

- $\ell_2 - \ell_1$: piecewise linear (Mairal-Yu 2012)
- $f - \ell_1$ with f non-linear: piecewise smooth (Bieker-Gebken-Peitz 2022)

Continuous setting:

- $\ell_2 - |\mu|$: piecewise linear (Courbot-Colicchio 2021)

Homotopy-SFW (Courbot-Colicchio 2021)

Algorithm 1: Homotopy-Sliding Frank Wolfe Algorithm

Input: $\mathbf{y} > 0$, $\mathbf{b} > 0$, Φ, $c > 0$, $\sigma_{\text{target}} > 0$

Output: estimation $\hat{\mu}$

Initialization: $\hat{\mu}_0 = 0$ and $\lambda_1 = \left\| \left(\Phi^* \left(\frac{\mathbf{y}-\mathbf{b}}{\mathbf{b}} \right) \right)_+ \right\|_\infty$

FOR $t = 1, \dots$ REPEAT

1. Compute $\hat{\mu}_t$ solution of $(\mathcal{P}_{\lambda_t})$ with SFW with warm start $\mu_t^{[0]} = \hat{\mu}_{t-1}$.
2. Compute σ_t from the residual:
$$\sigma_t = \mathcal{D}_{KL}(\Phi\hat{\mu}_t + \mathbf{b}, \mathbf{y})$$
3. IF $\sigma_t < \sigma_{\text{target}}$
   $\hat{\mu}_t$ is a solution $\Rightarrow$ STOP
4. ELSE IF $\sigma_t \geq \sigma_{\text{target}}$
$$\text{Update} \quad \lambda_{t+1} = \frac{\lambda_t \left\| \left(\eta_t(\hat{\mu}_t) \right)_+ \right\|_\infty}{c+1}, \qquad \eta_t = \frac{1}{\lambda_t} \Phi^* \partial_1 \mathcal{D}_{KL}(\Phi\hat{\mu}_t + \mathbf{b}, \mathbf{y})$$

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

Homotopy-SFW (Courbot-Colicchio 2021)

- If $\lambda > \lambda_1$, the dual certificate $\|\left(\eta_\lambda(\hat{\mu}_0\right)_+\|_\infty < 1$. No spikes are reconstructed.
- The sequence of estimated regularisation parameter is strictly decreasing: $\lambda_{t+1} < \lambda_t$.
- At the beginning of each homotopy iteration $t+1$:

$$\left\|\left(\eta_{t+1}(\mu_{t+1}^{[0]})\right)_+\right\|_\infty = \left\|\left(\eta_{t+1}(\hat{\mu}_t)\right)_+\right\|_\infty = \frac{\lambda_t}{\lambda_{t+1}} \left\|\left(\eta_t(\hat{\mu}_t)\right)_+\right\|_\infty = 1 + c > 1,$$

hence an inner iteration of SFW is always performed.

The homotopy-SFW algorithm for the $\mathcal{D}_{KL} - |\mu|$ off-the-grid problem produces a strictly decreasing sequence $(\sigma_t)_t$.

Homotopy-SFW (Courbot-Colicchio 2021)

$$\mu^*(\lambda) = \arg\min_{\mu} C(y, \lambda, \mu) = \mathcal{D}_{KL}(\Phi\mu + b, y) + \lambda|\mu|(\mathcal{X}) \tag{1}$$

$$\frac{\partial C(y,\lambda,\mu)}{\partial \mu}\Big|_{\mu=\mu^*} = \Phi^*\left(\mathbf{I} - \frac{y}{\Phi\mu^* + b}\right) + \lambda = 0 \qquad (\mu^*(\lambda) \text{ minimizer of } (1))$$

$$0 = \frac{\partial}{\partial \lambda}\left(\frac{\partial C(y,\lambda,\mu)}{\partial \mu}\Big|_{\mu=\mu^*}\right) \qquad (\text{Deriving again with respect to } \lambda)$$

$$0 = \frac{\partial}{\partial \mu}\left(\Phi^*\left(\mathbf{I} - \frac{y}{\Phi\mu + b}\right)\right)\Big|_{\mu=\mu^*} \cdot \frac{\partial \mu^*(\lambda)}{\partial \lambda} + 1$$

$$0 = \Phi^*\Phi \cdot \frac{y}{\|\Phi\mu^* + b\|^2} \cdot \frac{\partial \mu^*(\lambda)}{\partial \lambda} + 1 \qquad \Rightarrow \frac{\partial \mu^*(\lambda)}{\partial \lambda} < 0$$

Homotopy-SFW (Courbot-Colicchio 2021)

$$\frac{\partial C(y,\lambda,\mu)}{\partial \mu}\Big|_{\mu=\mu^*} = \Phi^*\left(\mathbf{I} - \frac{y}{\Phi\mu^* + b}\right) + \lambda = 0 \qquad \Rightarrow \Phi^*\left(\mathbf{I} - \frac{y}{\Phi\mu^* + b}\right) = -\lambda$$

$$\frac{\partial}{\partial \lambda}\mathcal{D}_{KL}(\Phi\mu^*(\lambda) + b, y) = \Phi^*\left(\mathbf{I} - \frac{y}{\Phi\mu^* + b}\right) \cdot \underbrace{\frac{\partial \mu^*(\lambda)}{\partial \lambda}}_{<0} > 0$$

This entails that $\lambda \to \mathcal{D}_{KL}(\Phi\mu^*(\lambda) + b, y)$ is strictly increasing.
Since $\lambda_{t+1} < \lambda_t$ and $\hat{\mu}_t = \mu^*(\lambda_t)$,

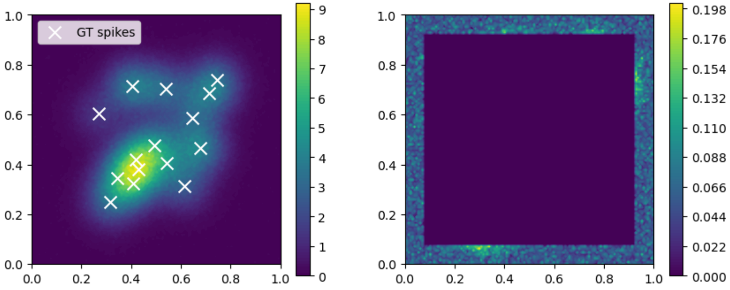
$$\sigma_{t+1} = \mathcal{D}_{KL}(\Phi\hat{\mu}_{t+1} + b, y) < \sigma_t = \mathcal{D}_{KL}(\Phi\hat{\mu}_t + b, \mathbf{y}). \quad \square$$

1D comparison between $\ell_2 - |\mu|$ and $\mathcal{D}_{KL} - |\mu|$ with Homotopy-SFW

Mean value	$\ell_2 - \|\mu\|$	$\mathcal{D}_{KL} - \|\mu\|$
Jaccard index	0.72	**0.74**
Number of TP	**3.32**	3.30
Number of FN	0.68	**0.70**
Number of FP	0.73	**0.54**
RMSE on amplitudes of TP	0.50	**0.42**
RMSE on positions of TP	0.012	**0.011**
Final estimated λ	6.09	40.21
Number of homotopy iterations	4.55	3.93
Value of σ_{target}	4.09	77.16

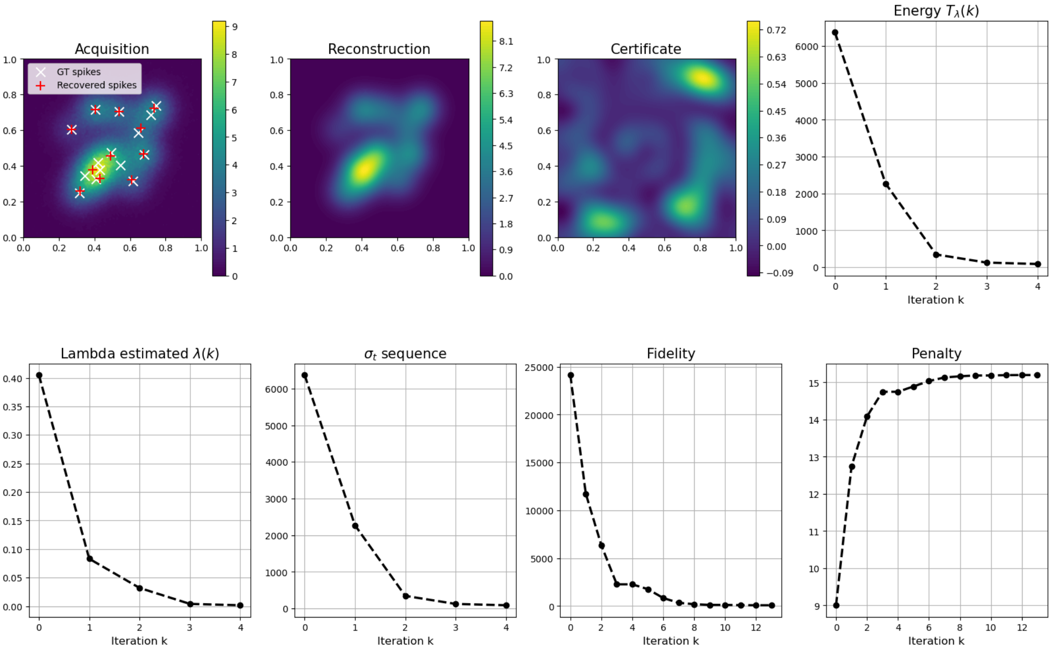
Parameters	$\ell_2 - \|\mu\|$	$\mathcal{D}_{KL} - \|\mu\|$
Initial value λ_0	$\frac{1}{5}\|\Phi^* y\|_\infty$	$\frac{1}{5}\left\|\left(\Phi^*\left(\frac{b-y}{b}\right)\right)_+\right\|_\infty$
Max number of homotopy iterations	$2 * N_{molecules}$	$2 * N_{molecules}$
Max number of inner SFW iterations	1	1
Homotopy parameter c	13	45
Choice of σ_{target}	$1.5 * \frac{1}{2}\|\Phi\mu_{gt} + b - y\|^2$	$1.5 * \mathcal{D}_{KL}(\Phi\mu_{gt} + b, y)$

Homotopy-SFW: choice of σ_{target} for $\mathcal{D}_{KL} - |\mu|$



Theoretical value (knowledge of ground truth)	$\mathcal{D}_{KL}(\Phi\mu_{gt} + b, y)$	82.74
(Bertero 2010) [1]	$M/2$	8192
(Bertero 2010) + knowledge of the *Poisson noise level nv*	$M/(2 * nv)$	81.92
Estimation based only on the background	$\mathcal{D}_{KL}(b, y_{bg})/M_{bg} * M$	104.66

[1] Bertero M., Boccacci P., Talenti G., Zanella R., Zanni L., *A discrepancy principle for Poisson data*, Inverse Problems, 2010.


Acquisition
GT spikes
Recovered spikes
Reconstruction
Certificate
Energy $T_\lambda(k)$
Iteration k
Lambda estimated $\lambda(k)$
Iteration k
σ_t sequence
Iteration k
Fidelity
Iteration k
Penalty
Iteration k

3D real data results

3D off-the-grid volume deblurring

Volume acquired in widefield microscopy of yeasts expressing fluorescent proteins (*SEC16-sfGFP* and a *SEC24-sfGFP*) localized at the Endoplasmic Reticulum exit sites (ERES) [a]

[a] Toader B., Boulanger J., Korolev Y., Lenz M.O., Manton J., Schonlieb C.B., Muresan L., *Image reconstruction in light-sheet microscopy: spatially varying deconvolution and mixed noise*, Journal of Mathematical Imaging and Vision, 2022.

- 3D volume blurred and noisy acquisition: $190 \times 190 \times 17$ voxels
- Voxel size: 65nm in xy and 250nm in z

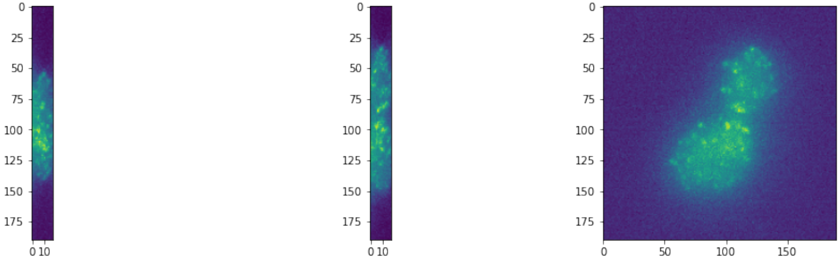
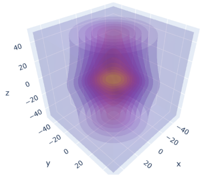


Figure: Maximum Intensity projection over *xz*, *yz*, *yx* planes.

PSF estimation

- 3D Gaussian PSF:

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



- $FWHM = 0.61\frac{\lambda_{wavelength}}{NA}$ with $\lambda_{wavelength} = 509$nm and $NA = 1.49$ on the xy plane. To recover σ: $FWHM = 2.355 * \sigma$.
- $\sigma_x = \sigma_y = 89$nm and $\sigma_z = 2 * \sigma_x$.

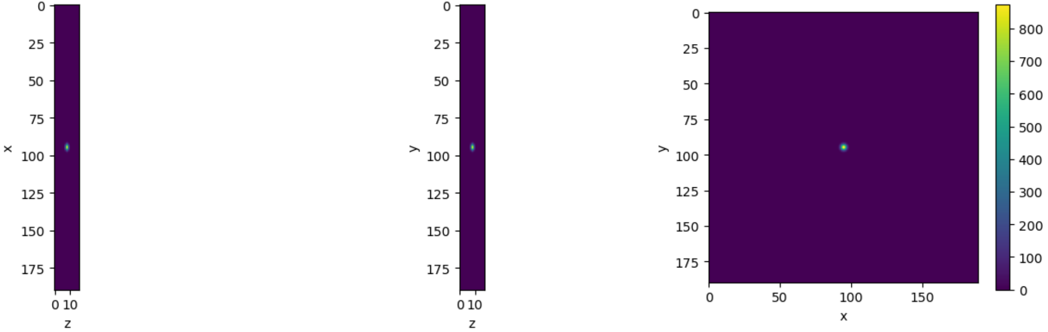


Figure: Maximum Intensity projection over xz, yz, yx planes.

Background estimation

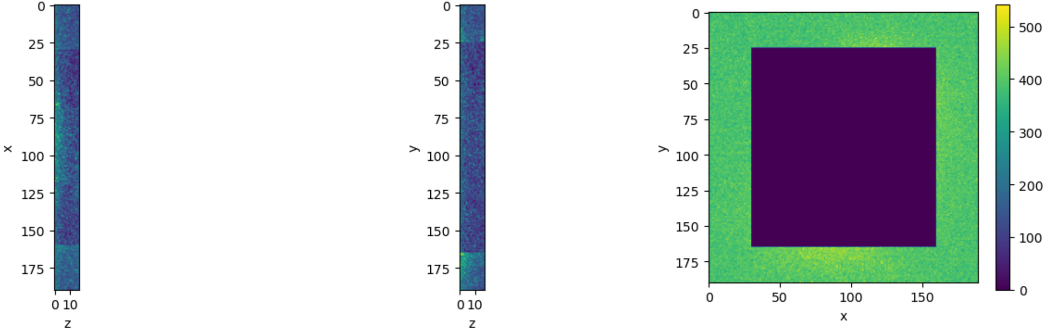


Figure: Maximum Intensity projection over *xz*, *yz*, *yx* planes.

- Constant background estimation: $b = 337.77$
- Estimation of $\sigma_{\text{target}} = 1102067.75$
- Bertero's estimate: $M/2 = 306850$

Reconstruction up to 130 spikes with H-SFW $\mathcal{D}_{KL} - |\mu|$

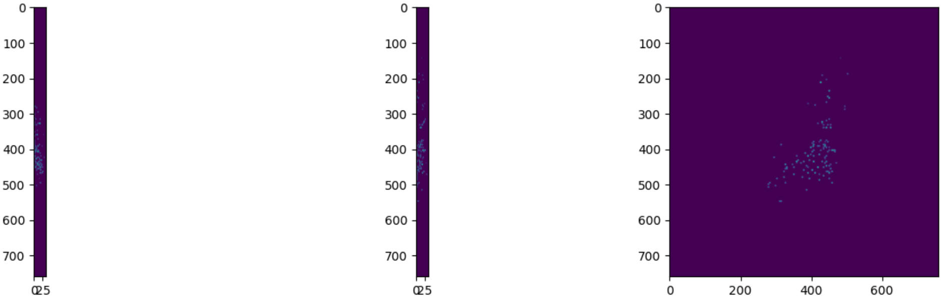


Figure: Maximum Intensity projection over *xz*, *yz*, *yx* planes.

Parameters	$\mathcal{D}_{KL} - \lvert\mu\rvert$
Initial value λ_0	$\left\| \left(\Phi^* \left(\frac{\mathbf{y} - \mathbf{b}}{\mathbf{b}} \right) \right)_+ \right\|_\infty$
Max number of homotopy iterations	10
Max number of inner SFW iterations	50
Homotopy parameter c	0.15
Choice of σ_{target}	based on the background

Conclusion and future work

Contributions

- Study of the dual problem and optimality conditions for off-the-grid variational model with Kullback-Leibler divergence as fidelity term and positivity constraints
- Practical implementation of off-the-grid approaches for Poisson noise formulation with the Kullback-Leibler divergence as fidelity term and positivity constraints
- Automatic selection of the regularisation parameter thanks to homotopy strategies
- 3D PSF modelled by Gaussian blur

Future work

- more general PSFs
- test on real-microscopy data with low photon count (e.g., fluctuation-based approaches)

Thanks for your attention!

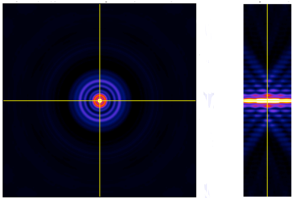


Figure: Airy disks (3D fitting with defocus effect)

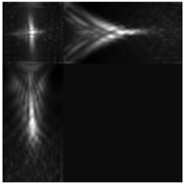


Figure: Light-sheet beam with Zernike polynomials

References

- Bertero M., Boccacci P., Talenti G., Zanella R., Zanni L., *A discrepancy principle for Poisson data*, Inverse Problems, 2010.
- Bieker K., Gebken B., Peitz S., *On the Treatment of Optimization Problems With L1 Penalty Terms via Multiobjective Continuation*, IEEE transactions on pattern analysis and machine intelligence, 2021.
- Boyer C., De Castro Y., Salmon J., *Adapting to unknown noise level in sparse deconvolution*, Information and Inference: A Journal of the IMA, 2017.
- Bredies K., Pikkarainen H. K., *Inverse problems in spaces of measures*, ESAIM: Control, Optimisation and Calculus of Variations, 2013.
- Candès E.J., Fernandez-Granda C., *Towards a Mathematical Theory of Super-resolution*, Communications on Pure and Applied Mathematics, 2013.
- Courbot J.B., Colicchio B.. *A Fast Homotopy Algorithm for Gridless Sparse Recovery*, Inverse Problems, 2021.
- De Castro Y., Gamboa F., *Exact reconstruction using Beurling minimal extrapolation*, Journal of Mathematical Analysis and Applications, 2012.
- Denoyelle Q. Duval V., Peyré G., *Support Recovery for Sparse Super-Resolution of Positive Measures*, Journal of Fourier Analysis and Applications, 2017.
- Denoyelle Q., Duval V., Peyré G., Soubies E., *The Sliding Frank Wolfe Algorithm and its Application to Super-Resolution*, Inverse Problems, 2019.

References

- Donoho D.L., Tsaig Y., *Fast Solution of ‘1-norm Minimization Problems When the Solution May be Sparse*, 2006.
- Duval V. Peyré G., *Exact Support Recovery for Sparse Spikes Deconvolution*, Foundations of Computational Mathematics, 2014.
- Frank M., Wolfe P, *An algorithm for quadratic programming*, Naval Res. Logist. Quart., 1956.
- Kirshner H., Aguet F.,Sage D., Unser M., *3-D PSF fitting for fluorescence microscopy: implementation and localization application*, Journal of Microscopy, 2013.
- Laville B., Blanc-Féraud L. and Aubert. G, *Off-the-grid variational sparse spike recovery: methods and algorithms*, Journal of Imaging, 2021.
- Mairal J., Bin Y., *Complexity Analysis of the Lasso Regularization Path*, International conference on Machine Learning, 2012.
- Malioutov D. M., Cetin M., Willsky A. S., *Homotopy continuation for sparse signal representation*, ICASSP 2005.
- Poon C., Peyré G., *Multi-dimensional sparse super-resolution*, SIAM J. Math. Anal., 2019.
- Toader B., Boulanger J., Korolev Y., Lenz M.O., Manton J., Schonlieb C.B., Muresan L., *Image reconstruction in light-sheet microscopy: spatially varying deconvolution and mixed noise*, Journal of Mathematical Imaging and Vision, 2022.

Dual problem of $\mathcal{D}_{KL} - |\mu|$

$$\underset{\mu \in \mathcal{M}(\mathcal{X})}{\arg\min} \underbrace{\frac{1}{\lambda}\mathcal{D}_{KL}(\Phi\mu + b, y)}_{G(\Phi\mu)} + \underbrace{|\mu|(\mathcal{X}) + \mathbb{1}_{\mathcal{M}(\mathcal{X})^+}(\mu)}_{F(\mu)} \qquad (\mathcal{D}_{KL} - |\mu|)$$

- $G : L^2(\mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$

$$G(s) := \begin{cases} \frac{1}{\lambda}\int_{\mathcal{X}} \Big(s+b\Big)(x) - y(x) + y(x)\log(y(x)) - y(x)\log\Big[\Big(s+b\Big)(x)\Big] \,\mathrm{d}x & s+b \in L^2(\mathcal{X})^+ \\ +\infty & s+b \notin L^2(\mathcal{X})^+ \end{cases}$$

- $F : \mathcal{M}(\mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$, $F(\cdot) = |\cdot|(\mathcal{X}) + \mathbb{1}_{\mathcal{M}(\mathcal{X})^+}(\cdot)$
- $\Phi : \mathcal{M}(\mathcal{X}) \to L^2(\mathcal{X})$

$$\underset{p^* \in \left(L^2(\mathcal{X})\right)^*}{\arg\max} \; -F^*(\Phi^* p^*) - G^*(-p^*) \qquad (\text{Dual of } \mathcal{D}_{KL} - |\mu|)$$

Convex conjugate of $\mathcal{D}_{KL}$

Given $t, b, \lambda > 0$, consider the one-dimensional Kullback-Leibler function, defined by

$$f_{t,b}(s) = \begin{cases} \frac{1}{\lambda}\big(s + b - t + t\log(t) - t\log(s+b)\big) & s + b > 0 \\ +\infty & s + b \leq 0 \end{cases}$$

$$f^*_{t,b}(s^*) = \sup_{s\in\mathbb{R}} ss^* - f_{t,b}(s) = \sup_{s+b>0} ss^* - \frac{1}{\lambda}\big(s + b - t + t\log(t) - t\log(s+b)\big) =$$
$$= \sup_{s+b>0} \underbrace{s\left(s^* - \frac{1}{\lambda}\right) - \frac{b}{\lambda} + \frac{t}{\lambda}\log(s+b) + \frac{t}{\lambda} - \frac{t}{\lambda}\log(t)}_{h(s)}$$

Convex conjugate of $\mathcal{D}_{KL}$

- If $s^* \geq \frac{1}{\lambda}$, then $\lim_{s\to+\infty} h(s) = +\infty$ implies $\sup_{s>0} h(s) = +\infty \Rightarrow f_{t,b}(s^*) = +\infty$.
- If $s^* < \frac{1}{\lambda}$, then $\lim_{s\to\pm\infty} h(s) = -\infty$.

$$h'(s) = s^* - \frac{1}{\lambda} + \frac{t}{\lambda(s+b)} = \frac{\lambda(s+b)s^* - (s+b) + t}{\lambda(s+b)} > 0$$

$$\iff \lambda(s+b)s^* - (s+b) + t > 0 \iff s > \frac{t}{1-\lambda s^*} - b$$

$$f^*_{t,b}(s^*) = h\left(\frac{t}{1-\lambda s^*} - b\right) = \frac{b}{\lambda}(1-\lambda s^*) - \frac{t}{\lambda}\log(1-\lambda s^*)$$

which is well-defined since $1 - \lambda s^* > 0 \iff s^* < \frac{1}{\lambda}$.

Convex conjugate of $\mathcal{D}_{KL}$

- If $s^* \geq \frac{1}{\lambda}$, then $\lim_{s \to +\infty} h(s) = +\infty$ implies $\sup_{s>0} h(s) = +\infty \Rightarrow f_{t,b}\left(s^*\right) = +\infty$.
- If $s^* < \frac{1}{\lambda}$, then $\lim_{s \to \pm\infty} h(s) = -\infty$.

$$h'(s) = s^* - \frac{1}{\lambda} + \frac{t}{\lambda(s+b)} = \frac{\lambda(s+b)s^* - (s+b) + t}{\lambda(s+b)} > 0$$

$$\iff \lambda(s+b)s^* - (s+b) + t > 0 \iff s > \frac{t}{1-\lambda s^*} - b$$

$$f^*_{t,b}(s^*) = h\left(\frac{t}{1-\lambda s^*} - b\right) = \frac{b}{\lambda}(1-\lambda s^*) - \frac{t}{\lambda}\log(1-\lambda s^*)$$

which is well-defined since $1 - \lambda s^* > 0 \iff s^* < \frac{1}{\lambda}$.

$$f^*_{t,b}(s^*) = \begin{cases} +\infty & s^* \geq \frac{1}{\lambda} \\ \frac{b}{\lambda}(1-\lambda s^*) - \frac{t}{\lambda}\log(1-\lambda s^*) & s^* < \frac{1}{\lambda} \end{cases}$$

Convex conjugate of $\mathcal{D}_{KL}$

- If $s^* \geq \frac{1}{\lambda}$, then $\lim_{s\to+\infty} h(s) = +\infty$ implies $\sup_{s>0} h(s) = +\infty \Rightarrow f_{t,b}(s^*) = +\infty$.
- If $s^* < \frac{1}{\lambda}$, then $\lim_{s\to\pm\infty} h(s) = -\infty$.

$$h'(s) = s^* - \frac{1}{\lambda} + \frac{t}{\lambda(s+b)} = \frac{\lambda(s+b)s^* - (s+b) + t}{\lambda(s+b)} > 0$$

$$\iff \lambda(s+b)s^* - (s+b) + t > 0 \iff s > \frac{t}{1-\lambda s^*} - b$$

$$f_{t,b}^*(s^*) = h\left(\frac{t}{1-\lambda s^*} - b\right) = \frac{b}{\lambda}(1-\lambda s^*) - \frac{t}{\lambda}\log(1-\lambda s^*)$$

which is well-defined since $1 - \lambda s^* > 0 \iff s^* < \frac{1}{\lambda}$.

$$G^*(p^*) = \begin{cases} +\infty & p^* \geq \frac{1}{\lambda} \\ \frac{\mathbf{b}}{\lambda}(1-\lambda p^*) - \frac{\mathbf{y}}{\lambda}\log(1-\lambda p^*) & p^* < \frac{1}{\lambda} \end{cases}$$

Convex conjugate of the penalty

$$F(\cdot) = A(\cdot) + B(\cdot) \qquad F, A, B : \mathcal{M}(\mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$$

$$F(\cdot) = \underbrace{|\cdot|(\mathcal{X})}_{A(\cdot)} + \underbrace{\mathbb{1}_{\mathcal{M}(\mathcal{X})^+}(\cdot)}_{B(\cdot)}$$

Convex conjugate of F can be obtained with the infimal convolution, see (Urruty 2006), (Fajardo 2012):

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

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

Convex conjugate of the indicator function $\mathbb{1}_{\mathcal{M}(\mathcal{X})^+}(\cdot)$

$$
\begin{aligned}
B^*(\psi) &= \sup_{m \in \mathcal{M}(\mathcal{X})} \langle \psi, m \rangle_{\mathcal{C}(\mathcal{X}) \times \mathcal{M}(\mathcal{X})} - B(m) \\
&= sup_{m \in \mathcal{M}(\mathcal{X})^+} \langle \psi, m \rangle_{\mathcal{C}(\mathcal{X}) \times \mathcal{M}(\mathcal{X})} \\
&\geq \langle \psi, m \rangle_{\mathcal{C}(\mathcal{X}) \times \mathcal{M}(\mathcal{X})} \forall m \in \mathcal{M}(\mathcal{X})^+
\end{aligned}
$$

If $\exists \bar{x} \in \mathcal{X}$ such that $\psi(\bar{x}) > 0$, consider $\bar{m} = \alpha \delta_{\bar{x}}$ with $\alpha > 0$.

$$
B^*(\psi) \geq \langle \psi, \bar{m} \rangle = \alpha \psi(\bar{x}) \xrightarrow{\alpha \to +\infty} +\infty \Rightarrow B^*(\psi) = +\infty.
$$

Notice that, if $\psi(x) \leq 0 \ \forall x \in \mathcal{X}$ $\langle \psi, m \rangle = \int_{\mathcal{X}} \psi \mathrm{d}m \leq 0 \ \forall m \in \mathcal{M}(\mathcal{X})^+$. Moreover, $\langle \psi, 0 \rangle = 0$. Thus, $B^*(\psi) = 0$ if $\psi(x) \leq 0 \ \forall x \in \mathcal{X}$.

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

Convex conjugate of the penalty

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

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

Convex conjugate of the penalty

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

- If $\|\psi\|_\infty > 1$ and $\exists \bar{x}$ such that $\psi(\bar{x}) > 1$, we have $F^*(\psi) = +\infty$.
  Assume there exist $\psi_1, \psi_2 \in \mathcal{C}(\mathcal{X})$ such that $\psi_1 + \psi_2 = \psi$ and $\|\psi_1\|_\infty \leq 1$ and $\psi_2(x) \leq 0 \ \forall x \in \mathcal{X}$.

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

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

Convex conjugate of the penalty

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

$$= \min_{\psi_1+\psi_2=\psi} \begin{cases} 0 & \|\psi_1\|_\infty \leq 1 \\ +\infty & \|\psi_1\|_\infty > 1 \end{cases} + \begin{cases} 0 & \psi_2(x) \leq 0 \ \forall x \in \mathcal{X} \\ +\infty & \exists x \in \mathcal{X} \text{ s.t. } \psi_2(x) > 0 \end{cases}$$

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

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

$\psi = \psi^+ + \psi^-$ and $\|\psi^+\|_\infty \leq 1$ and $\forall\ x \in \mathcal{X}\ \psi^-(x) \leq 0$, thus $F^*(\psi) = 0$.

Convex conjugate of the penalty

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

$$= \min_{\psi_1+\psi_2=\psi} \begin{cases} 0 & \|\psi_1\|_\infty \leq 1 \\ +\infty & \|\psi_1\|_\infty > 1 \end{cases} + \begin{cases} 0 & \psi_2(x) \leq 0 \ \forall x \in \mathcal{X} \\ +\infty & \exists x \in \mathcal{X} \text{ s.t. } \psi_2(x) > 0 \end{cases}$$

$$F^*(\psi) = \begin{cases} 0 & \forall x \in \mathcal{X} \ \psi(x) \leq 1 \\ +\infty & \exists x \in \mathcal{X} \ \psi(x) > 1 \end{cases}$$

Dual problem of $\mathcal{D}_{KL} - |\mu|$

$$\underset{\mu \in \mathcal{M}(\mathcal{X})}{\arg\min} \underbrace{\frac{1}{\lambda}\mathcal{D}_{KL}(\Phi\mu + b, y)}_{G(\Phi\mu)} + \underbrace{|\mu|(\mathcal{X}) + \mathbb{1}_{\mathcal{M}(\mathcal{X})^+}(\mu)}_{F(\mu)} \qquad (\mathcal{D}_{KL} - |\mu|)$$

$$\underset{p \in \mathcal{S}}{\arg\max} \underbrace{-\frac{b}{\lambda}(1 + \lambda p) + \frac{y}{\lambda}\log(1 + \lambda p)}_{-G^*(-p^*)} \qquad (\text{Dual of } \mathcal{D}_{KL} - |\mu|)$$

$$\mathcal{S} = \{p \in L^2(\mathcal{X}) : \underbrace{p > -\frac{1}{\lambda}}_{-G^*(-p^*)} \text{ and } \underbrace{\forall x \in \mathcal{X} \ \Phi^* p(x) \leq 1}_{-F^*(\Phi^* p^*)}\}$$

Extremality conditions

$$\begin{cases} -p_\lambda = \frac{1}{\lambda}\left(\frac{y}{\Phi\mu_\lambda + b} - I\right) \\ \Phi^* p_\lambda \in \partial|\mu_\lambda|(\mathcal{X}) + \partial\mathbb{1}_{\mathcal{M}(\mathcal{X})^+}(\mu_\lambda) \end{cases}$$

Algorithm 2: Sliding Frank Wolfe Algorithm for $\ell_2 - |\mu|$

- **Input:** initialisation $\mu_{[0]}=0$ and maximum number of iterations $K_{\max} \in \mathbb{N}$
- **Repeat:** For $0 \leq k \leq K_{\max}$
  1. **Insertion step:**

$$x_*^{[k]} \in \underset{x \in \mathcal{X}}{\operatorname{argmax}} |\eta^{[k]}(x)| \text{ with } \eta^{[k]}(x) = \frac{1}{\lambda}\Phi^*(\Phi\mu_{[k]} - y)$$

     - If $|\eta^{[k]}(x_*^{[k]})| < 1$, then $\mu_{[k]}$ is the solution of BLASSO problem
     - Else $\mu_{[k]} = \sum_{i=1}^{k} a_i^{[k]} \delta_{x_i^{[k]}}$ has to be updated
  2. **Update positions and amplitudes:**

$$x^{[k+1/2]} = (x_1^{[k]}, \cdots, x_k^{[k]}, x_*^{[k]})$$

$$a^{[k+1/2]} \in \underset{a \in \mathbb{R}^{k+1}}{\operatorname{argmin}} \frac{1}{2}\|y - \Phi_{x^{[k+1/2]}}(a)\|_2^2 + \lambda\|a\|_1$$

  3. **Sliding step:**

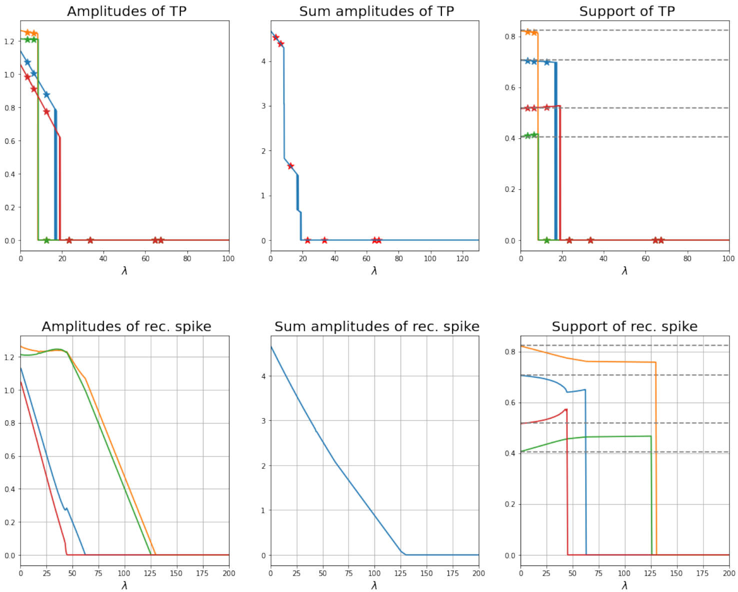
$$(a^{[k+1]}, x^{[k+1]}) \in \underset{(a,x) \in \mathbb{R}^{k+1} \times \mathcal{X}^{k+1}}{\operatorname{argmin}} \frac{1}{2}\|y - \Phi_x(a)\|_2^2 + \lambda\|a\|_1$$

- **Until:** $k = K_{\max}$;
- **Output:**
  - $K \in \mathbb{N}$ number of spikes;
  - $a \in \mathbb{R}^K$ s.t. $a_i > 0$ for $i = 1, \cdots, K$ amplitudes;
  - $x \in \mathcal{X}^K$ positions of the spikes.

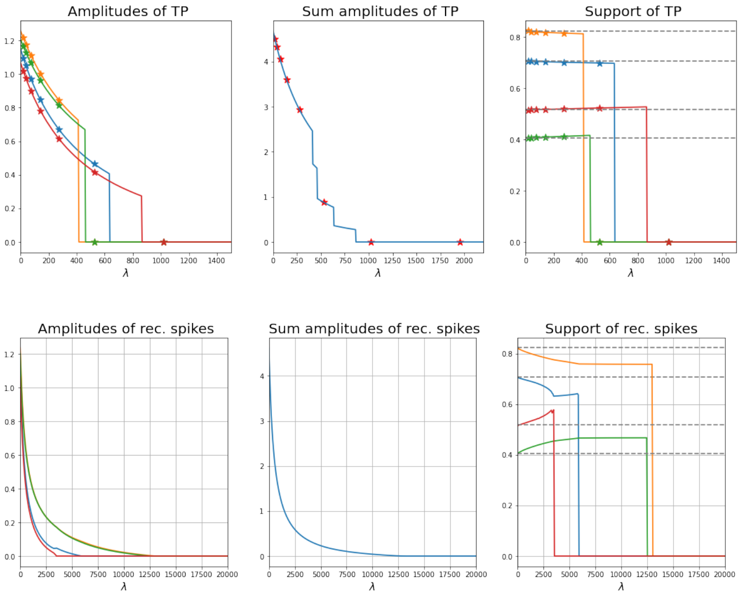
Algorithm 3: Sliding Frank Wolfe Algorithm for $\mathcal{D}_{KL} - |\mu|$

- **Input:** initialisation $\mu_{[0]}=0$ and maximum number of iterations $K_{\max} \in \mathbb{N}$
- **Repeat:** For $0 \leq k \leq K_{\max}$
  1. **Insertion step:**
  $$x_*^{[k]} \in \underset{x \in \mathcal{X}}{\text{argmax}} \left(\eta^{[k]}(x)\right)_+ \text{ with } \eta^{[k]}(x) = \frac{1}{\lambda} \Phi^* \left(I - \frac{y}{\Phi \mu_{[k]} + b} \right)$$
     - **If** $|\eta^{[k]}(x_*^{[k]})| < 1$, then $\mu_{[k]}$ is the solution of $\mathcal{D}_{KL} - |\mu|$ problem
     - **Else** $\mu_{[k]} = \sum_{i=1}^{k} a_i^{[k]} \delta_{x_i^{[k]}}$ has to be updated
  2. **Update positions and amplitudes:**
  $$x^{[k+1/2]} = (x_1^{[k]}, \cdots, x_k^{[k]}, x_*^{[k]})$$
  $$a^{[k+1/2]} \in \underset{a \in \mathbb{R}^{k+1}}{\text{argmin}}\, \mathcal{D}_{KL}(\Phi_{x^{[k+1/2]}}(a) + b, y) + \lambda \|a\|_1$$
  3. **Sliding step:**
  $$(a^{[k+1]}, x^{[k+1]}) \in \underset{(a,x) \in \mathbb{R}^{k+1} \times \mathcal{X}^{k+1}}{\text{argmin}} \mathcal{D}_{KL}(\Phi_x(a) + b, y) + \lambda \|a\|_1$$
- **Until:** $k = K_{\max}$;
- **Output:**
  - $K \in \mathbb{N}$ number of spikes;
  - $a \in \mathbb{R}^K$ s.t. $a_i > 0$ for $i = 1, \cdots, K$ amplitudes;
  - $x \in \mathcal{X}^K$ positions of the spikes.

Regularisation path for $\ell_2 - |\mu|$



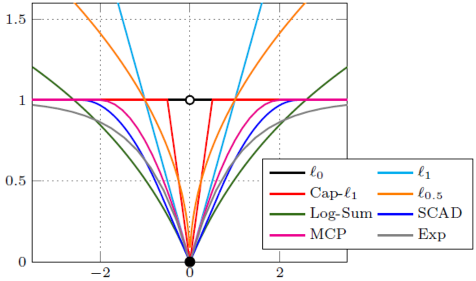
Regularisation path for $\mathcal{D}_{KL} - |\mu|$



On-the-grid Super-Resolution

$$\arg\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2 + \lambda\|\mathbf{x}\|_0$$

- ℓ_2**-fidelity** corresponds to a **Gaussian noise** formulation
- **Penalty** to enforce **sparsity** in the reconstruction: ℓ_0-norm $\|\mathbf{x}\|_0 = \#\{j|\ x_j \neq 0\}$



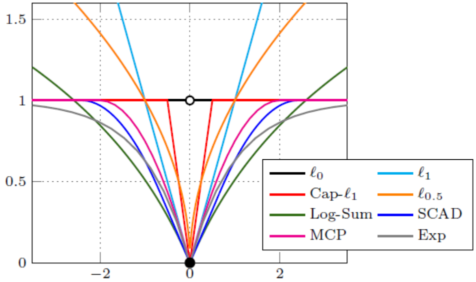
Relaxations of ℓ_0-norm

- ℓ_1-norm: $\|\mathbf{x}\|_1 = \sum_{i=1}^N |x_i|$
- $\phi_{capped-\ell_1}(\mathbf{x},\theta) = \min\{\theta|\mathbf{x}|, 1\}, \theta > 0$
- ℓ_p-norm: $\|\mathbf{x}\|_p = \left(\sum_{i=1}^N |x_i|^p\right)^{1/p}, 0 < p < 1$
- $\phi_{log-sum}(\mathbf{x},\delta) = \log(\delta + |\mathbf{x}|),\ \delta > 0$

On-the-grid Super-Resolution

$$\arg\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2 + \lambda\|\mathbf{x}\|_0$$

- ℓ_2**-fidelity** corresponds to a **Gaussian noise** formulation
- **Penalty** to enforce **sparsity** in the reconstruction: ℓ_0-norm $\|\mathbf{x}\|_0 = \#\{j|\ x_j \neq 0\}$



Relaxations of ℓ_0-norm

- $\phi_{MCP}(\mathbf{x},\lambda,\beta) = \lambda\left(\frac{\beta\lambda}{2}\mathbb{1}_{\{|\mathbf{x}|>\beta\lambda\}} + \left(|\mathbf{x}| - \frac{\mathbf{x}^2}{2\beta\lambda}\right)\mathbb{1}_{\{|\mathbf{x}|\leq\beta\lambda\}}\right)$, $\theta > 0,\ \lambda > 0$
- $\phi_{SCAD}(\mathbf{x},\lambda,a) = \begin{cases} \lambda|\mathbf{x}| & \text{if } |\mathbf{x}| \leqslant \lambda \\ -\frac{\lambda^2-2a\lambda|\mathbf{x}|+\mathbf{x}^2}{2(a-1)} & \text{if } \lambda < |\mathbf{x}| \leq a\lambda,\ a > 2,\ \lambda > 0 \\ \frac{(a+1)\lambda^2}{2} & \text{if } |\mathbf{x}| > a\lambda \end{cases}$
- $\phi_{CEL0}(\mathbf{x},\lambda,a) = \lambda - \frac{a^2}{2}\left(|\mathbf{x}| - \frac{\sqrt{2\lambda}}{a}\right)^2 \mathbb{1}_{\left\{|\mathbf{x}|\leqslant\frac{\sqrt{2\lambda}}{a}\right\}}, \lambda > 0, a > 0$

Sudifferential of the penalty

If $\mu = \sum_{i=1}^{N} a_i \delta_{x_i}$ with $a_i > 0$, $x_i \in \mathcal{X}$ we have

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

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

Thus,

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

Fenchel duality

For $\Lambda : V \to Y$ linear, $F : V \to \mathbb{R}$ and $G : Y \to \mathbb{R}$ convex, the primal problem

$$\arg\min_{u \in V} F(u) + G(\Lambda u) \tag{Primal}$$

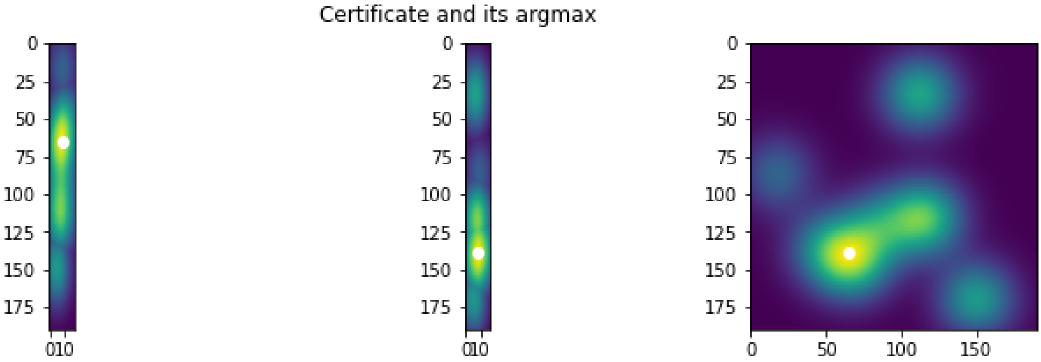
has a dual problem which reads

$$\arg\max_{p^* \in Y^*} -F^*(\Lambda^* p^*) - G^*(-p^*) \tag{Dual}$$

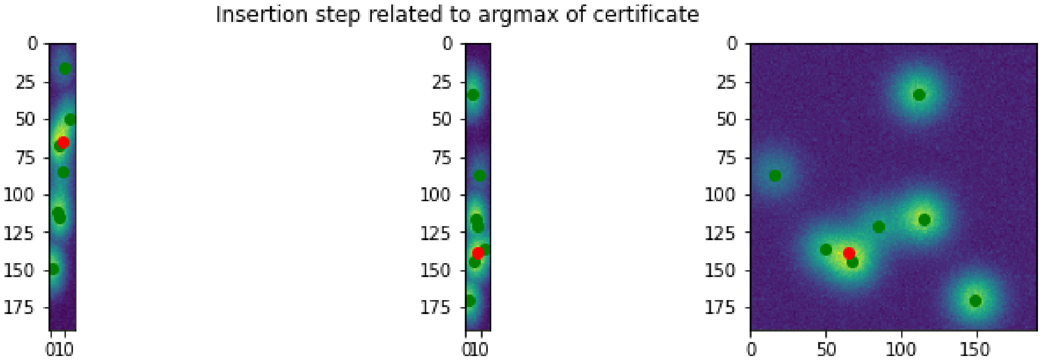
Moreover, if $u \in V$ and $p^* \in Y^*$ are respectively solutions of the primal and dual, the following **extremality conditions** hold:

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

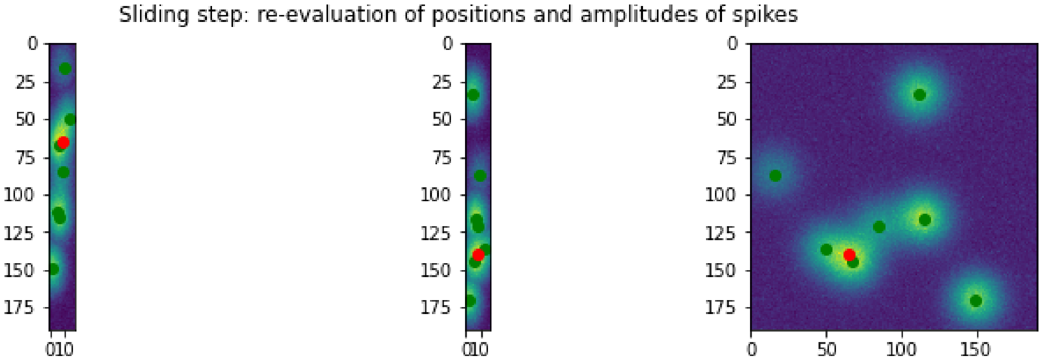
SFW algorithm



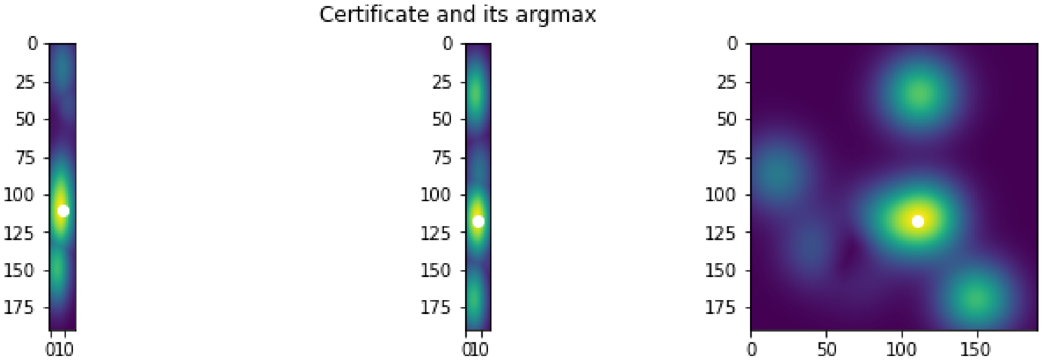
SFW algorithm



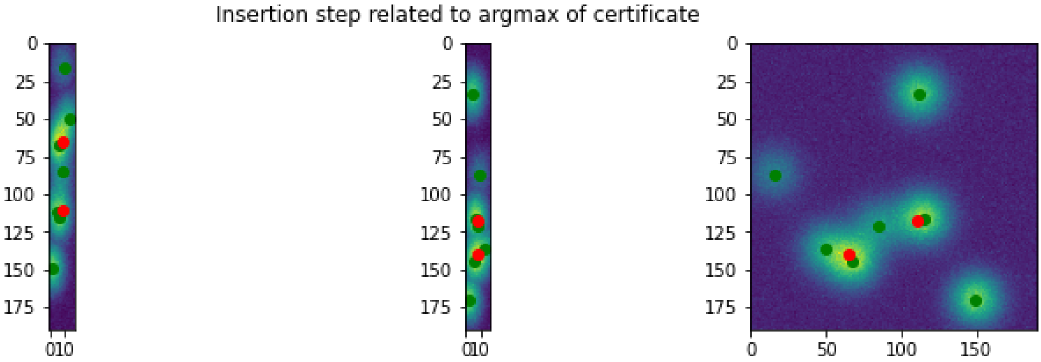
SFW algorithm



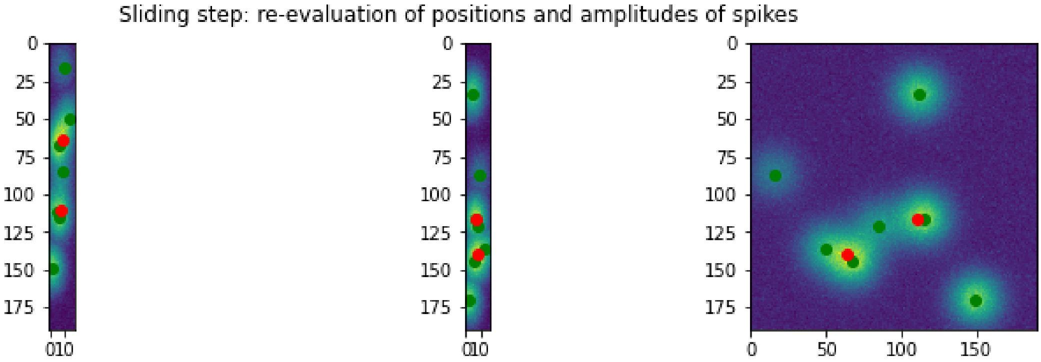
SFW algorithm



SFW algorithm



SFW algorithm



SFW algorithm