Exploiting the low rank property in off-the-grid sparse super-resolution

Paul Catala
ENS

Vincent Duval
INRIA Paris
(MOKAPLAN)

Gabriel Peyré
CNRS/ ENS

CIMI Workshop
November 7th, 2019

Outline

Summary

Inverse problems for imaging

Measuring devices have a non sharp impulse response: our observations are **blurred** of a "true ideal scene".

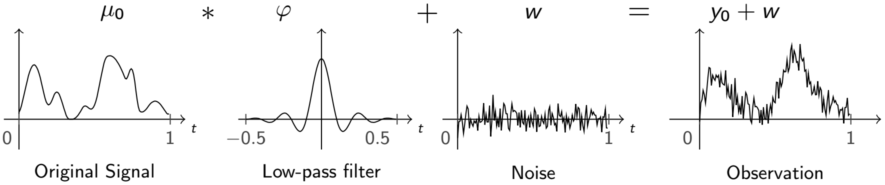
- Geophysics,
- Astronomy,
- Microscopy,
- Spectroscopy,
- ...

Image courtesy of S. Ladjal

Goal: Obtain as much detail as we can from given measurements.

The Deconvolution Problem

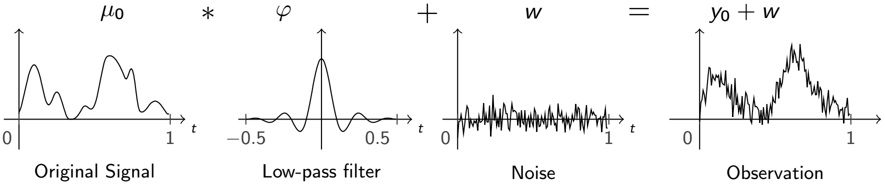
- Consider a signal μ_0 defined on $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$ (i.e. $[0,1)^d$ with periodic boundary condition).
- Perturbation model:



- Goal: recover μ_0 from the observation $y_0 + w = \varphi * \mu_0 + w$ (or simply $y_0 = \varphi * \mu_0$)

The Deconvolution Problem

- Consider a signal μ_0 defined on $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$ (i.e. $[0,1)^d$ with periodic boundary condition).
- Perturbation model:



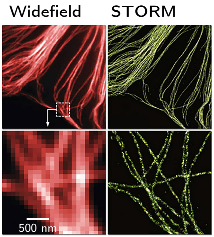
- Goal: recover μ_0 from the observation $y_0 + w = \varphi * \mu_0 + w$ (or simply $y_0 = \varphi * \mu_0$)
- Ill-posed problem:
  - the low pass filter might not be invertible ($\hat{\varphi}_n = 0$ for some frequency n)
  - even though, the problem is ill-conditioned ($|\hat{\varphi}_n| \ll |\hat{\varphi}_0|$ for high frequencies n)

The Deconvolution Problem

Assumption: the signal μ_0 is **sparse**.

In other words, we want to recover *point sources* (amplitudes and locations)

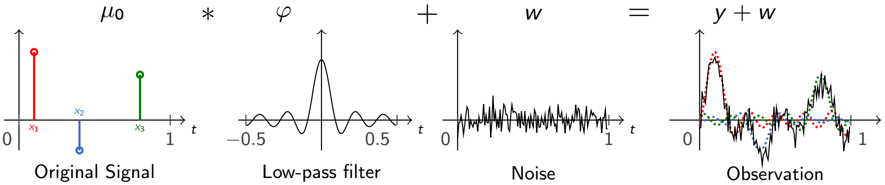
- Spectral estimation,
- Seismic imaging,
- EEG,
- Direction of Arrival,
- Super-resolution microscopy (PALM/STORM)
- ...



[www.microscopyu.com]

The Deconvolution Problem

Assumption: the signal μ_0 is **sparse**.



$$\mu_0 = \sum_{i=1}^{N} a_i \delta_{x_i}, \quad \text{where} \begin{cases} a_i \in \mathbb{C}, \\ x_i \in \mathbb{T}, \\ N \in \mathbb{N} \text{ is small.} \end{cases}$$

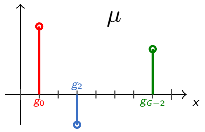
so that we observe $y + w = \sum_{i=1}^{N} a_i \varphi(\cdot - x_i) + w$.

Idea: Look for a **sparse** signal μ such that $\varphi * \mu \approx y_0 + w$ (or y_0).

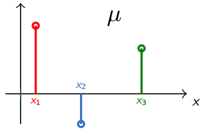
Possible approaches

.

- Define a grid $\mathcal{G} = \{g_i \;:\; 0 \leqslant i \leqslant G-1\}$ an try to recover a signal of the form $\mu = \sum_{i=0}^{G-1} a_i \delta_{g_i}$ using LASSO or Matching Pursuit...
  - Well understood algorithms
  - Large and ill-conditioned problems when using thin grids
  - Discretization artifacts, basis mismatch



- Use a fully continuous approach (Prony, MUSIC, **Beurling LASSO**)
  - Nice theoretical properties
  - Numerical resolution not straightforward



Towards the continuous approach

Define the **total variation of the measure** $m \in \mathcal{M}(\mathbb{T}^d)$ as:

$$|\mu|(\mathbb{T}^d) = \sup \left\{ \mathcal{R}\mathrm{e} \left(\int_{\mathbb{T}^d} \psi^* dm \right) ; \psi \in \mathscr{C}(\mathbb{T}^d, \mathbb{C}), \|\psi\|_\infty \leqslant 1 \right\}$$

Example :

- If $\mu = \sum_{i=1}^r a_i \delta_{x_i}$, then $|\mu|(\mathbb{T}^d) = \sum_{i=1}^M |a_i|$.
- If $\mu = fd\mathcal{L}$, then $|\mu|(\mathbb{T}^d) = \int_{\mathbb{T}^d} |f(t)| \mathrm{d}t$.

Rationale: the extreme points of $\{\mu \in \mathcal{M}(\mathbb{T}^d), |\mu|(\mathbb{T}^d) \leqslant 1\}$ are the Dirac masses: $\alpha\delta_x$ for $x \in \mathbb{T}^d$, $|\alpha| = 1$.

Continuous sparse recovery

Given a linear observation operator $\Phi : \mathcal{M}(\mathbb{T}^d) \to \mathbb{C}^M$, consider

- **Basis Pursuit for measures** [de Castro & Gamboa (12), Candès & Fernandez-Granda (13)],

$$\inf_{\mu \in \mathcal{M}(\mathbb{T}^d)} |\mu|(\mathbb{T}^d) \text{ such that } \Phi\mu = y_0 \qquad (\mathcal{P}_0(y_0))$$

- **LASSO for measures, or BLASSO** [Recht et al. (12), Bredies & Pikkarainen (13), Azais et al. (13)]

$$\inf_{\mu \in \mathcal{M}(\mathbb{T}^d)} \lambda|\mu|(\mathbb{T}^d) + \frac{1}{2}\left\|\Phi\mu - (y_0 + w)\right\|^2 \qquad (\mathcal{P}_\lambda(y_0 + w))$$

Observation framework

For the rest of the talk, we assume that Φ is a **partial Fourier operator**

$$\Phi\mu = \mathcal{F}_{\Omega_c}\mu, \quad \text{where} \quad \Omega_c = \left\{ j \in \mathbb{N}^d \;:\; \|j\|_\infty \leqslant f_c \right\},$$

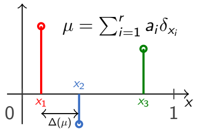
$$(\mathcal{F}_{\Omega_c}\mu)_j \stackrel{\text{def.}}{=} \int_{\mathbb{T}^d} e^{-2\mathrm{i}\pi\langle j, x\rangle} \mathrm{d}\mu(x).$$

- Ideal Low-Pass Filter (convolution w/ Dirichlet kernel), spectral estimation,
- Extensions to more general observation operators are possible.

Identifiability for discrete measures

Minimum separation distance of μ:

$$\Delta(\mu) = \min_{\substack{x,x' \in \operatorname{Supp}\mu, \\ x \neq x'}} \|x - x'\|_\infty$$



Theorem (Candès & Fernandez-Granda (2013))

There exists a constant $C_d > 0$ such that, for any (discrete) measure μ_0 with $\Delta(\mu_0) \geqslant \frac{C_d}{f_c}$, μ_0 is the unique solution of

$$\inf_{\mu \in \mathcal{M}(\mathbb{T}^d)} |\mu|(\mathbb{T}^d) \text{ such that } \Phi\mu = y_0 \qquad (\mathcal{P}_0(y_0))$$

where $y_0 = \Phi\mu_0$.

Remark: $1 \leqslant C \leqslant 1.26$ for $d = 1$.

Robustness

Question: if w is small and $\lambda > 0$ is small, can we recover a solution $\mu \approx \mu_0$ where $y_0 = \Phi\mu_0$?

$$\inf_{\mu \in \mathcal{M}(\mathbb{T}^d)} \lambda|\mu|(\mathbb{T}^d) + \frac{1}{2}\left\|\Phi\mu - (y_0 + w)\right\|^2 \qquad (\mathcal{P}_\lambda(y_0 + w))$$

☞ **Yes,** provided μ_0 is the unique solution to

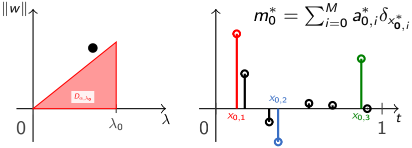
$$\inf_{\mu \in \mathcal{M}(\mathbb{T}^d)} |\mu|(\mathbb{T}^d) \text{ such that } \Phi\mu = y_0 \qquad (\mathcal{P}_0(y_0))$$

(+ technical conditions)

- Weak-* convergence results [Bredies & Pikkarainen (13)],
- Estimation on the local averages of μ [Azais et al. (13), Fernandez-Granda (13)].

Robustness

Consider an input measure
$\mu_0 = \sum_{i=1}^r a_{0,i}\delta_{x_{0,i}}$



Theorem (D.-Peyré'15)

*Assume that μ_0 is **"non-degenerate"**.*
Then there exists, $\alpha > 0$, $\lambda_0 > 0$ such that for $0 \leqslant \lambda \leqslant \lambda_0$ and $\|w\| \leqslant \alpha\lambda$,

- *the solution $\mu_{(\lambda,w)}$ to $\mathcal{P}_\lambda(y+w)$ is unique and has exactly r spikes, $\mu_{(\lambda,w)} = \sum_{i=1}^r a_i(\lambda,w)\delta_{x_i(\lambda,w)}$,*
- *the mapping $(\lambda, w) \mapsto (a, x)$ is $\mathscr{C}^1$.*
- *the solution has the Taylor expansion*

$$\begin{pmatrix} a(\lambda,w) \\ x(\lambda,w) \end{pmatrix} = \begin{pmatrix} a_0 \\ x_0 \end{pmatrix} + \begin{pmatrix} I & 0 \\ 0 & \mathrm{diag} a_0^{-1} \end{pmatrix} (\Gamma_{x_0}^* \Gamma_{x_0})^{-1} \left[\begin{pmatrix} \mathrm{sign}(a_0) \\ 0 \end{pmatrix} \lambda - \Gamma_{x_0}^* w \right] + o \begin{pmatrix} \lambda \\ w \end{pmatrix}$$

Summary

Numerical methods for the BLASSO

- Discretization of the domain + proximal algorithm [Donoho'92,. . .]
- Greedy method [Bredies & Pikkarainen'13, Boyd et al.'15]
- Moment-Sum of Squares hierarchies (following [Lasserre'00])
  - In [De Castro et al.'17, Josz et al. '17]: a relaxation method taylored for real-valued measures.
  - We use a relaxation for **complex-valued measures** μ. Based on the reformulation [Tang et al. '13] in the 1D-case.

Reformulation

$$\min_{\mu \in \mathcal{M}(\mathbb{T}^d)} \lambda \left|\mu\right| \left(\mathbb{T}^d\right) + \frac{1}{2} \left\| y - \mathcal{F}_{\Omega_c} \mu \right\|^2$$

Reformulation

$$\min_{z \in \mathbb{C}^{(2f_c+1)^d}} \frac{1}{2} \|y - z\|^2 + \lambda \Big(\min_{\mu \in \mathcal{M}(\mathbb{T}^d)} |\mu|(\mathbb{T}^d) \quad \text{s.t.} \quad (\mathcal{F}\mu)_k = z_k \quad \forall k \in \Omega_c \Big).$$

It is sufficient to study the problem

$$\min_{\mu \in \mathcal{M}(\mathbb{T}^d)} |\mu|(\mathbb{T}^d) \quad \text{s.t.} \quad (\mathcal{F}\mu)_k = z_k \quad \forall k \in \Omega_c = [\![-f_c, f_c]\!]^d \qquad (\mathcal{Q}_0(z))$$

Moment based relaxation (motivation)

Let $\nu = |\mu|$ and consider its moment matrix $\mathbb{M}_\ell[\nu]$,

$$\forall i, j \in [\![-\ell, \ell]\!]^d, \quad (\mathbb{M}_\ell[\nu])_{i,j} = \int_{\mathbb{T}^d} e^{-2\mathrm{i}\pi\langle i, x\rangle} e^{2\mathrm{i}\pi\langle j, x\rangle} \mathrm{d}\nu(x)$$

Then,

- $\mathbb{M}_\ell[\nu]$ is **positive semi-definite** ($\mathbb{M}_\ell[\nu] \succeq 0$).

$$\forall q \in \mathbb{C}^{(2\ell+1)^d}, \; q^* \mathbb{M}_\ell[\nu] q = \int_{\mathbb{T}^d} \left(\sum_{\|i\|_\infty \leqslant \ell} q_i e^{2\mathrm{i}\pi\langle i, x\rangle} \right)^* \left(\sum_{\|j\|_\infty \leqslant \ell} q_j e^{2\mathrm{i}\pi\langle j, x\rangle} \right) \mathrm{d}\nu(x)$$

$$= \int_{\mathbb{T}^d} \left| \sum_{\|j\|_\infty \leqslant \ell} q_j e^{2\mathrm{i}\pi\langle j, x\rangle} \right|^2 \mathrm{d}\nu(x) \geqslant 0.$$

Moment based relaxation (motivation)

Let $\nu = |\mu|$ and consider its moment matrix $\mathbb{M}_\ell[\nu]$,

$$\forall i, j \in [\![-\ell, \ell]\!]^d, \quad (\mathbb{M}_\ell[\nu])_{i,j} = \int_{\mathbb{T}^d} e^{-2\mathrm{i}\pi\langle i, x\rangle} e^{2\mathrm{i}\pi\langle j, x\rangle} \mathrm{d}\nu(x)$$

Then,

- $\mathbb{M}_\ell[\nu]$ is **positive semi-definite** ($\mathbb{M}_\ell[\nu] \succeq 0$).
- $\mathbb{M}_\ell[\nu]$ is **multi-level Toeplitz**, a.k.a. Toeplitz-Block-Toeplitz ($\mathbb{M}_\ell[\nu] \in \mathcal{T}_\ell$).

$$\begin{aligned}(\mathbb{M}_\ell[\nu])_{i+k,j+k} &= \int_{\mathbb{T}^d} e^{-2\mathrm{i}\pi\langle i+k, x\rangle} e^{2\mathrm{i}\pi\langle j+k, x\rangle} \mathrm{d}\nu(x) \\ &= \int_{\mathbb{T}^d} e^{-2\mathrm{i}\pi\langle i, x\rangle} e^{2\mathrm{i}\pi\langle j, x\rangle} \mathrm{d}\nu(x) = (\mathbb{M}_\ell[\nu])_{i,j}\end{aligned}$$

for all i, j, k such that $\|i\|_\infty \leqslant \ell$, $\|j\|_\infty \leqslant \ell$, $\|i+k\|_\infty \leqslant \ell$, $\|j+k\|_\infty \leqslant \ell$.

Moment based relaxation (motivation)

Let $\nu = |\mu|$ and consider its moment matrix $\mathbb{M}_\ell[\nu]$,

$$\forall i,j \in [\![-\ell,\ell]\!]^d, \quad (\mathbb{M}_\ell[\nu])_{i,j} = \int_{\mathbb{T}^d} e^{-2\mathrm{i}\pi\langle i,x\rangle} e^{2\mathrm{i}\pi\langle j,x\rangle} \mathrm{d}\nu(x)$$

Then,

- $\mathbb{M}_\ell[\nu]$ is **positive semi-definite** ($\mathbb{M}_\ell[\nu] \succeq 0$).
- $\mathbb{M}_\ell[\nu]$ is **multi-level Toeplitz**, a.k.a. Toeplitz-Block-Toeplitz ($\mathbb{M}_\ell[\nu] \in \mathcal{T}_\ell$).
- $\tau\mathbb{M}_\ell[\nu] - \tilde{z}\tilde{z}^* \succeq 0$ where $\tau = \nu(\mathbb{T}^d) = (\mathbb{M}_\ell[\nu])_{(0,0)}$ and $\tilde{z} = \mathcal{F}_{[\![-\ell,\ell]\!]^d}\mu$.

$$\begin{aligned}
\forall q \in \mathbb{C}^{(2\ell+1)^d}, \quad q^*(zz^*)q &= \left| \int_{\mathbb{T}^d} \sum_{\|j\|_\infty \leqslant \ell} q_j e^{2\mathrm{i}\pi\langle j,x\rangle} \mathrm{d}\mu(x) \right|^2 \\
&\leqslant \left(\int_{\mathbb{T}^d} \left| \sum_{\|j\|_\infty \leqslant \ell} q_j e^{2\mathrm{i}\pi\langle j,x\rangle} \right|^2 \mathrm{d}\,|\mu|\,(x) \right) \left(\int_{\mathbb{T}^d} 1^2 \mathrm{d}\,|\mu| \right) \\
&= (q^*\mathbb{M}_\ell[\nu]q) \times \tau.
\end{aligned}$$

Moment based relaxation (motivation)

Let $\nu = |\mu|$ and consider its moment matrix $\mathbb{M}_\ell[\nu]$,

$$\forall i, j \in [\![-\ell, \ell]\!]^d, \quad (\mathbb{M}_\ell[\nu])_{i,j} = \int_{\mathbb{T}^d} e^{-2i\pi\langle i, x\rangle} e^{2i\pi\langle j, x\rangle} \mathrm{d}\nu(x)$$

Then,

- $\mathbb{M}_\ell[\nu]$ is **positive semi-definite** ($\mathbb{M}_\ell[\nu] \succeq 0$).
- $\mathbb{M}_\ell[\nu]$ is **multi-level Toeplitz**, a.k.a. Toeplitz-Block-Toeplitz ($\mathbb{M}_\ell[\nu] \in \mathcal{T}_\ell$).
- $\tau\mathbb{M}_\ell[\nu] - \tilde{z}\tilde{z}^* \succeq 0$ where $\tau = \nu(\mathbb{T}^d) = (\mathbb{M}_\ell[\nu])_{(0,0)}$ and $\tilde{z} = \mathcal{F}_{[\![-\ell,\ell]\!]^d}\mu$.

In other words, $\begin{pmatrix} \mathbb{M}_\ell[\nu] & \tilde{z} \\ \tilde{z}^* & \tau \end{pmatrix} \succeq 0.$

Moment based relaxation

Given $z \in \mathbb{C}^{(2f_c+1)^d}$, consider the problem on measures

$$\min_{\mu \in \mathcal{M}(\mathbb{T}^d)} |\mu|(\mathbb{T}^d) \quad \text{s.t.} \quad (\mathcal{F}\mu)_k = z_k, \ \forall k \in \Omega_c = [\![-f_c, f_c]\!]^d \qquad (\mathcal{Q}_0(z))$$

or the semi-definite program ($\ell \geqslant f_c$)

$$\min_{\substack{R \succeq 0, \\ \tilde{z} \in \mathbb{C}^{(2\ell+1)^d}}} \left(\frac{1}{(2\ell+1)^d} \operatorname{Tr}(R) + \tau \right) \quad \text{s.t.} \quad \begin{cases} \forall k \in \Omega_c, & \tilde{z}_k = z_k, \\ \begin{pmatrix} R & \tilde{z} \\ \tilde{z}^* & \tau \end{pmatrix} & \succeq 0, \\ R & \in \mathcal{T}_\ell. \end{cases} \qquad (\mathcal{Q}_0^{(\ell)}(z))$$

Proposition (same as [Lasserre '00])

$$\min \mathcal{Q}_0^{(\ell)}(z) \leqslant \min \mathcal{Q}_0^{(\ell+1)}(z) \leqslant \min \mathcal{Q}_0(z)$$

$$\textit{and} \ \lim_{\ell \to +\infty} \left(\min \mathcal{Q}_0^{(\ell)}(z) \right) = (\min \mathcal{Q}_0(z))$$

Flatness criterion

We say that R is **flat** if $\operatorname{rank}\left([R]_{[\![-\ell+1,\ell-1]\!]^d}\right) = \operatorname{rank} R$.

Proposition

If R is flat, then R has a representing measure: $R = \mathbb{M}_\ell[\nu]$ for some measure $\nu \geqslant 0$. Moreover $\operatorname{card}\operatorname{Supp}(\nu) = \operatorname{rank}(R)$.

Note: Similar to [Curto & Fialkow'96], but the degree is

$$\deg_\infty(i) = \max(|i_1|, \ldots |i_d|)$$

instead of

$$\deg_1(i) = |i_1| + \ldots + |i_d|.$$

☞ we rely on [Laurent & Mourrain'09] for flat extensions with general monomial sets.

Remark: For $d = 1$, R already has a representing measure for $\ell = f_c$.

Tightness of the relaxation

Let $(R, \tilde{z})$ be a solution to

$$\min_{\substack{R\succeq 0,\\ \tilde{z}\in\mathbb{C}^{(2\ell+1)^d}}} \left(\frac{1}{(2\ell+1)^d}\operatorname{Tr}(R)+\tau\right) \quad \text{s.t.} \quad \begin{cases} \forall k\in\Omega_c, & \tilde{z}_k = z_k, \\ \begin{pmatrix} R & \tilde{z} \\ \tilde{z}^* & \tau \end{pmatrix} & \succeq 0, \\ R & \in \mathcal{T}_\ell. \end{cases} \qquad (\mathcal{Q}_0^{(\ell)}(z))$$

Proposition

Assume that R is flat, and let $\nu \geqslant 0$ s.t. $R = \mathbb{M}_\ell[\nu]$. Then, there exists $\mu \in \mathcal{M}(\mathbb{T}^d)$, such that

- $\operatorname{card}\operatorname{Supp}(\mu) = \operatorname{rank}(R)$,
- $\tilde{z} = \mathcal{F}_{[\![-\ell,\ell]\!]^d}\mu$, *and* $\nu = |\mu|$.
- $\min \mathcal{Q}_0^{(\ell)}(z) = \min \mathcal{Q}_0(z)$ *and μ is a solution to $\mathcal{Q}_0(z)$.*

Conversely, if μ is a solution to $\mathcal{Q}_0(z)$ and $\min \mathcal{Q}_0^{(\ell)}(z) = \min \mathcal{Q}_0(z)$, then $(\mathbb{M}_\ell[|\mu|], \mathcal{F}_{[\![-\ell,\ell]\!]^d}\mu)$ is a solution to $\mathcal{Q}_0^{(\ell)}(z)$.

Summary

What we have seen so far

We want to solve the relaxation of the BLASSO:

$$\min_{\substack{R\succeq 0,\\ \tilde{z}\in\mathbb{C}^{(2\ell+1)^d}}} \left(\lambda\left(\frac{1}{(2\ell+1)^d}\operatorname{Tr}(R)+\tau\right)+\frac{1}{2}\left\|y-\tilde{z}_{[\![-f_c,f_c]\!]^d}\right\|^2\right)$$

$$\text{s.t.} \quad \begin{cases} \begin{pmatrix} R & \tilde{z} \\ \tilde{z}^* & \tau \end{pmatrix} \succeq 0, \\ R\in\mathcal{T}_\ell. \end{cases} \qquad (\mathcal{Q}_\lambda^{(\ell)}(y))$$

That SDP has a large size ($m \overset{\text{def.}}{=} (2\ell+1)^d+1$). But...

- R has low rank (sparsity of μ_λ, if the relaxation is tight)
- R has the (multi-level) Toeplitz property

What we have seen so far

We want to solve the relaxation of the BLASSO:

$$\min_{\substack{R\succeq 0,\\ \tilde{z}\in\mathbb{C}^{(2\ell+1)^d}}} \left(\lambda\left(\frac{1}{(2\ell+1)^d}\operatorname{Tr}(R)+\tau\right)+\frac{1}{2}\left\|y-\tilde{z}_{[\![-f_c,f_c]\!]^d}\right\|^2\right)$$

$$\text{s.t.}\quad \begin{cases} \begin{pmatrix} R & \tilde{z} \\ \tilde{z}^* & \tau \end{pmatrix} \succeq 0, \\ R \in \mathcal{T}_\ell. \end{cases} \qquad (\mathcal{Q}_\lambda^{(\ell)}(y))$$

That SDP has a large size ($m \stackrel{\text{def.}}{=} (2\ell+1)^d+1$). But...

- R has low rank (sparsity of μ_λ, if the relaxation is tight)
- R has the (multi-level) Toeplitz property

We use

- a **conditional gradient / Frank-Wolfe** algorithm to exploit the low rank property.
- the **Fast Fourier Transform** in the calculations involving the Toeplitz matrix R.

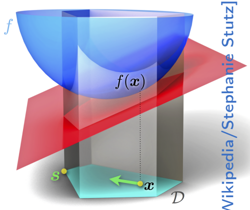
The Frank-Wolfe algorithm

Goal: Minimize a convex differentiable function f on a compact convex set $\mathcal{D} \subset \mathbb{R}^P$

Algorithm (Frank-Wolfe/Conditional gradient)

For all $k \in \mathbb{N}$, iterate

1. Linear minimization:
   $s_k \in \operatorname{argmin}_{s\in\mathcal{D}} f(x_k) + \langle \nabla f(x_k), s - x_k \rangle$
2. Line search: $x_{k+1} \in \operatorname{argmin}_{x\in[x_k,s_k]} f(x)$



[Wikipedia/Stephanie Stutz]

Remarks:

- If ∇f is Lipschitz, $f(x_k) - \min_{\mathcal{D}} f = O\left(\frac{1}{k}\right)$.
- At each step, $x_k \in \operatorname{conv}(x_0, s_1, \dots, s_{k-1})$.
- In step 2, one may choose $x_{k+1} \in \mathcal{D}$ with $f(x_{k+1}) \leqslant \min_{x\in[x_k,s_k]} f(x)$
- Minimization of a linear form: OK if we can handle the **extreme points** of $\mathcal{D}$.

  ☹ What are the extreme point of $\mathcal{T}_\ell \cap \{R \succeq 0\}$?

We truncate the PSD cone (w.l.o.g.), and we penalize the Toeplitz constraint

$$\min_{\substack{R \succeq 0, \\ \tilde{z} \in \mathbb{C}^{(2\ell+1)^d}}} \left(\lambda \left(\frac{1}{(2\ell+1)^d} \operatorname{Tr}(R) + \tau \right) + \frac{1}{2} \left\| y - \tilde{z}_{[\![-f_c, f_c]\!]^d} \right\|^2 + \frac{1}{2\rho} \| R - P_{\mathcal{T}_\ell} R \|^2 \right)$$

$$\text{s.t.} \quad \begin{cases} \begin{pmatrix} R & \tilde{z} \\ \tilde{z}^* & \tau \end{pmatrix} \succeq 0, \\ \frac{1}{(2\ell+1)^d} \operatorname{Tr} R + \tau \leqslant C \end{cases}$$

We truncate the PSD cone (w.l.o.g.), and we penalize the Toeplitz constraint

$$\min_{\substack{R\succeq 0,\\ \tilde{z}\in\mathbb{C}^{(2\ell+1)^d}}} \left(\lambda\left(\frac{1}{(2\ell+1)^d}\operatorname{Tr}(R)+\tau\right)+\frac{1}{2}\left\|y-\tilde{z}_{[\![-f_c,f_c]\!]^d}\right\|^2+\frac{1}{2\rho}\|R-P_{\mathcal{T}_\ell}R\|^2\right)$$

$$\text{s.t.}\quad \begin{cases} \begin{pmatrix} R & \tilde{z} \\ \tilde{z}^* & \tau \end{pmatrix} \succeq 0, \\ \frac{1}{(2\ell+1)^d}\operatorname{Tr} R+\tau\leqslant C \end{cases} \iff \hat{R}\overset{\text{def.}}{=}\begin{pmatrix} R & \tilde{z} \\ \tilde{z}^* & \tau \end{pmatrix}\in K$$

☞ K is a **truncated PSD cone**. Its extreme points are 0 or of the form αuu^* where $u\in\mathbb{C}^{(2\ell+1)^d+1}$.

Consequence:

- If $\hat{R}_0=0$, at each iteration, $\hat{R}_k$ is of the form $\sum_{i=1}^{k-1}\alpha_i u_i u_i^*$.
- Instead of storing $\hat{R}_k$, we store $U_k\in\mathbb{C}^{m\times k}$ where $\hat{R}_k=U_kU_k^*$, $m=((2\ell+1)^d+1)$.

Step 1: linear minimization

At each iteration k,

$$\text{Find} \quad \underset{\hat{S}\in K}{\operatorname{argmin}} \ \operatorname{Tr}(M\hat{S}) \qquad \text{where } M \stackrel{\text{def.}}{=} \nabla f(\hat{R}_k) \in \mathcal{H}_n(\mathbb{C}).$$

- A solution is given by $\hat{S}_{k+1} = \alpha v_{k+1} v_{k+1}^*$, where v_{k+1} is obtained by **power iterations** on $M = \nabla f(\hat{R}_k)$ (up to a diagonal rescaling)
- To compute Mv:

$$\nabla f(\hat{R}_k)v = \underbrace{\left(\text{terms involving } \hat{R}_k v\right)}_{\text{use the factorization by } U_k} + \underbrace{(\text{terms involving } (P_{\mathcal{T}_\ell} R_k)v)}_{\text{use the } \textbf{Fast Fourier Transform} \text{ (FFT)}}$$

- **Complexity:** $O(k\ell^d \log \ell)$ (instead of $O(\ell^{2d})$).
- **Storage:** we only need to store variables of size $m \times k$ (instead of m^2).

Step 2: Line-search and refinement

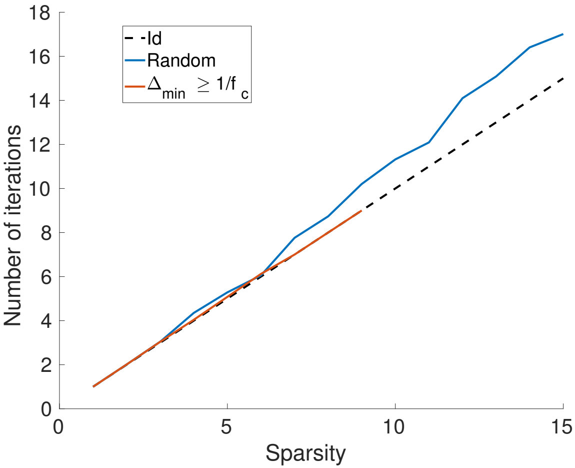
- Update $\tilde{U}_{k+1} \stackrel{\text{def.}}{=} [\alpha U_k \ (1-\alpha) v_{k+1}]$ where $\alpha \in [0,1]$ is chosen to minimize $f(\tilde{U}_{k+1}\tilde{U}_{k+1}^*)$ (closed form expression).
- **Non convex update** (as in [Boyd et al.'15, Bredies & Pikkarainen'13])

$$U_{k+1} = \mathrm{BFGS}(U \mapsto f(UU^*), U_{k+1})$$

Remarks:

- Complexity of each BFGS inner step $O(k^2\ell^d + k\ell^d \log \ell)$.
- The non convex step does not break the theoretical convergence of the algorithm.
- It improves **a lot** the practical convergence of the algorithm: convergence in r outer iterations where r is the number of Dirac masses of the solution.

Finite number of iterations



Number of outer iterations w.r.t. sparsity of solution (averaged over 200 trials)

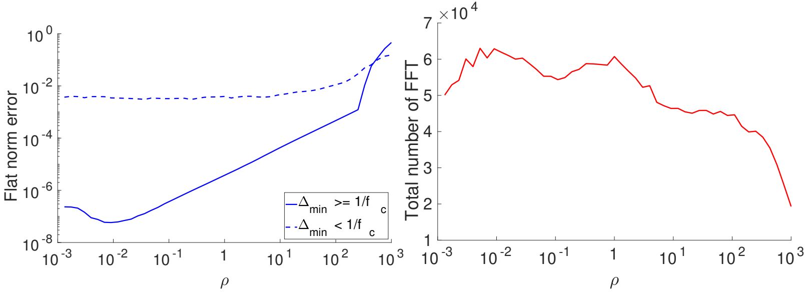
Epilogue

Once U_k (or $\hat{R}_k$) has converged, we need to recover the measure $\mu = \sum_{i=1}^{r} \alpha_i \delta_{x_i}$ from its moments. We apply the procedure described in [Lasserre'09] (see also [Harmouch et al.'17, Josz et al.'17]).

- Compute $\tilde{U}_k$, the reduced column echelon form of U_k.
- From $\tilde{U}_k$, build the "multiplication" matrices $N_1, \dots, N_d$ (they commute).
- The eigenvalues of N_j are the $e^{2\mathrm{i}\pi \langle e_j, x \rangle}$ for $x \in \operatorname{Supp} \mu$ ($e_j = (0, \dots, 1, 0, \dots, 0)$).
  $\to$ recover each $x \in \operatorname{Supp} \mu$ by jointly diagonalizing $N_1, \dots, N_d$.

Impact of the Toeplitz penalization

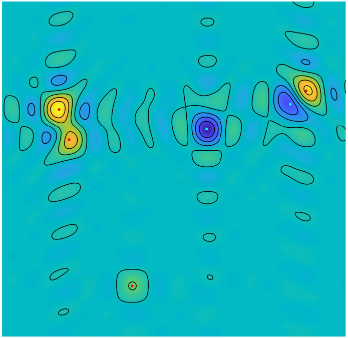
$$\lim_{\rho \to 0^+} \frac{1}{2\rho} \|R - P_{\mathcal{T}_\ell} R\|^2 = \chi_{\mathcal{T}_\ell}(R)$$

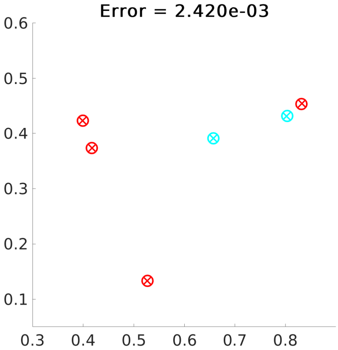


(a) Error w.r.t. MOSEK solution (b) Total number of fft wrt ρ

1D example (results averaged over $\sim$ 700 trials)

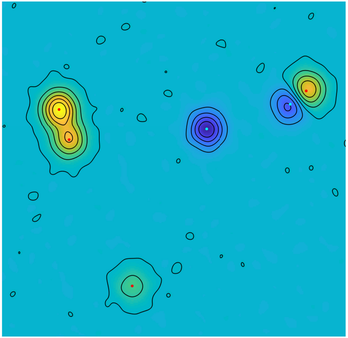
Synthetic Data - Examples

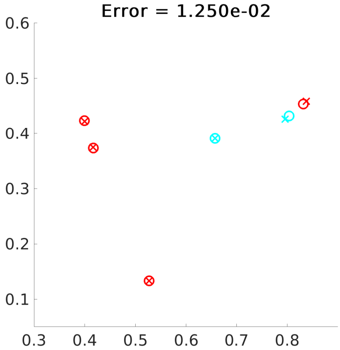




Dirichlet, $\frac{\|w\|}{\|y_0\|} = 10^{-4}$

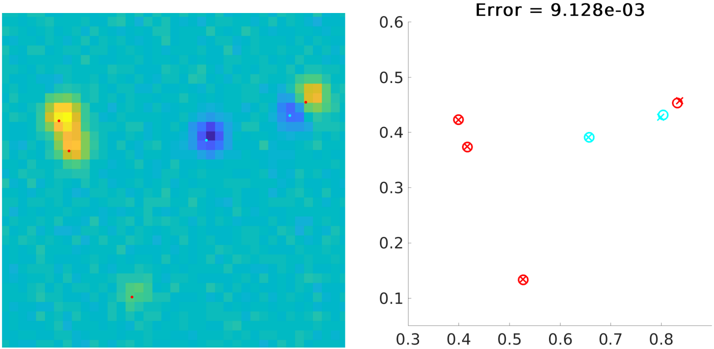
Synthetic Data - Examples





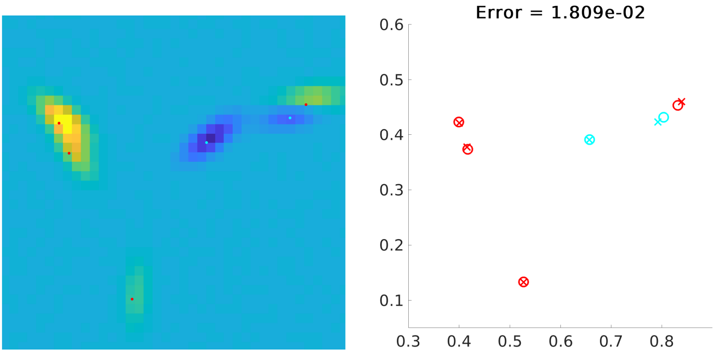
Gaussian - $f_c = 30$, $\frac{\|w\|}{\|y_0\|} = 10^{-4}$

Synthetic Data - Examples



Subsampled Gaussian - $f_c = 30$, $\frac{\|w\|}{\|y_0\|} = 10^{-3}$, $\mathcal{G} = 64 \times 64$

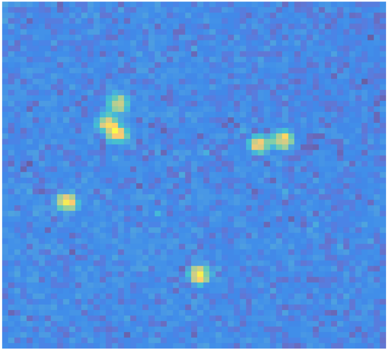
Synthetic Data - Examples

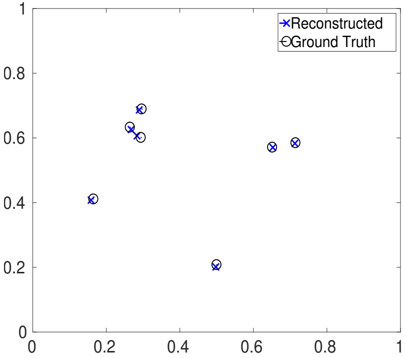


Foveation - $f_c = 30$, $\frac{\|w\|}{\|y_0\|} = 10^{-3}$, $\mathcal{G} = 64 \times 64$

SMLM Data - Example

Observation = sampled convolution
(from the microscopy challenge `http://bigwww.epfl.ch/palm`)

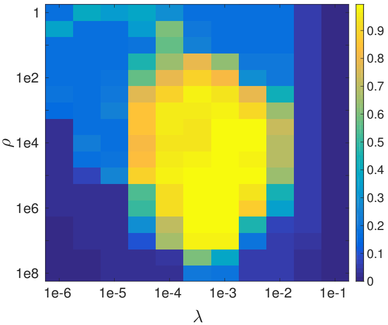




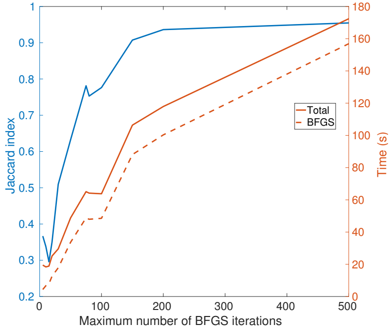
Reconstruction error: $\|x_{rec} - x_0\| \,/\, \|x_0\| = 1.57 \times 10^{-2}$

SMLM Data - Performance

$$\text{Jaccard index} \overset{\text{def.}}{=} \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive} + \text{False Negative}}$$



(c) Jaccard index wrt λ and ρ (up to normalization factors). Each pixel is obtained by averaging over 20 images.



(d) Jaccard index (blue) and time (red) wrt number of BFGS iterations. Values are averaged over 20 images.

Conclusion

- A SDP hierarchy to solve the BLASSO which yields large SDP problems...
- A fast solver which exploits
  - the low rank of the solutions
  - the Toeplitz structure of moment matrices
  - allows to solve the BLASSO in 2D for moderate f_c.
- Ongoing/future work: apply this kind of methods to the recovery of higher dimensional objects (curves...)

Thank you for your attention!

Paper:
A Low-rank Approach to Off-the-Grid Sparse Super-resolution
P. Catala, V. Duval, G. Peyré, SIIMS, 2019, Vol. 12, Issue 3.

Thank you for your attention!