A Low Rank Approach to Off-the-Grid Sparse Super-Resolution

Paul Catala [1]
Joint work with Vincent Duval [2,3] and Gabriel Peyré [1]

[1]DMA, Ecole Normale Supérieure, PSL, CNRS UMR 8553

[2]Mokaplan, Inria Paris

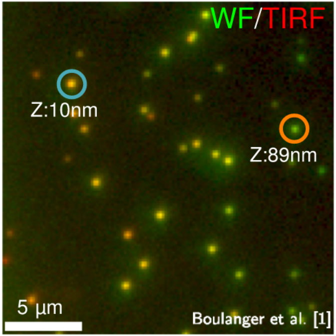
[3]CEREMADE, Université Paris-Dauphine, PSL

ITWIST, November 2018

Sparse Super-Resolution

Recover pointwise sources from low-resolution and noisy observations.

Astrophysics (2D)



Molecule fluorescence (3D)

Also neural spikes (1D), seismic imaging (1.5D), ...

Single Molecule Localization Microscopy

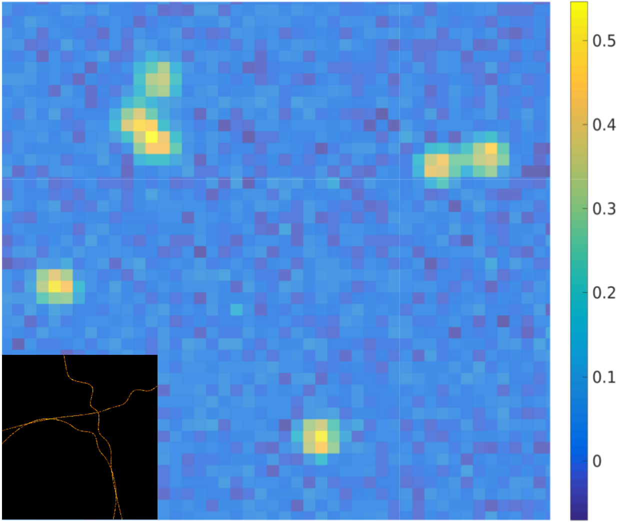


Figure: `bigwww.epfl.ch/smlm/`

Overview

Model

Low-Rank Semidefinite Relaxations

Algorithm: FFT-Based Conditional Gradient

Numerics

Model

Degradation Model

<u>**Signal to recover:**</u> discrete Radon measure on $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$:

$$\mu_0 = \sum_{i=1}^r a_i \delta_{x_i}, \quad a_i \in \mathbb{R}, \quad x_i \in \mathbb{T}^d$$

Degradation Model

Signal to recover: discrete Radon measure on $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$:

$$\mu_0 = \sum_{i=1}^r a_i \delta_{x_i}, \quad a_i \in \mathbb{R}, \quad x_i \in \mathbb{T}^d$$

Forward operator: $\Phi : \mathcal{M}(\mathbb{T}^d) \to \mathbb{C}^N$, such that

$$\Phi\mu \stackrel{\text{def.}}{=} \int_{\mathbb{T}^d} \varphi(x) \mathrm{d}\mu(x), \quad \varphi(x) = (\varphi_1(x), \dots, \varphi_N(x))^\top$$

with φ continuous function.

Fourier Approximation Of Operators

Important case: Φ **Fourier operator**, *i.e.*

$$\Phi\mu = \mathcal{F}\mu \stackrel{\text{def.}}{=} \left(\int_{\mathbb{T}^d} e^{-2i\pi\langle k, x\rangle} \mathrm{d}\mu(x) \right)_{k\in\Omega_c}, \quad \Omega_c = [\![-f_c, f_c]\!]^d,$$

for some cutoff frequency $f_c \in \mathbb{N}^*$ (Candès and Fernandez-Granda [2014])

Fourier Approximation Of Operators

Important case: Φ **Fourier operator**, *i.e.*

$$\Phi\mu = \mathcal{F}\mu \overset{\text{def.}}{=} \left(\int_{\mathbb{T}^d} e^{-2i\pi\langle k, x\rangle} \mathrm{d}\mu(x) \right)_{k \in \Omega_c}, \quad \Omega_c = [\![-f_c, f_c]\!]^d,$$

for some cutoff frequency $f_c \in \mathbb{N}^*$ (Candès and Fernandez-Granda [2014])

General case: $\Phi : \mathcal{M}(\mathbb{T}^d) \to \mathbb{C}^N$ **any integral operator**, spectral approximation:

$$\Phi \approx \mathcal{A} \circ \mathcal{F}$$

$$\text{where } \mathcal{A} \text{ solves} \quad \min_{\mathcal{A} \in \mathcal{M}_{N,|\Omega_c|}(\mathbb{C})} \|\Phi - \mathcal{A} \circ \mathcal{F}\|$$

Ideal Low-Pass Filtering

Convolution with Dirichlet kernel

$$\boldsymbol{\mathcal{A}} = \mathbf{Id}$$

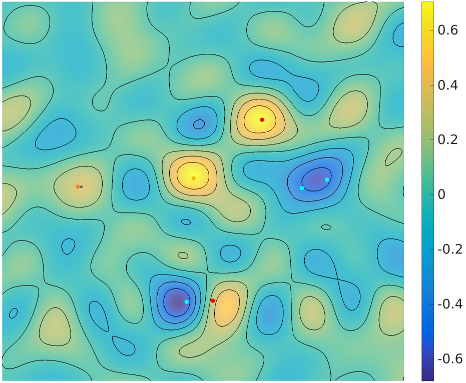


Figure: $\mathcal{F}^* y$

Gaussian Filtering

Convolution with (periodized) Gaussian kernel ψ

$$\mathcal{A} = \mathbf{Diag}\left(\hat{\psi}(-\boldsymbol{k})\right)_{\boldsymbol{k}\in\boldsymbol{\Omega_c}}$$

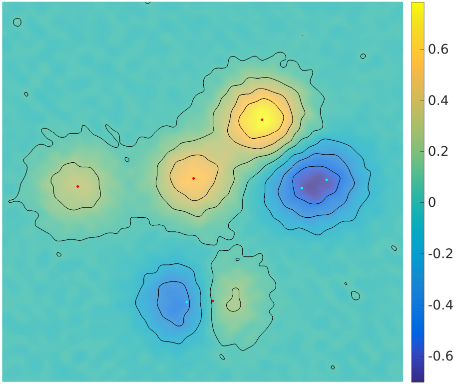


Figure: $\mathcal{F}^* y$

Gaussian Filtering + Subsampling

Not a convolution, sampling grid $\mathcal{G}$

$$\boldsymbol{\mathcal{A}} = \left(\hat{\psi}(-\boldsymbol{k}) e^{2i\pi \langle \boldsymbol{k}, \boldsymbol{t} \rangle} \right)_{(\boldsymbol{t},\boldsymbol{\omega}) \in \mathcal{G} \times \boldsymbol{\Omega_c}}$$

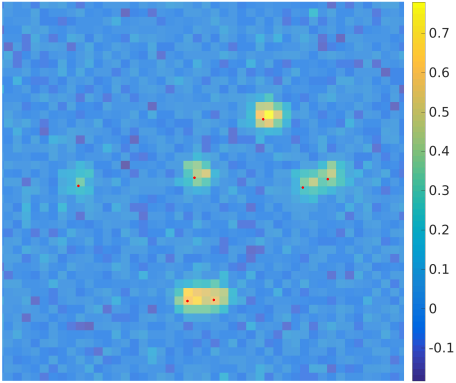


Figure: y, $\mathcal{G} = 64 \times 64$

Foveation

Not a convolution, sampling grid $\mathcal{G}$

$$\mathcal{A} = \ldots$$

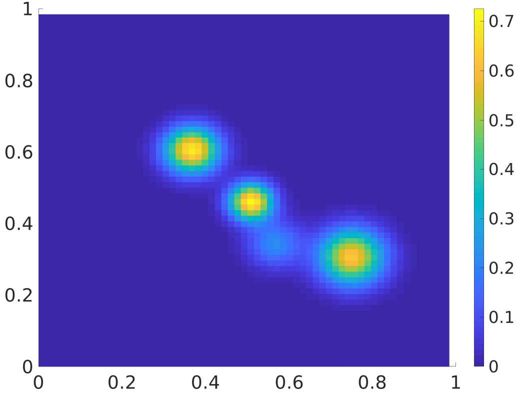


Figure: y, $\mathcal{G} = 64 \times 64$

Sparse Recovery

Measurements:

$$y = \mathcal{A}\mathcal{F}\mu_0 + w \in \mathbb{C}^N$$

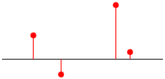
Sparse Recovery

Measurements:

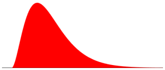
$$y = \mathcal{A}\mathcal{F}\mu_0 + w \in \mathbb{C}^N$$

Grid-free regularization: Total Variation of measures

$$|\mu|(\mathbb{T}^d) \overset{\text{def.}}{=} \sup \left\{ \int \eta \mathrm{d}\mu \ ; \ \eta \in \mathcal{C}(\mathbb{T}^d), \|\eta\|_\infty \leqslant 1 \right\}$$



$$|\mu|(\mathbb{T}^d) = \|a\|_{\ell^1}$$



$$|\mu|(\mathbb{T}^d) = \|f\|_{\mathrm{L}^1}$$

BLASSO (Azaïs et al. [2015])

$$\min_{\mu \in \mathcal{M}(\mathbb{T}^d)} \frac{1}{2}\|y - \mathcal{A}\mathcal{F}\mu\|^2 + \lambda|\mu|(X) \qquad (\mathcal{P}_\lambda)$$

Related Works

- Support discretization $\longrightarrow$ **LASSO** - *prox methods* (Donoho [1992])
  $\longrightarrow$ fast, inaccurate

- **Greedy methods** - *MP, FW* (Bredies and Pikkarainen [2013])
  $\longrightarrow$ continuous setting, slow convergence

- **SDP relaxation** (Candès and Fernandez-Granda [2014])
  $\longrightarrow$ simple, stable, not scalable

Off-the-Grid (applies to Greedy methods and SDP relaxation)

Contributions

- SDP approach combined with conditional gradient algorithm
- scalable FFT-based computations

Low-Rank Semidefinite Relaxations

Moment Matrices

Let $\ell \geqslant f_c$, $m = (2\ell + 1)^d$.

Definition (Moment matrices)

Given $\nu \in \mathcal{M}(\mathbb{T}^d)$, the moment matrix of order ℓ of ν is the matrix $R(\nu) \in \mathbb{C}^{m \times m}$ such that for every multi-indices $i, j \in [\![-\ell, \ell]\!]^d$,

$$R(\nu)_{i,j} = \int_{\mathbb{T}^d} e^{-2i\pi\langle i-j, x\rangle} \mathrm{d}\nu(x)$$

Definition (Generalized Tœplitz matrices)

$R \in \mathbb{C}^{m \times m}$ is a generalized Tœplitz matrix, denoted $R \in \mathcal{T}_m$, if for every multi-indices $i, j, k \in [\![-\ell, \ell]\!]^d$ such that $\|i + k\|_\infty \leqslant \ell$ and $\|j + k\|_\infty \leqslant \ell$,

$$R_{i+k,j+k} = R_{i,j}$$

Semidefinite Hierarchies

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

Lasserre [2001]

- BLASSO only involves a few **trigonometric moments** of μ and $|\mu|$
- It may be cast over the cone of moment sequences ...
- ... approximable by a hierarchy of **semidefinite** cones ...
- ... involving **PSD + GT matrices** of increasing size

Semidefinite Hierarchies

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

Lasserre [2001]

- BLASSO only involves a few **trigonometric moments** of μ and $|\mu|$
- It may be cast over the cone of moment sequences ...
- ... approximable by a hierarchy of **semidefinite** cones ...
- ... involving **PSD + GT matrices** of increasing size

Moment relaxation at order ℓ

$$\min_{R,z,\tau} \quad \frac{1}{2}\|y - \mathcal{A}z\|^2 \; + \; \frac{\lambda}{2}\left(\frac{1}{m}\mathrm{Tr}(R) + \tau\right) \qquad (\mathcal{P}_\lambda^{(\ell)})$$

$$\text{s.t.} \begin{cases} (a) \quad \mathcal{R} = \begin{bmatrix} R & \tilde{z} \\ \tilde{z}^* & \tau \end{bmatrix} \succeq 0, \;\; \tilde{z}_k = z_k, \; \forall k \in \Omega_c \\ (b) \quad R \in \mathcal{T}_m \end{cases}$$

Semidefinite Hierarchies

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

Lasserre [2001]

- BLASSO only involves a few **trigonometric moments** of μ and $|\mu|$
- It may be cast over the cone of moment sequences ...
- ... approximable by a hierarchy of **semidefinite** cones ...
- ... involving **PSD + GT matrices** of increasing size

Moment relaxation at order ℓ

$$\min_{R,z,\tau} \quad \frac{1}{2}\|y-\mathcal{A}z\|^2 + \frac{\lambda}{2}\left(\frac{1}{m}\mathrm{Tr}(R)+\tau\right) \qquad (\mathcal{P}_\lambda^{(\ell)})$$

$$\text{s.t.} \begin{cases} (a) \quad \mathcal{R}=\begin{bmatrix} R & \tilde{z} \\ \tilde{z}^* & \tau \end{bmatrix} \succeq 0, \quad \tilde{z}_k = z_k, \forall k\in\Omega_c \\ (b) \quad R\in\mathcal{T}_m \end{cases}$$

Proposition

For any $\ell \geqslant f_c$, $\min(\mathcal{P}_\lambda^{(\ell)}) \leqslant \min(\mathcal{P}_\lambda^{(\ell+1)}) \leqslant \min(\mathcal{P}_\lambda)$. Moreover, $\lim_{\ell\to\infty}\min(\mathcal{P}_\lambda^{(\ell)}) = \min(\mathcal{P}_\lambda)$

Collapsing Of The Hierarchy

Proposition

Let $\ell \geqslant f_c$. *Then* $\min(\mathcal{P}_\lambda^{(\ell)}) = \min(\mathcal{P}_\lambda)$ *iff there exists* (R, z, τ) *solutions to* $(\mathcal{P}_\lambda^{(\ell)})$ *and* μ *solution to* $(\mathcal{P}_\lambda)$ *such that*

$$\tau = |\mu|(\mathbb{T}^d) \quad \textit{and} \quad R = R(|\mu|)$$

Collapsing Of The Hierarchy

Proposition

Let $\ell \geqslant f_c$. Then $\min(\mathcal{P}_\lambda^{(\ell)}) = \min(\mathcal{P}_\lambda)$ iff there exists (R, z, τ) solutions to $(\mathcal{P}_\lambda^{(\ell)})$ and μ solution to $(\mathcal{P}_\lambda)$ such that

$$\tau = |\mu|(\mathbb{T}^d) \quad \textit{and} \quad R = R(|\mu|)$$

When do we have $\min(\mathcal{P}_\lambda^{(\ell)}) = \min(\mathcal{P}_\lambda)$?

- When $d = 1$, it holds for **any** $\ell \geqslant f_c$. (Tang et al. [2013])
- When $d = 2$, there **exists** $\ell \geqslant f_c$ such that the relaxation is tight.
- When $d > 2$, we do not know in general.

Collapsing detected via **flatness criterion** on R (Curto and Fialkow [1996])

Collapsing Of The Hierarchy

Proposition

Let $\ell \geqslant f_c$. Then $\min(\mathcal{P}_\lambda^{(\ell)}) = \min(\mathcal{P}_\lambda)$ *iff there exists* (R, z, τ) *solutions to* $(\mathcal{P}_\lambda^{(\ell)})$ *and* μ *solution to* $(\mathcal{P}_\lambda)$ *such that*

$$\tau = |\mu|(\mathbb{T}^d) \quad \textit{and} \quad R = R(|\mu|)$$

When do we have $\min(\mathcal{P}_\lambda^{(\ell)}) = \min(\mathcal{P}_\lambda)$?

- When $d = 1$, it holds for **any** $\ell \geqslant f_c$. (Tang et al. [2013])
- When $d = 2$, there **exists** $\ell \geqslant f_c$ such that the relaxation is tight.
- When $d > 2$, we do not know in general.

Collapsing detected via **flatness criterion** on R (Curto and Fialkow [1996])

How to retrieve μ from $R(|\mu|)$?
Algebraic method (Laurent [2010], Lasserre [2010], Josz et al. [2017])

Low-Rank Structure

Proposition

In the case of collapsing, $(\mathcal{P}_\lambda^{(\ell)})$ *always admits a solution* $\mathcal{R}_\lambda$ *such that* $\operatorname{rank} \mathcal{R}_\lambda \leqslant r$*,* r *being the number of spikes in a solution of* $(\mathcal{P}_\lambda)$*.*

Proof.

Results from the fact that if $\nu = \sum_{i=1}^{r} a_i \delta_{x_i}$, then $\operatorname{rank} R(\nu) \leqslant r$. □

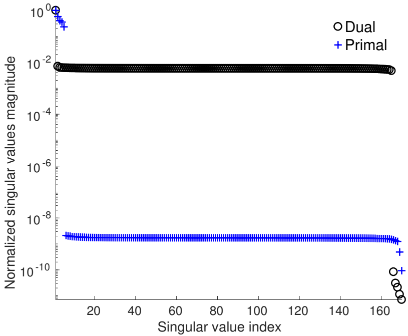


Figure: ($r = 5$ spikes, $f_c = 5$, $d = 2$). Singular values of primal and dual matrices

Algorithm: FFT-Based Frank-Wolfe

Penalized Problem

$$(\mathcal{P}_\lambda^{(\ell)}) \quad \min_{R,z,\tau} \frac{1}{2}\|y - \mathcal{A}z\|^2 + \frac{\lambda}{2}\left(\frac{1}{m}\mathrm{Tr}(R) + \tau\right)$$

$$\text{s.t.} \begin{cases} (a) \begin{bmatrix} R & \tilde{z} \\ \tilde{z}^* & \tau \end{bmatrix} \succeq 0, \quad \tilde{z}_k = z_k, \forall k \in \Omega_c \\ (b) \; R \in \mathcal{T}_m \end{cases}$$

Frank-Wolfe (aka Conditional Gradient):

- handles well low-rank iterates
- cannot handle the geometry induced by (a) + (b)

Penalized Problem

$$(\mathcal{P}_\lambda^{(\ell)}) \quad \min_{R,z,\tau} \frac{1}{2}\|y - \mathcal{A}z\|^2 + \frac{\lambda}{2}\left(\frac{1}{m}\mathrm{Tr}(R) + \tau\right)$$

$$\text{s.t.} \begin{cases} (a) \begin{bmatrix} R & \tilde{z} \\ \tilde{z}^* & \tau \end{bmatrix} \succeq 0, \quad \tilde{z}_k = z_k, \forall k \in \Omega_c \\ (b) \; R \in \mathcal{T}_m \end{cases}$$

<u>Frank-Wolfe</u> (aka Conditional Gradient):

- handles well low-rank iterates
- cannot handle the geometry induced by (a) + (b)

$\Longrightarrow$ **Penalize Toeplitz constraint (b)**

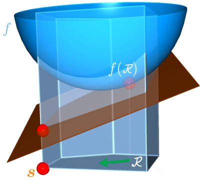
$$\min_{R,z,\tau} \quad \frac{1}{2}\|y - \mathcal{A}z\|^2 + \frac{\lambda}{2}\left(\frac{1}{m}\mathrm{Tr}(R) + \tau\right) + \frac{1}{2\rho}\|R - P_{\mathcal{T}_m}(R)\|^2$$

$$\text{s.t.} \quad \mathcal{R} = \begin{bmatrix} R & \tilde{z} \\ \tilde{z}^* & \tau \end{bmatrix} \succeq 0, \quad \tilde{z}_k = z_k, \forall k \in \Omega_c. \qquad (\mathcal{P}_{\lambda,\rho}^{(\ell)})$$

Alternating Descent Conditional Gradient Method

Frank-Wolfe steps:

1. $\mathcal{S}_\star \in \underset{\mathcal{S}\in\mathcal{D}}{\text{argmin}} \langle \nabla f(\mathcal{R}_t), \mathcal{S} \rangle$
2. $\mathcal{R}_{t+1} = \mathcal{R}_t + c(\mathcal{S}_\star - \mathcal{R}_t)$, with $c \in [0, 1]$



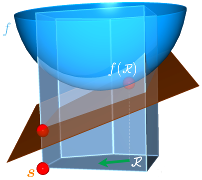
Jaggi [2013]

- ⊕ Sparse iterates
- ⊕ Simple LM
- ⊖ Slow convergence: $f(\mathcal{R}_t) - f(\mathcal{R}^\star) \leqslant O(\frac{1}{t})$

Alternating Descent Conditional Gradient Method

Frank-Wolfe steps:

1. $\mathcal{S}_\star \in \underset{\mathcal{S}\in\mathcal{D}}{\text{argmin}} \langle \nabla f(\mathcal{R}_t), \mathcal{S}\rangle$
2. $\mathcal{R}_{t+1} = \mathcal{R}_t + c(\mathcal{S}_\star - \mathcal{R}_t)$, with $c \in [0,1]$



Jaggi [2013]

- ⊕ Sparse iterates
- ⊕ Simple LM
- ⊖ Slow convergence: $f(\mathcal{R}_t) - f(\mathcal{R}^\star) \leqslant O(\frac{1}{t})$

- $\mathcal{D} = \{X \succeq 0 \,;\, \operatorname{tr} X \leqslant 1\}$
- **Step 1. := $O(f_c^d \log f_c)$ FFT-based Power Iterations to compute a leading eigenvector of ∇f**
- Low-cost storage: $\mathcal{R}_t = \mathcal{U}_t\mathcal{U}_t^*$
- **Non-convex BFGS step** (Boyd et al. [2015]) on $F : \mathcal{U} \mapsto f(\mathcal{U}\mathcal{U}^*)$.

Algo: Building Moment Matrix

Set: $\mathcal{U}_0 = [0...0]^\top$, D_0: $\operatorname{tr}\mathcal{R}_\star \leqslant D_0$

For $t = 1, \ldots$ **do**

1. $v_t = D_0 \underset{\|v\|\leqslant 1}{\text{argmin}}\, v^\top \nabla f\,[\mathcal{U}_t\mathcal{U}_t^*]\, v$
2. $\widehat{\mathcal{U}}_{t+1} = [\alpha_t\mathcal{U}_t, \beta_t v_t]$, with ls on (α_t, β_t) (closed form)
3. $\mathcal{U}_{t+1} = \textbf{bfgs}(F(\mathcal{U}), \text{start at } \widehat{\mathcal{U}}_{t+1})$

Numerics

Numerics

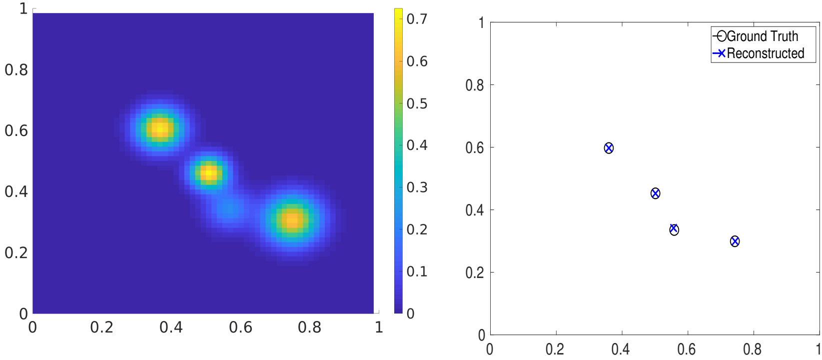


Figure: $f_c = 25$, $\lambda = 5.10^{-4}\|\Phi^* y\|_\infty$, $\rho = 10^3$
Support localization relative error $= 4.7 \times 10^{-3}$

Numerics

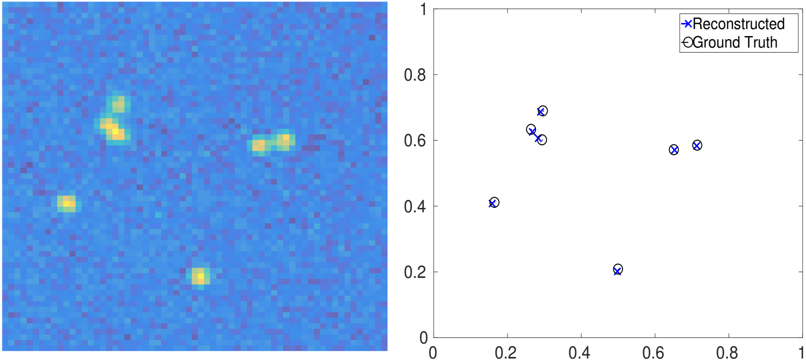


Figure: $f_c = 30$, $\lambda = 5.10^{-3} \|\Phi^* y\|_\infty$, $\rho = 5.10^5 \|\Phi^* y\|_\infty^{-1}$.

Support localization error $\frac{\|x_{\text{recov}} - x_0\|}{\|x_0\|} = 1.57 \times 10^{-2}$

Numerics

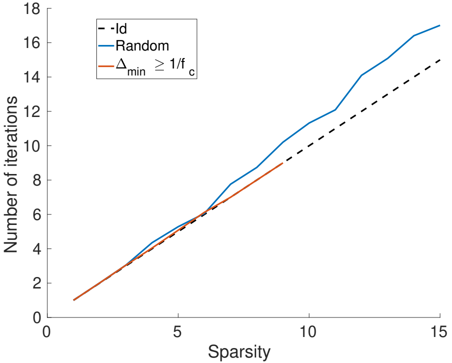
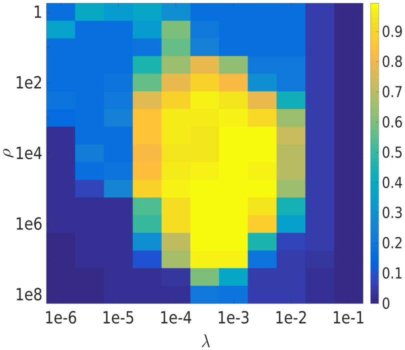


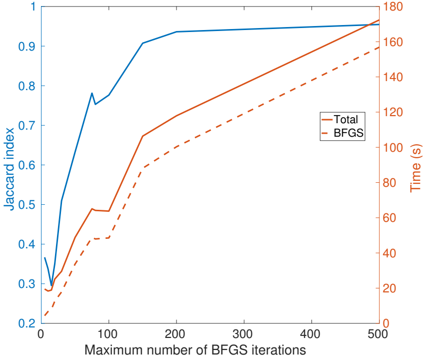
Figure: Mean number of iterations before convergence (over 200 random trials), with respect to sparsity of the solution measure

Numerics

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



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



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

Conclusion

- SDP formulation for problem the problem of spikes superresolution...
- ... which admits low-rank solutions
- Scalable method in 2D, based on a conditional gradient approach
- **Future works:** Lasserre's hierarchy encompasses large class of problems (polynomial optimization, optimal transport, etc...)
  $\longrightarrow$ possible extensions for our algorithm

Thank you for your attention!

Fast-Fourier-Tranforms-Based Computations

- Leading eigenvector is computed using **Power Iteration**.
- Requires only computing $\nabla f \cdot v$, with

$$\nabla f(\mathcal{U}\mathcal{U}^*) = \begin{bmatrix} \frac{1}{n} I_n & p \\ p^* & 1 \end{bmatrix} + \frac{1}{\rho}\mathcal{U}\mathcal{U}^* - \frac{1}{\rho} P_{V_\Theta}(\mathcal{U}\mathcal{U}^*)$$

- Main costly operation: $P_{V_\Theta}(\mathcal{U}\mathcal{U}^*) \cdot v$

Fast-Fourier-Tranforms-Based Computations

- Leading eigenvector is computed using **Power Iteration**.
- Requires only computing $\nabla f \cdot v$, with

$$\nabla f(\mathcal{U}\mathcal{U}^*) = \begin{bmatrix} \frac{1}{n} I_n & p \\ p^* & 1 \end{bmatrix} + \frac{1}{\rho}\mathcal{U}\mathcal{U}^* - \frac{1}{\rho} P_{V_\Theta}(\mathcal{U}\mathcal{U}^*)$$

- Main costly operation: $P_{V_\Theta}(\mathcal{U}\mathcal{U}^*) \cdot v$

Key Ingredient: $O(f_c^d \log(f_c))$ FFT-Based Computations

Toeplitz-Vector Multiplication

Let $x \in \mathbb{C}^{(n+1)^d}$, $t \in \mathbb{C}^{(2n+1)^d}$, and $T = \mathrm{T\oe p}(t)$. Then

$$Tx = \mathrm{Pad}^{-1} \circ \mathcal{F}^{-1}\Big(\langle \mathcal{F} \circ \mathrm{Pad}(x), \mathcal{F}(t)\rangle\Big)$$

Toeplitz Projection

Let $U = [U_1, \ldots, U_r] \in \mathbb{C}^{(n+1)^d \times r}$. Then $P_{\mathcal{T}}[UU^*] = \mathrm{T\oe p}(t)$, with

$$t_i \propto \left[\sum_k \mathcal{F}^{-1}(|\mathcal{F} \circ \mathrm{Pad}(U_k)|^2)\right]_i$$

Support Recovery via Root-Finding

Dual polynomial $\eta_\lambda = \sum p_k e^{2i\pi\langle k, x\rangle}$

Root-finding:

- $P(X) = \sum_k p_k X^k$, $X \in \mathbb{C}^d$
- Solve $|P(X)|^2 - 1 = 0$
- Select roots s.t. $|X| = 1$

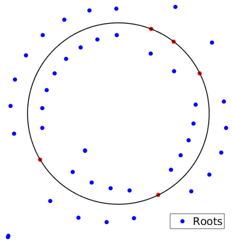


Figure: Roots of $1 - |P|^2$, with $P = \sum p_k X^k$

Sensitivity Analysis

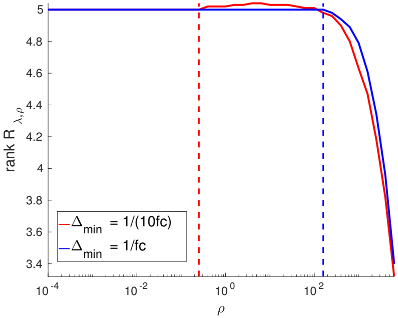


Figure: Rank of $\mathcal{R}_{\lambda,\rho}$ w.r.t. ρ

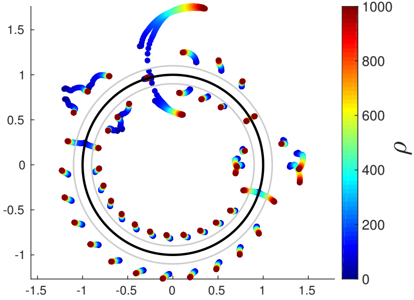
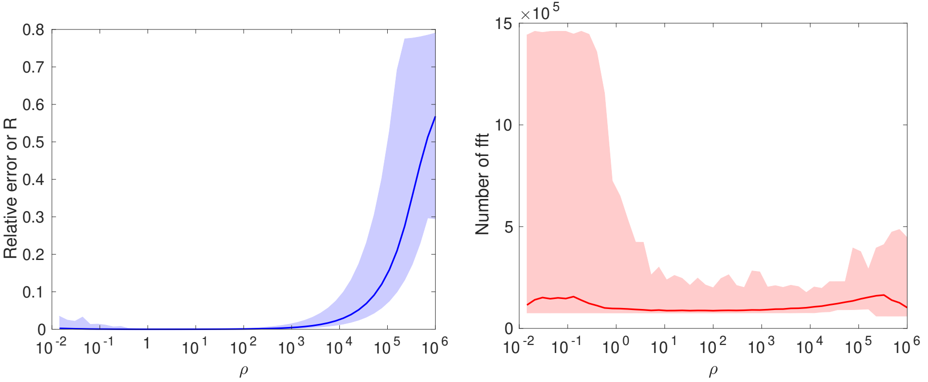


Figure: Roots trajectory w.r.t ρ.


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Relative error or R
10^{-2} 10^{-1} 1 10^{1} 10^{2} 10^{3} 10^{4} 10^{5} 10^{6}
ρ
$\times 10^{5}$
15
10
5
0
Number of fft
10^{-2} 10^{-1} 10^{0} 10^{1} 10^{2} 10^{3} 10^{4} 10^{5} 10^{6}
ρ

J. M. Azaïs, Y. de Castro, and F. Gamboa. Spike detection from inaccurate sampling. Applied and Computational Harmonic Analysis, 38(2):177–195, 2015.

N. Boyd, G. Schiebinger, and B. Recht. The alternating descent conditional gradient method for sparse inverse problems. In CAMSAP, pages 57–60, 2015.

K. Bredies and H. K. Pikkarainen. Inverse problems in spaces of measures. ESAIM: Control, Optimization and Calculus of Variations, 19(1):190–218, 2013.

E.J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. Comm. Pure Appl. Math, 67(6):906–956, 2014.

R.E. Curto and L.A. Fialkow. Solution of the truncated complex moment problem for flat data. Memoirs of the AMS, (568), 1996.

D. L. Donoho. Superresolution via sparsity constraints. SIAM Journal on Mathematical Analysis, 23(5):1309–1331, 1992.

M. Jaggi. Revisiting frank-wolfe: Projection-free sparse convex optimization. In ICML, volume 28, 2013.

C. Josz, J.B. Lasserre, and B. Mourrain. Sparse polynomial interpolation: Compressed sensing, super resolution, or prony? arXiv:1708.06187, 2017.

J. B. Lasserre. Global optimization with polynomials and the problem of moments. SIAM Journal on Optimization, 11(3):796–817, 2001.

J. B. Lasserre. Moments, positive polynomials and their applications. Imperial College Press Optimization Series. Imperial College Press, London, 2010. ISBN 978-1-84816-445-1. URL http://opac.inria.fr/record=b1130505.

M. Laurent. Sums of squares, moment matrices and optimization over polynomials. In Emerging Applications of Algebraic Geometry, volume 149. Springer new York, 2010.

G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. IEEE Trans. Inf. Theory, 59(11):7465–7490, 2013.