Off-the-Grid Wasserstein Group Lasso

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

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

[2]Inria Paris

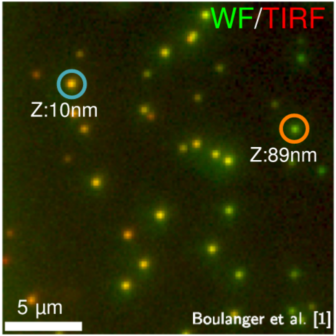
[3]CEREMADE, Université Paris-Dauphine, PSL, CNRS

GRETSI 2019, 26-29 août, Lille

Motivation: Sparse Super-Resolution

Recover pointwise sources from low-resolution and noisy observations.

Astrophysics (2D)



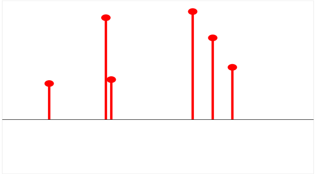
Molecule fluorescence (3D)

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

Model

Super-Resolution of Measures

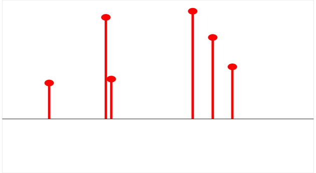
<u>**Signal to recover**</u>: discrete positive Radon measure on d-dimensional torus $\mathbb{T}^d$



$$\mu_0 = \sum_{k=1}^{r} a_k \delta_{x_k} \in \mathcal{M}_+(\mathbb{T}^d)$$

Super-Resolution of Measures

Signal to recover: discrete positive Radon measure on d-dimensional torus $\mathbb{T}^d$



$$\mu_0 = \sum_{k=1}^{r} a_k \delta_{x_k} \in \mathcal{M}_+(\mathbb{T}^d)$$

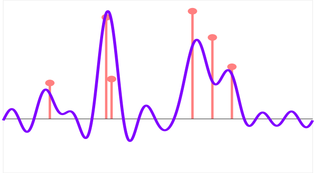
Linear Fourier measurements:

$$y = \mathcal{F}(\mu_0) + w \in \mathbb{C}^n$$

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

with $\Omega_c \overset{\text{def.}}{=} [\![-f_c, f_c]\!]^d$.

$\iff$ convolution with low-pass filter



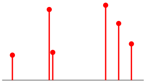
$\mathcal{F}^* y$

Off-The-Grid Recovery

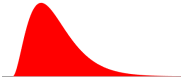
Inverse problem: $y = \mathcal{F}(\mu_0) + w \in \mathbb{C}^n$

Grid-free regularization: total variation (TV) of measures

$$|\mu|(\mathbb{T}^d) = \sup \left\{ \int_{\mathbb{T}^d} \eta \mathrm{d}\mu \ ; \ \eta \in \mathcal{C}(\mathbb{T}^d) \quad \text{and} \quad \|\eta\|_\infty \leqslant 1 \right\}$$



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



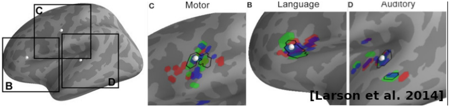
$|\mu|(\mathbb{T}^d) = \|f\|_{L^1}$

BLASSO (Azaïs et al. [2015])

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

Multi-Task Off-the-Grid Recovery

Inverse problem: $u = \mathcal{F}(\mu_0) + w$ and $v = \mathcal{F}(\nu_0) + \varepsilon$, $\mu_0 \simeq \nu_0$



regularization: TV + Wasserstein (Janati et al. [2018])

$$\mathcal{W}_c(\mu, \nu) = \min_{\gamma \in \mathcal{M}_+(\mathbb{T}^d \times \mathbb{T}^d)} \left\{ \int_{\mathbb{T}^d \times \mathbb{T}^d} c \mathrm{d}\gamma \; ; \; \pi_1 \gamma = \mu \quad \text{and} \quad \pi_2 \gamma = \nu \right\}$$

Wasserstein-BLASSO

$$\min_{\mu, \nu \in \mathcal{M}_+(\mathbb{T}^d)} \frac{1}{2} \|u - \mathcal{F}(\mu)\|^2 + \lambda \mu(\mathbb{T}^d) + \frac{1}{2} \|v - \mathcal{F}(\nu)\|^2 + \lambda \nu(\mathbb{T}^d) + \tau \mathcal{W}_c(\mu, \nu) \qquad (\mathcal{P}_{\lambda, \tau})$$

Off-the-grid extension of Janati et al. [2018]

Semidefinite Hierarchies

Lasserre [2001], Parrilo [2003], Dumitrescu [2017]

Moment Matrices

Let $\Omega_\ell = [\![0, \ell]\!]^d$, $\ell \geqslant f_c$, and $m = (\ell + 1)^d$.

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

$$R(\nu)_{k,l} = \int_{\mathbb{T}^d} e^{-2i\pi\langle k-l, x\rangle} \mathrm{d}\nu(x) \quad \forall k, l \in \Omega_\ell$$

Moment Matrices

Let $\Omega_\ell = [\![0, \ell]\!]^d$, $\ell \geqslant f_c$, and $m = (\ell + 1)^d$.

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

$$R(\nu)_{k,l} = \int_{\mathbb{T}^d} e^{-2i\pi\langle k-l, x\rangle} \mathrm{d}\nu(x) \quad \forall k, l \in \Omega_\ell$$

Definition (Generalized Toeplitz matrices $\mathcal{T}_m$) $R \in \mathcal{T}_m$ if for every multi-indices $j, k, l \in \Omega_\ell$ such that $\|k + j\|_\infty \leqslant \ell$ and $\|l + j\|_\infty \leqslant \ell$,

$$\boxed{R_{k+j,l+j} = R_{k,l}} \overset{\text{def.}}{=} z_{k-l}$$

In this case, we write $R = \mathrm{Toep}(z)$

Moment Matrices

Let $\Omega_\ell = [\![0, \ell]\!]^d$, $\ell \geqslant f_c$, and $m = (\ell+1)^d$.

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

$$R(\nu)_{k,l} = \int_{\mathbb{T}^d} e^{-2i\pi\langle k-l, x\rangle} \mathrm{d}\nu(x) \quad \forall k, l \in \Omega_\ell$$

Definition (Generalized Toeplitz matrices $\mathcal{T}_m$) $R \in \mathcal{T}_m$ if for every multi-indices $j, k, l \in \Omega_\ell$ such that $\|k+j\|_\infty \leqslant \ell$ and $\|l+j\|_\infty \leqslant \ell$,

$$\boxed{R_{k+j,l+j} = R_{k,l}} \overset{\text{def.}}{=} z_{k-l}$$

In this case, we write $R = \mathrm{Toep}(z)$

- $R(\nu) \in \mathcal{T}_m$
- If $\nu \geqslant 0$, $R(\nu) \succeq 0$
- If $\nu = \sum a_i \delta_{x_i}$, $R(\nu) = \sum a_i e(x_i) e(x_i)^*$, with $e(x) = [e^{-2i\pi\langle k, x\rangle}]_{k \in \Omega_\ell}$

Example: OT

$$\mathcal{W}_c(\mu,\nu) = \min_{\gamma \in \mathcal{M}_+(\mathbb{T}^d \times \mathbb{T}^d)} \int c \mathrm{d}\gamma \quad \text{s.t.} \quad \begin{cases} \pi_1\gamma = \mu \\ \pi_2\gamma = \nu \end{cases}$$

- assume cost is a trigonometric polynomial: $c = \sum_k \hat{c}_k e^{-2i\pi\langle k,x\rangle}$
- $\mathcal{W}_c$ only involves trigonometric moments of γ ($\gamma \geqslant 0$)
- Replace measures by (infinite) moment sequences ...
- ... truncate these sequences ...
- ... they will satisfy (necessary) **PSD** constraints

Example: OT

$$\mathcal{W}_c(\mu,\nu) = \min_{\gamma \in \mathcal{M}_+(\mathbb{T}^d \times \mathbb{T}^d)} \int c \,\mathrm{d}\gamma \quad \text{s.t.} \quad \begin{cases} \pi_1 \gamma = \mu \\ \pi_2 \gamma = \nu \end{cases}$$

- assume cost is a trigonometric polynomial: $c = \sum_k \hat{c}_k e^{-2i\pi\langle k, x\rangle}$
- $\mathcal{W}_c$ only involves trigonometric moments of γ ($\gamma \geqslant 0$)
- Replace measures by (infinite) moment sequences ...
- ... truncate these sequences ...
- ... they will satisfy (necessary) **PSD** constraints

«Change of variable:

$$z = \mathcal{F}_2(\gamma), \quad \textit{i.e.} \;\; z_{(s,t)} = \int_{\mathbb{T}^d \times \mathbb{T}^d} e^{-2i\pi\langle s, x\rangle} e^{-2i\pi\langle t, y\rangle} \mathrm{d}\gamma(x, y)$$

$$z_1 = \mathcal{F}(\pi_1\gamma) = z_{(\cdot,0)}, \quad \text{and} \quad z_2 = \mathcal{F}(\pi_2\gamma) = z_{(0,\cdot)} \quad »$$

Moment relaxation at order ℓ $\quad (m = (\ell+1)^d)$

$$\min_{z \in \mathbb{C}^{(2m-1)\times(2m-1)}} \langle \hat{c}, z\rangle \quad \text{s.t.} \begin{cases} \mathrm{Toep}(z) \succeq 0 \\ z_1 = u \\ z_2 = v \end{cases} \qquad (OT^{(\ell)})$$

SDP hierarchy for Wasserstein-BLASSO

$$\min_{\mu,\nu\in\mathcal{M}_+(\mathbb{T}^d)} \frac{1}{2}\|u-\mathcal{F}(\mu)\|^2+\lambda\mu(\mathbb{T}^d)+\frac{1}{2}\|v-\mathcal{F}(\nu)\|^2+\lambda\nu(\mathbb{T}^d)+\tau\mathcal{W}_c(\mu,\nu)$$

$\Updownarrow$ reformulation over product measures

$$\min_{\gamma\in\mathcal{M}_+(\mathbb{T}^d\times\mathbb{T}^d)} \frac{1}{2}\|u-\mathcal{F}(\pi_1\gamma)\|^2+\frac{1}{2}\|v-\mathcal{F}(\pi_2\gamma)\|^2+2\lambda\gamma(\mathbb{T}^d\times\mathbb{T}^d)+\tau\langle c,\,\gamma\rangle$$

$\Downarrow$ semidefinite relaxation

Moment relaxation at order ℓ

$$\begin{aligned}\min_{z\in\mathbb{C}^{(2m-1)\times(2m-1)}} \quad & \frac{1}{2}\|u-z_1\|^2+\frac{1}{2}\|v-z_2\|^2+2\lambda z_0+\tau\langle\hat{c},\,z\rangle \\ \text{s.t.} \quad & \mathrm{Toep}(z)\succeq 0\end{aligned} \qquad (\mathcal{P}_{\lambda,\tau}^{(\ell)})$$

Convergence of the hierarchy

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

Prop. (Collapsing) Let $\ell \geqslant f_c$. Then $\min(\mathcal{P}_{\lambda,\tau}^{(\ell)}) = \min(\mathcal{P}_{\lambda,\tau})$ iff there exist z solution to $(\mathcal{P}_{\lambda,\tau}^{(\ell)})$ and γ solution to $(\mathcal{P}_{\lambda,\tau})$ st $z = \mathcal{F}_2(\gamma)$ (z be the moments of γ).

We know how to detect collapsing via flatness criterion ($:=$ recurrence relations between columns of $\mathrm{Toep}(z)$) Curto and Fialkow [1996]

Low-Rank Solutions to $(\mathcal{P}^{(\ell)}_{\lambda,\tau})$

Proposition

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

Proof.

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

$\Rightarrow$ **efficient FFT-based Frank-Wolfe solver** (C.,Duval,Peyré [2019])

Support Recovery

Sparse Recovery with Prony

Let $R = R(\gamma) = \mathrm{Toep}(z)$

$$p \in \operatorname{Ker} R \Rightarrow p^* R p = 0 \Rightarrow \int_{\mathbb{T}^d} \Big| \sum_k p_k e^{2i\pi\langle k, x\rangle} \Big|^2 \mathrm{d}\gamma(x) = 0$$
$$\Rightarrow \operatorname{Supp} \gamma \subset \left\{ x \in \mathbb{T}^d \; ; \; p(x) = 0 \right\}$$

Let $\langle \operatorname{Ker} R \rangle \stackrel{\text{def.}}{=}$ ideal generated by $\operatorname{Ker} R$

Theorem (see e.g. Laurent [2010]) If the flatness criterion holds, then
$$\operatorname{Supp} \gamma = \left\{ x \in \mathbb{T}^d \; ; \; p(x) = 0 \quad \forall p \in \langle \operatorname{Ker} R \rangle \right\}$$

Solving system of polynomial equations $\Rightarrow$ (multivariate) Prony's method

Based on joint diagonalization (Harmouch et al. [2017], Josz et al. [2017])
- in 1-D, $\langle \operatorname{Ker} R \rangle = \langle p \rangle$, root-finding $\Leftrightarrow$ eigenvalues of companion matrix
- in d-D, joint diagonalization of (commuting) "multiplication matrices"

Recovering non-atomic measures

Approximate joint diagonalization

use optimization scheme to find best co-diagonalization basis for multiplication matrices

return a discrete measure

Christoffel polynomial

Pauwels and Lasserre [2019]

use regularized inverse of moment matrix

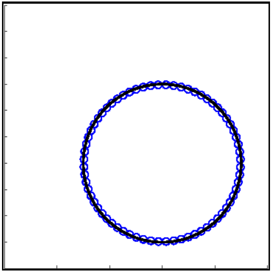
support $\subset$ level sets

Recovering non-atomic measures

Approximate joint diagonalization

use optimization scheme to find best co-diagonalization basis for multiplication matrices

return a discrete measure

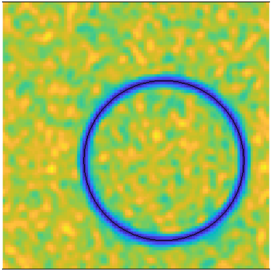


Christoffel polynomial

Pauwels and Lasserre [2019]

use regularized inverse of moment matrix

support $\subset$ level sets

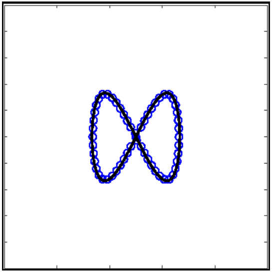


Recovering non-atomic measures

Approximate joint diagonalization

use optimization scheme to find best co-diagonalization basis for multiplication matrices

return a discrete measure

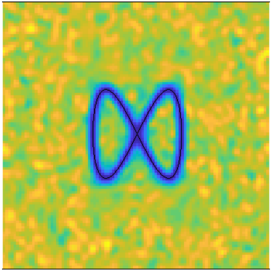


Christoffel polynomial

Pauwels and Lasserre [2019]

use regularized inverse of moment matrix

support $\subset$ level sets

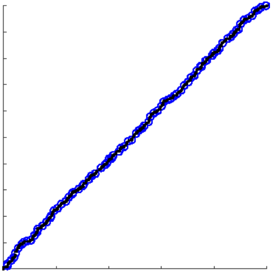


Recovering non-atomic measures

Approximate joint diagonalization

use optimization scheme to find best co-diagonalization basis for multiplication matrices

return a discrete measure

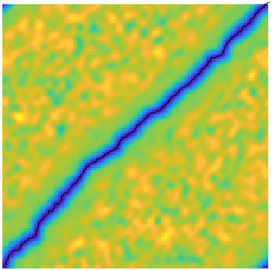


Christoffel polynomial

Pauwels and Lasserre [2019]

use regularized inverse of moment matrix

support $\subset$ level sets

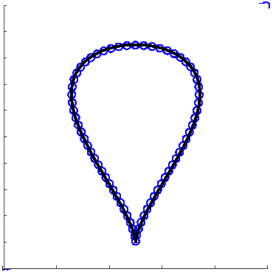


Recovering non-atomic measures

Approximate joint diagonalization

use optimization scheme to find best co-diagonalization basis for multiplication matrices

return a discrete measure

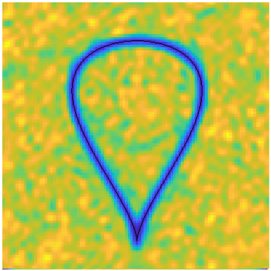


Christoffel polynomial

Pauwels and Lasserre [2019]

use regularized inverse of moment matrix

support $\subset$ level sets

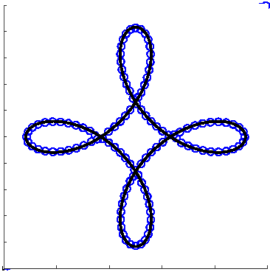


Recovering non-atomic measures

Approximate joint diagonalization

use optimization scheme to find best co-diagonalization basis for multiplication matrices

return a discrete measure

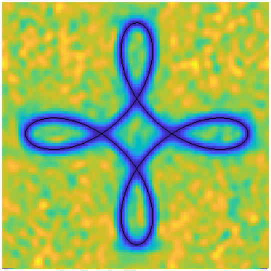


Christoffel polynomial

Pauwels and Lasserre [2019]

use regularized inverse of moment matrix

support $\subset$ level sets

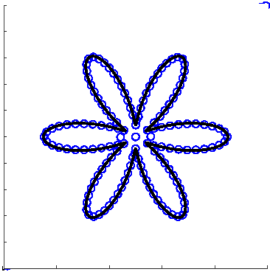


Recovering non-atomic measures

Approximate joint diagonalization

use optimization scheme to find best co-diagonalization basis for multiplication matrices

return a discrete measure

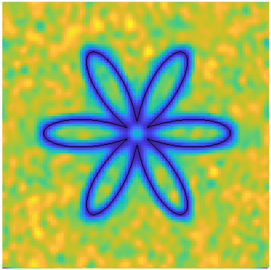


Christoffel polynomial

Pauwels and Lasserre [2019]

use regularized inverse of moment matrix

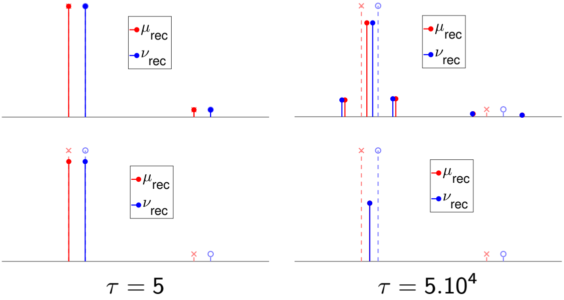
support $\subset$ level sets



Simulations

Simulations

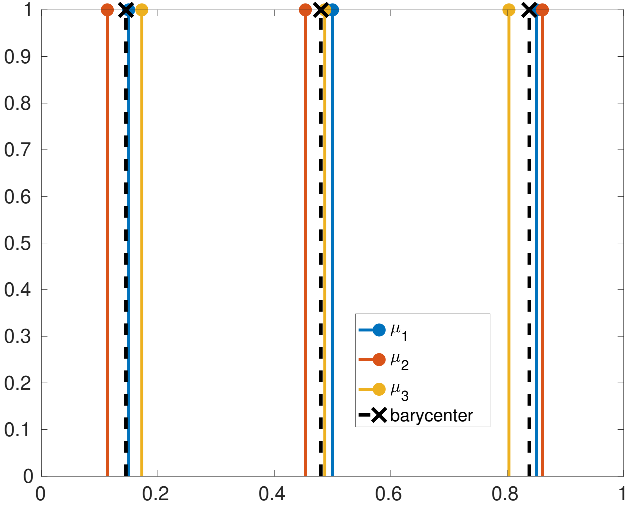
Wasserstein cost: $\sin^2(x-y)$



noiseless case, $\lambda = 10^{-2}$ (*top*) and $\lambda = 1$ (*bottom*)

Simulations

Multi observations penalization: $\sum_k \mathcal{W}_c(\mu_k, \mu_b)$



Conclusion

- off-the-grid solver for the multi-task super-resolution problem
- using the Wasserstein penalization introduced by Janati et al. [2018]
- and Lasserre's hierarchy

- **Future lines of work:**
  - extension to unbalanced transport
  - Lasserre's hierarchy for curve recovery

Thank You!

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

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

B. A. Dumitrescu. *Positive trigonometric Polynomials and Signal Processing Applications*. Signals and Communication Technology. Springer International Publishing, 2017.

J. Harmouch, H. Khalil, and B. Mourrain. Structured low rank decomposition of multivariate Hankel matrices. *Linear Algebra and its Applications*, 542:162 – 185, 2017.

H. Janati, M. Cuturi, and A. Gramfort. Wasserstein regularization for sparse multi-task regression. *ArXiv e-prints*, 2018.

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.

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

P. A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming*, 96(2):293–320, 2003.

E. Pauwels and J.B. Lasserre. The empirical christoffel function with applications in data analysis. *Advances in Computational Mathematics*, 45 (3):1439–1468, 2019.