Déconvolution Parcimonieuse Sans Grille: une Méthode de Faible Rang

Paul Catala [1]
Joint work with Vincent Duval [2,3] and Gabriel Peyré [1,4]

[1]DMA, ENS Ulm

[2]Mokaplan, Inria Paris

[3]CEREMADE, Univerité Paris-Dauphine

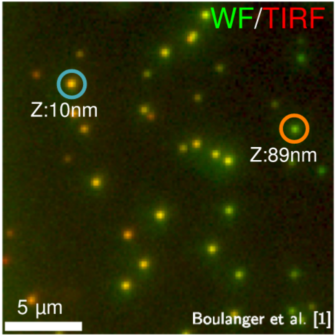
[4]CNRS, UMR 8553

Gretsi 2017, Juan-les-Pins

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), ...

Overview

Introduction

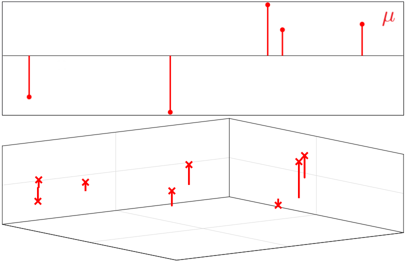
Low-Rank Semidefinite Primal

Toeplitz Relaxation

FFT-Based Conditional Gradient

Conclusion

Degradation Model



Radon measure on $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$

Initial measure:

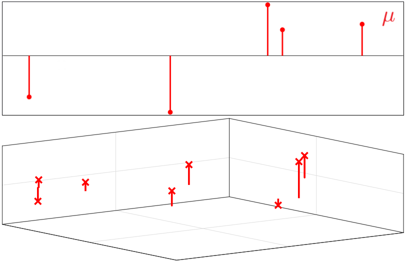
$\mu_{a,x} = \sum_{i=1}^{r} a_i \delta_{x_i}, a_i \in \mathbb{R}, x_i \in \mathbb{T}^d$

Degradation Model

Radon measure on $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$

Initial measure:

$$\mu_{a,x} = \sum_{i=1}^{r} a_i \delta_{x_i},\ a_i \in \mathbb{R},\ x_i \in \mathbb{T}^d$$



Linear measurements:

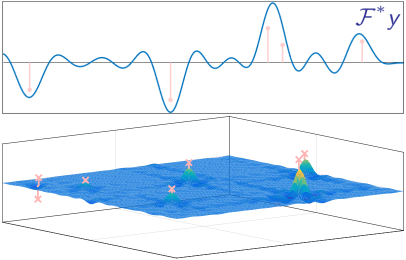
$$y = \mathcal{F}\mu_{a,x} + w$$

$\rightarrow$ Fourier measurements (cutoff frequency f_c)

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

for $k \in [\![-f_c; f_c]\!]^d$

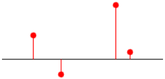
$\rightarrow$ Noise $w \in \mathbb{C}^{(2f_c+1)^d}$



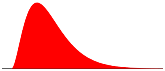
Sparse Recovery

Grid-free regularization: total variation of measures

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



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

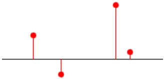


$|\mu|(\mathbb{T}^d) = \|f\|_{\mathrm{L}^1}$

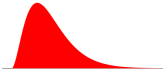
Sparse Recovery

Grid-free regularization: total variation of measures

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



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



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

BLASSO (de Castro and Gamboa [2012])

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

$$\max_{p \in \mathbb{C}^{n^d}} \left\{ \langle y, \, p \rangle - \frac{\lambda}{2}\|p\|^2 \; ; \; \|\mathcal{F}^* p\|_\infty \leqslant 1 \right\} \qquad (\mathcal{D}_\lambda)$$

Dual Certificate

$$\max\left\{\langle y,\ p\rangle - \tfrac{\lambda}{2}\|p\|^2\ ;\ \|\mathcal{F}^* p\|_\infty \leqslant 1\right\} \qquad (\mathcal{D}_\lambda)$$

$\Longrightarrow \boldsymbol{\eta_\lambda \overset{\text{def.}}{=} \mathcal{F}^* p_\lambda}$ **dual certificate**.

$\eta_\lambda(x) \;=\; \sum_{k\in[\![-f_c;f_c]\!]^d} p_k e^{-2i\pi\langle k,x\rangle},$
with $p_k = p^*_{-k}$.

Dual Certificate

$$\max \left\{ \langle y,\ p \rangle - \tfrac{\lambda}{2}\|p\|^2 \ ;\ \|\mathcal{F}^* p\|_\infty \leqslant 1 \right\} \quad (\mathcal{D}_\lambda)$$

$\Longrightarrow \boldsymbol{\eta_\lambda \overset{\text{def.}}{=} \mathcal{F}^* p_\lambda}$ **dual certificate**.

$$\eta_\lambda(x) \;=\; \sum_{k \in [\![-f_c; f_c]\!]^d} p_k e^{-2i\pi\langle k, x\rangle},$$

with $p_k = p_{-k}^*$.

Optimality relation: $\eta_\lambda \in \partial|\mu_\lambda|(X)$

$$\mu_{a,x} \text{ solves } (\mathcal{P}_\lambda) \Leftrightarrow \begin{cases} \eta_\lambda(x_i) = \operatorname{sign}(a_i) \\ \|\eta_\lambda\|_\infty \leqslant 1 \end{cases}$$

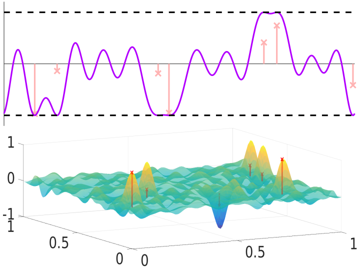


Figure: Dual polynomial

Related Works

- Support discretization $\longrightarrow$ LASSO - **Basis pursuit** (Donoho [1992])
  $\longrightarrow$ fast, inaccurate
- **Greedy support retrieval** (Bredies and Pikkarainen [2013])
  $\longrightarrow$ continuous setting, slow convergence
- **SDP relaxation** (Candès and Fernandez-Granda [2014])
  $\longrightarrow$ simple, stable, not scalable
- Non-variational schemes: **MUSIC** (Schmidt [1986])
  $\longrightarrow$ robust, scalable
  and **Prony's method** (Kunis et al. [2016]), **FRI-based methods** (Wei and Dragotti [2016]), ...

grid-free

Contributions

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

Overview

Semidefinite Hierarchies

- Encode measure in terms of moments
- Moment matrix satisfies **PSD** **+** **Toeplitz** **constraints**
- Keep only moments up to order m

Moment-Relaxations

$$\min_{u,z,\tau} \quad u_0 + \tau + \frac{1}{2}\left\|\frac{y}{\lambda} + 2z\right\|^2$$

$$\text{s.t.} \begin{cases} (a) \quad \mathcal{R} = \begin{bmatrix} R & z \\ z^* & \tau \end{bmatrix} \succeq 0, \\ (b) \quad R = \sum_{\boldsymbol{k} \in \rrbracket -m;m \llbracket^d} u_{\boldsymbol{k}} \Theta_{\boldsymbol{k}} \end{cases} \qquad (\widetilde{\mathcal{P}}_\lambda)$$

$$\Theta_{\boldsymbol{k}} = \theta_{k_d} \otimes \ldots \otimes \theta_{k_1}, \text{ with } \begin{cases} \otimes \quad \text{Kronecker product,} \\ \theta_k = (\delta_{i,i+k})_{i=1,\ldots,m} = \begin{bmatrix} 0 & \ldots & 1 & 0 & \ldots \\ 0 & \ldots & 0 & 1 & \ldots \\ & \ddots & & \ddots & \ddots \end{bmatrix} \end{cases}$$

When Is This Relaxation Tight?

Proposition (Dumitrescu [2017])

When $d = 1$, the relaxation is tight for $m \geqslant 2f_c + 1$.

Proof is based on either

- Fejér-Riesz theorem (Dumitrescu [2017])
- Carathéodory-Toeplitz theorem (Tang et al. [2013])

When Is This Relaxation Tight?

Proposition (Dumitrescu [2017])

When $d = 1$, the relaxation is tight for $m \geqslant 2f_c + 1$.

Proof is based on either

- Fejér-Riesz theorem (Dumitrescu [2017])
- Carathéodory-Toeplitz theorem (Tang et al. [2013])

When $d = 2$, there exists an m s.t. the relaxation is tight.
When $d > 2$, we do not know in general.

Support Recovery

Proposition

Let $\mathcal{R}_\lambda = \begin{bmatrix} R_\lambda & z_\lambda \\ z_\lambda^* & \tau \end{bmatrix}$ *be a solution of* $(\widetilde{\mathcal{P}}_\lambda)$ *(suppose the relaxation is tight). The coefficients of the dual polynomial are given by*

$$p_\lambda = \frac{y}{\lambda} + 2z_\lambda$$

Proof.

Comes from primal-dual optimality relations $\square$

$\Longrightarrow$ Support recovered via root-finding on the dual certificate $\mathcal{F}^* p_\lambda$.

Low-Rank Structure

Proposition

When $d = 1$, $(\widetilde{\mathcal{P}}_\lambda)$ *admits a solution* $\mathcal{R}_\lambda$ *such that* $\operatorname{rank} \mathcal{R}_\lambda \leqslant s$, *where* s *is the number of spikes in* μ_λ *(solution of* $(\mathcal{P}_\lambda)$*).*

Low-Rank Structure

Proposition

When $d = 1$, $(\widetilde{\mathcal{P}}_\lambda)$ *admits a solution* $\mathcal{R}_\lambda$ *such that* $\operatorname{rank} \mathcal{R}_\lambda \leqslant s$, *where* s *is the number of spikes in* μ_λ *(solution of* $(\mathcal{P}_\lambda)$*).*

In dimension $d \geqslant 2$, we conjecture that this results holds

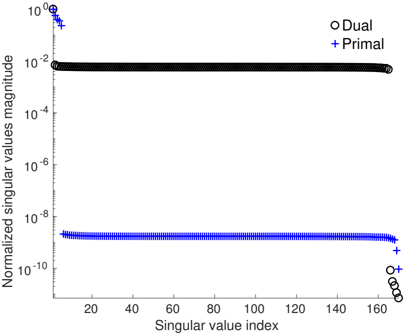


Figure: ($r = 5$ spikes, $f_c = 5$, $d = 2$). Singular values of primal and dual matrices

Overview

Penalized Primal

Numerical challenges:

- Handling both constraints simultaneously
- Toeplitz $\cap$ Low-Rank: analysis difficult

$$\boxed{\begin{aligned} &\min_{u,z,\tau} \; u_0 + \tau + \frac{1}{2}\Big\|\frac{y}{\lambda} + 2z\Big\|^2 \\ &\text{s.t.} \begin{cases} (a) & \mathcal{R} = \begin{bmatrix} R & z \\ z^* & \tau \end{bmatrix} \succeq 0 \\ (b) & R = \sum_{\boldsymbol{k}} u_{\boldsymbol{k}} \Theta_{\boldsymbol{k}} \end{cases} \end{aligned}} \quad (\widetilde{\mathcal{P}}_\lambda)$$

Penalized Primal

Numerical challenges:

- Handling both constraints simultaneously
- Toeplitz $\cap$ Low-Rank: analysis difficult

$$\min_{u,z,\tau} u_0 + \tau + \frac{1}{2}\|\frac{y}{\lambda} + 2z\|^2$$
$$\text{s.t.} \begin{cases} (a) & \mathcal{R} = \begin{bmatrix} R & z \\ z^* & \tau \end{bmatrix} \succeq 0 \\ (b) & R = \sum_{\boldsymbol{k}} u_{\boldsymbol{k}} \Theta_{\boldsymbol{k}} \end{cases} \qquad (\widetilde{\mathcal{P}}_\lambda)$$

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

$$\mathcal{R}_{\lambda,\rho} \in \operatorname*{argmin}_{R,u,z,\tau} \; u_0 + \tau + \frac{1}{2}\|\frac{y}{\lambda} + 2z\|^2 + \frac{1}{2\rho}\|R - \sum_{\boldsymbol{k}\in]\!]-m,m[\![^d} u_{\boldsymbol{k}}\Theta_{\boldsymbol{k}}\|^2 \qquad (\widetilde{\mathcal{P}}_{\lambda,\rho})$$
$$\text{s.t.} \quad \begin{bmatrix} R & z \\ z^* & \tau \end{bmatrix} \succeq 0.$$

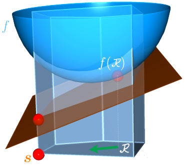
Overview

Alternating Descent Conditional Gradient Method

$$\left.\begin{array}{l} f \text{ convex, smooth} \\ K \text{ convex, bounded} \end{array}\right\} \min_{\mathcal{R}\in K} f(\mathcal{R})$$

Frank-Wolfe steps:

1. $\mathcal{S}^\star = \min_{\mathcal{S}\in\mathcal{D}} \langle \nabla f(\mathcal{R}_r), \mathcal{S} \rangle$
2. $\mathcal{R}_{r+1} = \mathcal{R}_r + c(\mathcal{S}^\star - \mathcal{R}_r)$, with $c \in [0, 1]$



Jaggi [2013]

Sparse iterates + simple LM

Slow convergence:

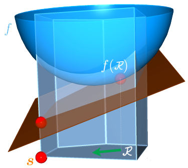
$f(x_r) - f(x^\star) \leqslant O(1/t)$

Alternating Descent Conditional Gradient Method

$\left.\begin{array}{l} f \text{ convex, smooth} \\ K \text{ convex, bounded} \end{array}\right\} \min_{\mathcal{R}\in K} f(\mathcal{R})$

Frank-Wolfe steps:

1. $\mathcal{S}^\star = \min_{\mathcal{S}\in\mathcal{D}} \langle \nabla f(\mathcal{R}_r), \mathcal{S}\rangle$
2. $\mathcal{R}_{r+1} = \mathcal{R}_r + c(\mathcal{S}^\star - \mathcal{R}_r)$, with $c \in [0,1]$



Jaggi [2013]

Sparse iterates + simple LM

Slow convergence:

$f(x_r) - f(x^\star) \leqslant O(1/t)$

- $K = \{X \succeq 0\,;\ \mathrm{tr}\, X \leqslant 1\}$
- Step 1. amounts to compute a leading eigenvector of ∇f.
- Low-cost storage: $\mathcal{R}_t = \mathcal{U}_t\mathcal{U}_t^*$
- Fixed-rank **BFGS** step (Boyd et al. [2015]) on $F : U \mapsto f(UU^*)$.

Algo: Recovering Dual Polynomial

Set: $U_0 = [0...0]^\top$, D_0: $\mathrm{tr}\,\mathcal{R}^\star \leqslant D_0$

For $r = 1, \ldots, N$ **do**

1. $v_r = D_0 \operatorname*{argmin}_{\|v\|\leqslant 1} v^\top \cdot \nabla f\,[U_r U_r^*] \cdot v$
2. $\widehat{U}_{r+1} = [\alpha_r U_r, \beta_r v_r]$, where $\alpha_r, \beta_r = \operatorname*{argmin}_{\alpha+\beta\leqslant 1} f(\alpha U_r U_r^* + \beta v_r v_r^*)$
3. $U_{r+1} \leftarrow$ **BFGS**$(F(U)$, init. at $\widehat{U}_{r+1})$

Fast-Fourier-Tranforms-Based Computations

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

$$\nabla f(UU^*) = \begin{bmatrix} \frac{1}{n} I_n & \frac{y}{\lambda} + 2z \\ \frac{y}{\lambda} + 2z & 1 \end{bmatrix} + \frac{1}{\rho}(UU^* - P_{\mathcal{T}}(UU^*))$$

- Main costly operation: $P_{\mathcal{T}}(UU^*) \cdot v$

Fast-Fourier-Tranforms-Based Computations

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

$$\nabla f(UU^*) = \begin{bmatrix} \frac{1}{n} I_n & \frac{y}{\lambda} + 2z \\ \frac{y}{\lambda} + 2z & 1 \end{bmatrix} + \frac{1}{\rho}(UU^* - P_{\mathcal{T}}(UU^*))$$

- Main costly operation: $P_{\mathcal{T}}(UU^*) \cdot v$

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

Toeplitz-Vector Multiplication

Let $x \in \mathbb{C}^{(n+1)^d}$, $t \in \mathbb{C}^{(2n+1)^d}$, and $T = \mathrm{Toep}_t$. 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 $P_{\mathcal{T}}[UU^*] = \mathrm{Toep}_t$, with

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

Numerics

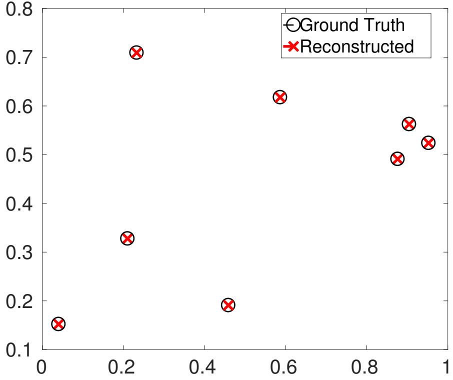


Figure: $r = 8$, $f_c = 13$, $\|w\| = 0.001\|y\|$, $\lambda = 1$, $\rho = 10$, total time: 315s, error (flat-norm): $\|\mu_{\lambda,\rho} - \mu_0\| = 4.57 \times 10^{-3}$

Numerics

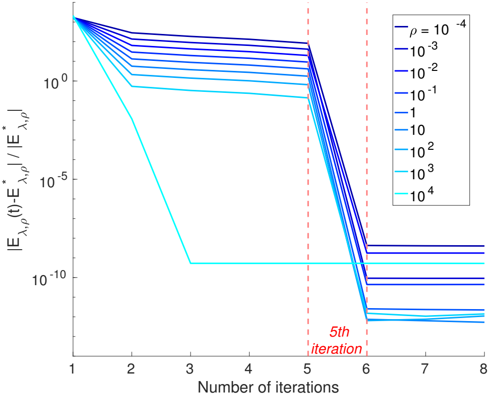


Figure: Relative error between $E_{\lambda,\rho}$ and E_λ. Initial measure has 5 spikes.

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
- Generalizable framework, to problems of the form

$$\min f(x) \quad \text{s.t.} \quad x \in A \cap B$$

with f smooth, A some convex hull, B an affine space on which we can project easily.

Merci pour votre attention!

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. $|X| = 1$

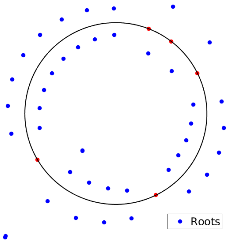


Figure: Roots of $1 - |P|^2$, with $P = \sum p_k X^k$

Sensitivity Analysis

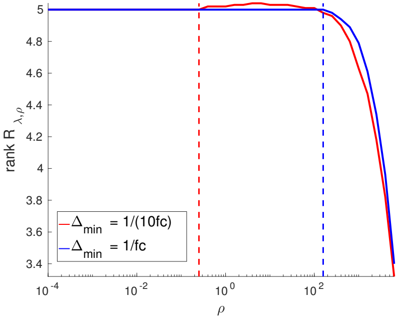


Figure: Rank of $\mathcal{R}_{\lambda,\rho}$ w.r.t. ρ

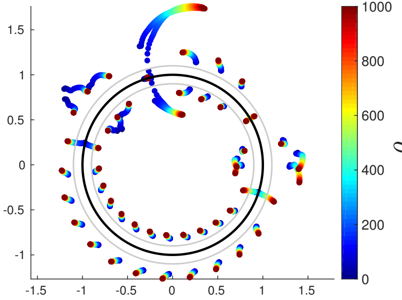


Figure: Roots trajectory w.r.t ρ.

Numerics

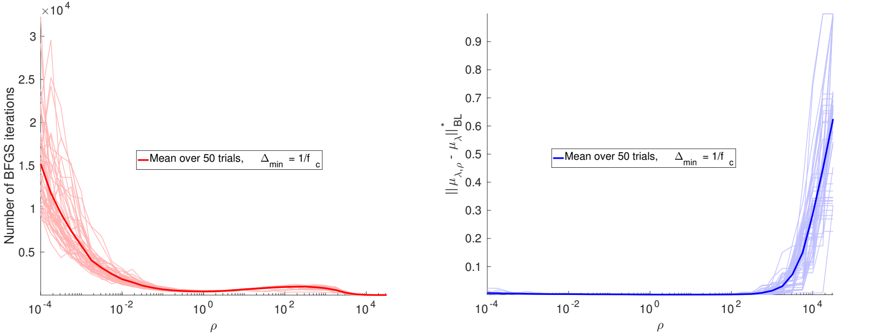


Figure: Left: Number of BFGS iterations, summed over all Frank-Wolfe steps. Right: Error between $\mu_{\lambda,\rho}$ and μ_{λ}, measured with the dual bounded Lipschitz norm.

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. Communications on pure and applied Mathematics, 67(6): 906–956, 2014.

Y. de Castro and F. Gamboa. Exact reconstruction using Beurling minimal extrapolation. Journal of Mathematical analysis and applications, 395(1): 336–354, 2012.

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

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

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

S. Kunis, T. Peter, T. Römer, and U. Von der Ohe. A multivariate generalization of prony's method. Linear Algebra and its Applications, 490:31–47, 2016.

R. Schmidt. Multiple emitter location and signal parameter estimation. Antennas and Propagation, IEEE Transactions on, 34(3):276–280, 1986.

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

X. Wei and P.L. Dragotti. FRESH – FRI-based single-image super-resolution algorithm. IEEE Trans. Image Processing, 25(8), 2016.