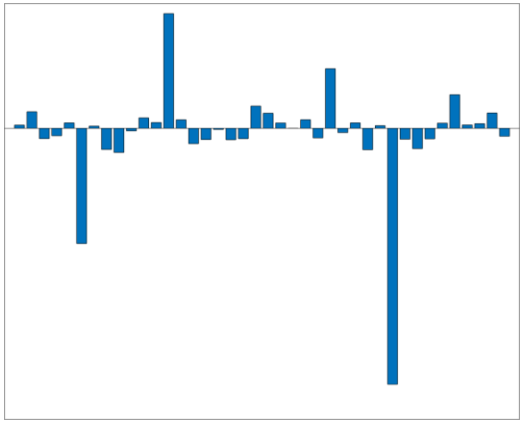
A dual certificate analysis of compressive off-the-grid recovery

Nicolas Keriven
Ecole Normale Supérieure (Paris)
CFM-ENS chair in Data Science

Joint work with **Clarice Poon** (Cambridge Uni.), **Gabriel Peyré** (ENS)

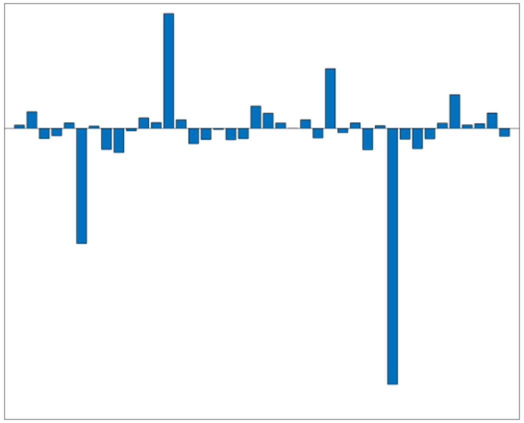
Journée GdR MIA, May 3rd 2018

Discrete compressive sensing

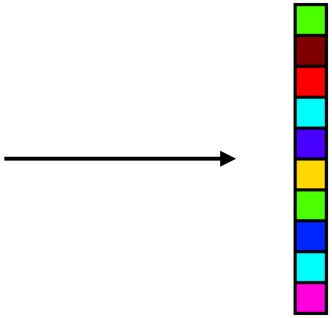


$$x \in \mathbb{R}^n$$

Discrete compressive sensing

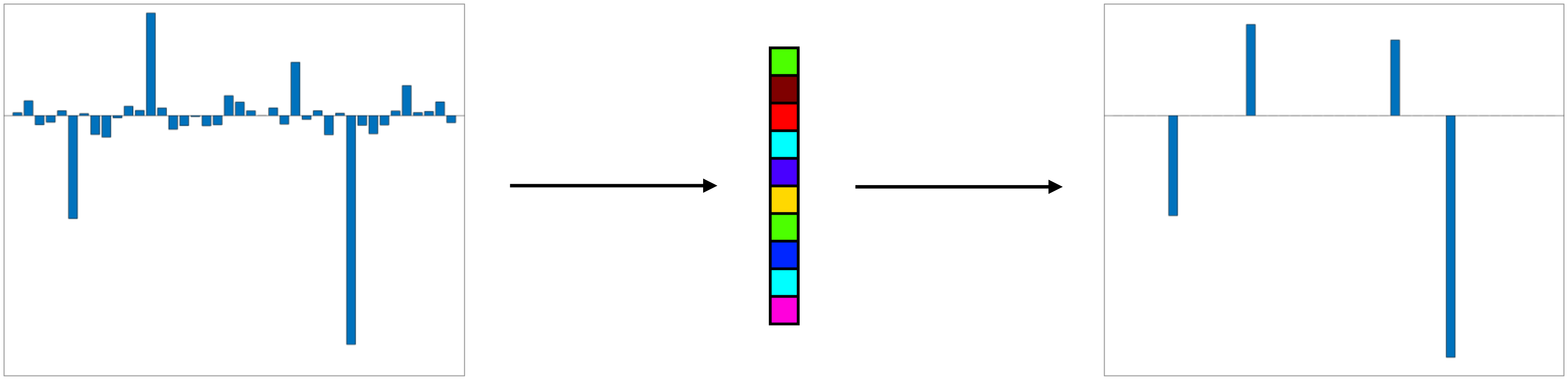


$x \in \mathbb{R}^n$



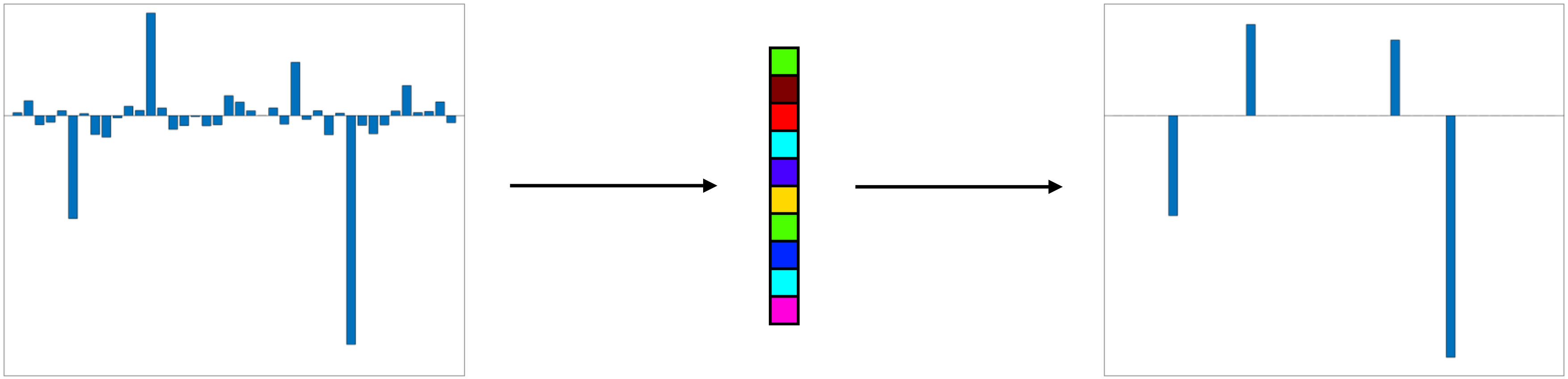
$y = Mx + e$

Discrete compressive sensing



$x \in \mathbb{R}^n$ $\quad$ $y = Mx + e$ $\quad$ $\tilde{x}$

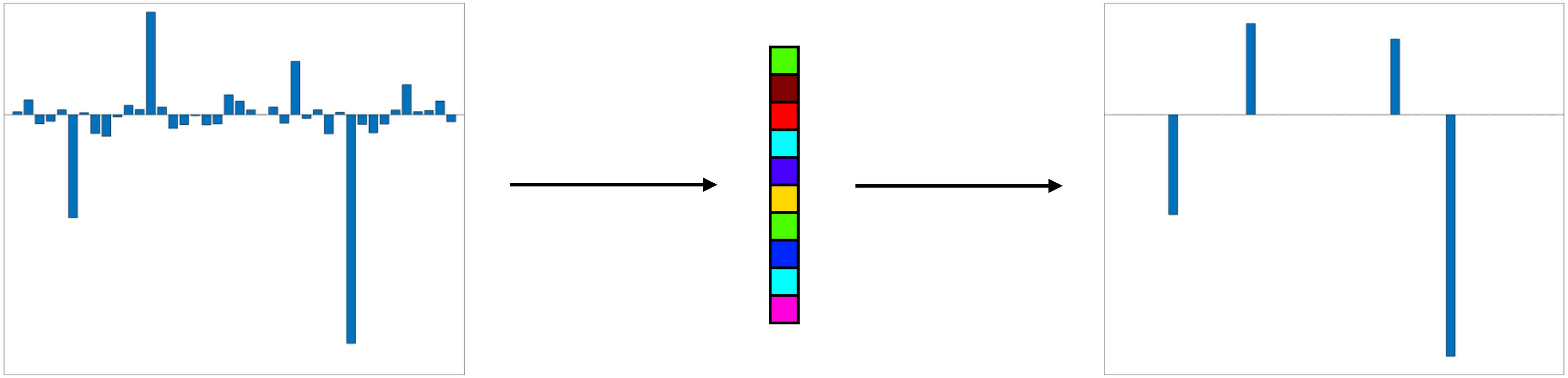
Discrete compressive sensing



$x \in \mathbb{R}^n$ $\qquad$ $y = Mx + e$ $\qquad$ $\tilde{x}$

- **Signal**: vector

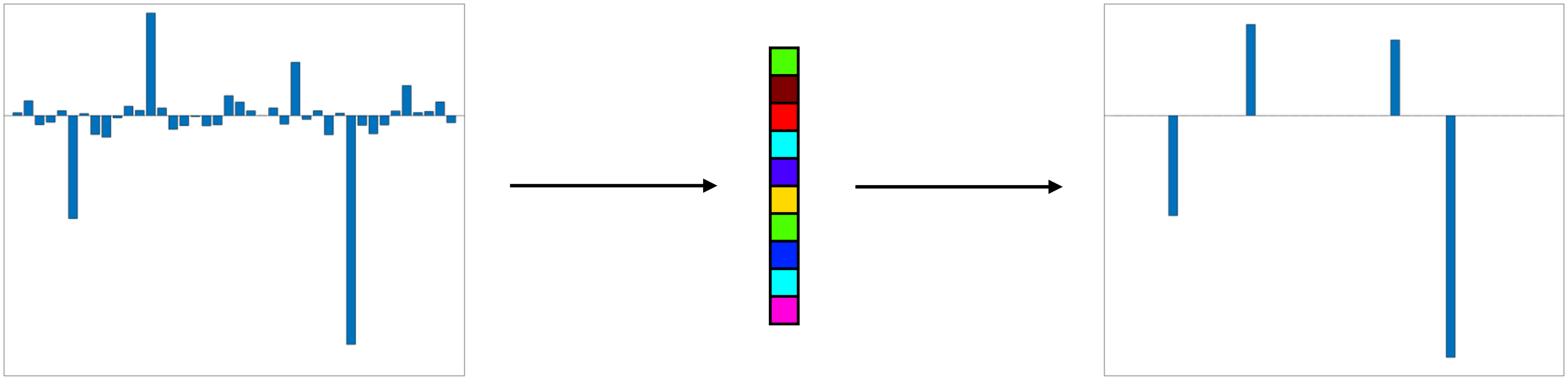
Discrete compressive sensing



$x \in \mathbb{R}^n$ $\quad$ $y = Mx + e$ $\quad$ $\tilde{x}$

- **Signal**: vector
- **Sparsity**: few non-zeros coefficients

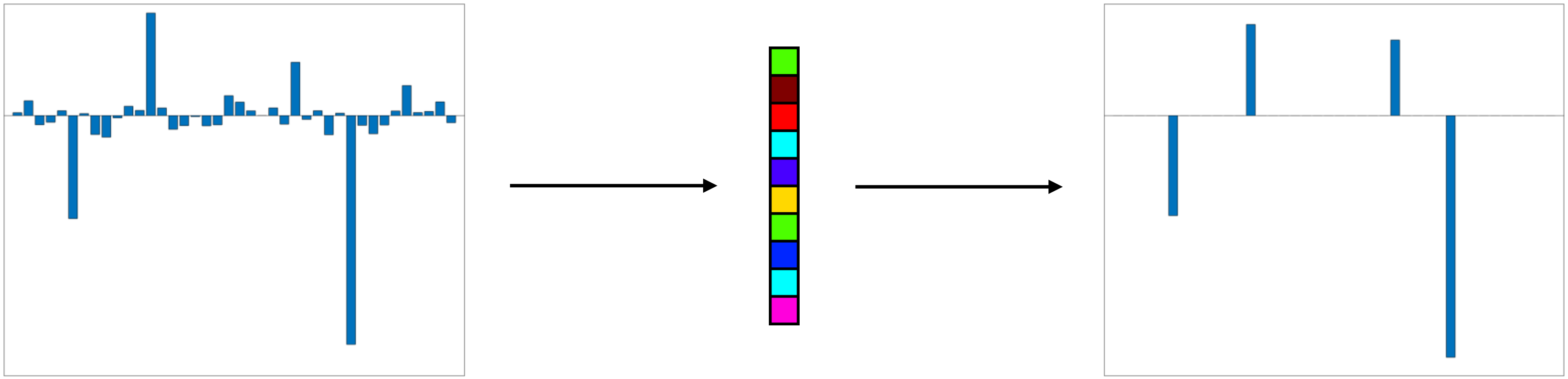
Discrete compressive sensing



$x \in \mathbb{R}^n$ $\quad$ $y = Mx + e$ $\quad$ $\tilde{x}$

- **Signal**: vector
- **Sparsity**: few non-zeros coefficients
- **Dimensionality reduction** (often random matrix)

Discrete compressive sensing



$x \in \mathbb{R}^n$ $\quad y = Mx + e \quad$ $\tilde{x}$

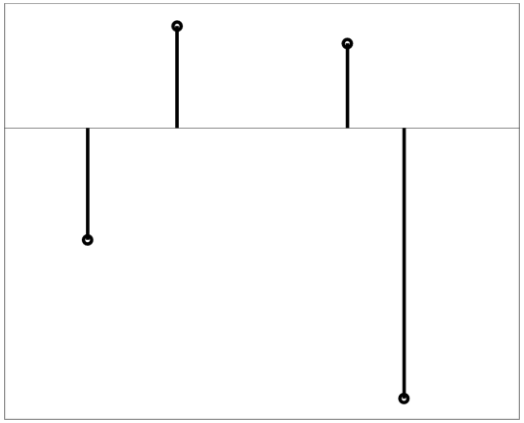
- **Signal**: vector
- **Sparsity**: few non-zeros coefficients
- **Dimensionality reduction** (often random matrix)
- **Recovery**: convex relaxation

$$\min_{\|x\|_0 \leq s} \|Mx - y\| \longrightarrow$$

LASSO

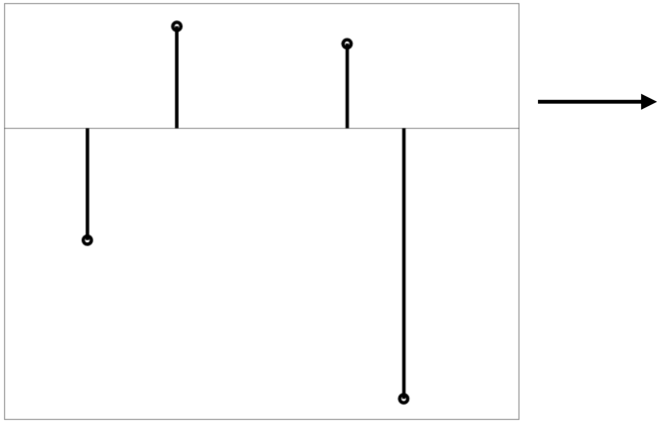
$$\min_x \frac{1}{2}\|Mx - y\|_2^2 + \lambda\|x\|_1$$

Off-the-grid recovery: « super-resolution »



$$\mu \in \mathcal{M}(\mathcal{X})$$

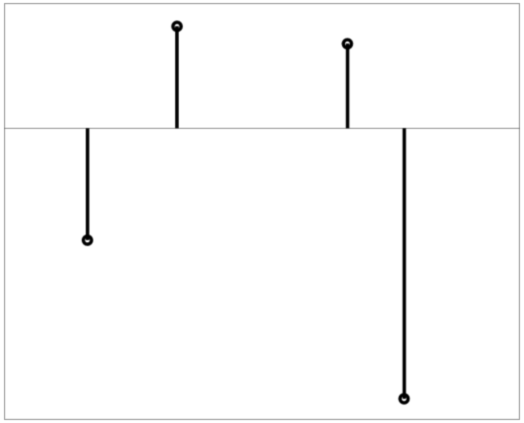
Off-the-grid recovery: « super-resolution »



$$y = \Phi\mu + e$$

$$\mu \in \mathcal{M}(\mathcal{X})$$

Off-the-grid recovery: « super-resolution »



$$\mu \in \mathcal{M}(\mathcal{X})$$

$$\longrightarrow \quad y = \Phi\mu + e$$

$$\Phi\mu = \int \varphi(x) d\mu(x)$$

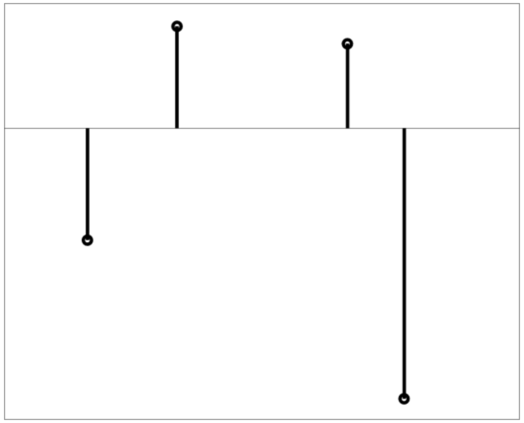
- Fourier

$$\varphi(x) = [e^{2i\pi xk}]_{|k|\leq f_c} \in \mathbb{C}^{2f_c+1}$$

- Convolution

$$\varphi(x) = f(x - \cdot) \in L^2(\mathcal{X})$$

Off-the-grid recovery: « super-resolution »



$\mu \in \mathcal{M}(\mathcal{X})$

$$y = \Phi\mu + e$$

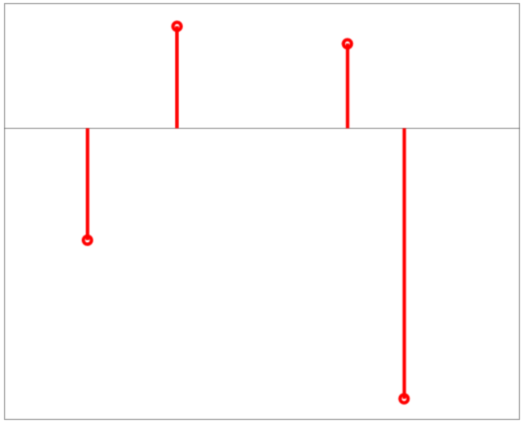
$$\Phi\mu = \int \varphi(x) d\mu(x)$$

- Fourier

$\varphi(x) = [e^{2i\pi xk}]_{|k| \leq f_c} \in \mathbb{C}^{2f_c+1}$

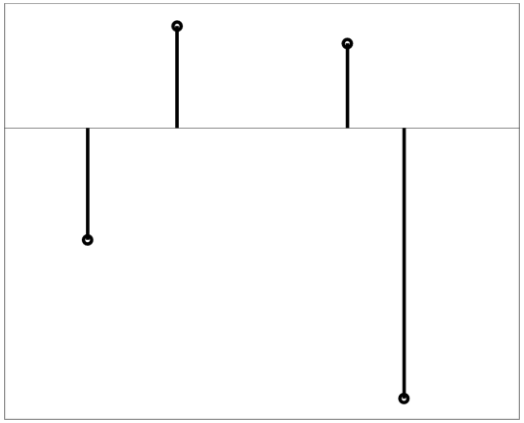
- Convolution

$\varphi(x) = f(x - \cdot) \in L^2(\mathcal{X})$



$\tilde{\mu}$

Off-the-grid recovery: « super-resolution »



$\mu \in \mathcal{M}(\mathcal{X})$

$$\longrightarrow y = \Phi\mu + e \longrightarrow$$

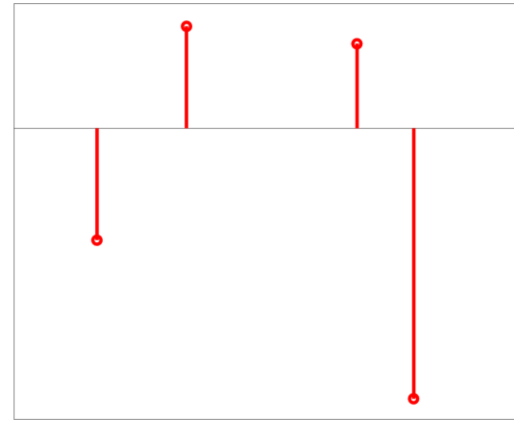
$$\Phi\mu = \int \varphi(x) d\mu(x)$$

- Fourier

$\varphi(x) = [e^{2i\pi xk}]_{|k| \leq f_c} \in \mathbb{C}^{2f_c+1}$

- Convolution

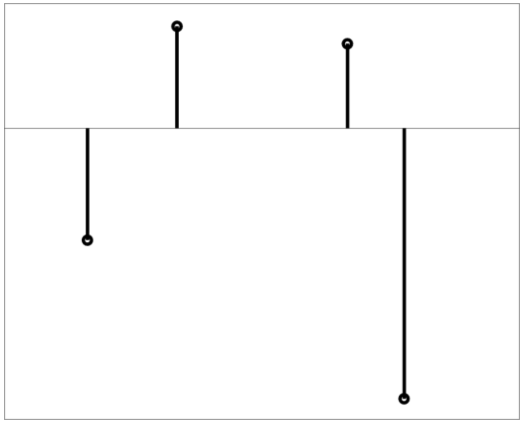
$\varphi(x) = f(x - \cdot) \in L^2(\mathcal{X})$



$\tilde{\mu}$

- **Signal**: Radon measure

Off-the-grid recovery: « super-resolution »



$\mu \in \mathcal{M}(\mathcal{X})$

$\longrightarrow y = \Phi\mu + e \longrightarrow$

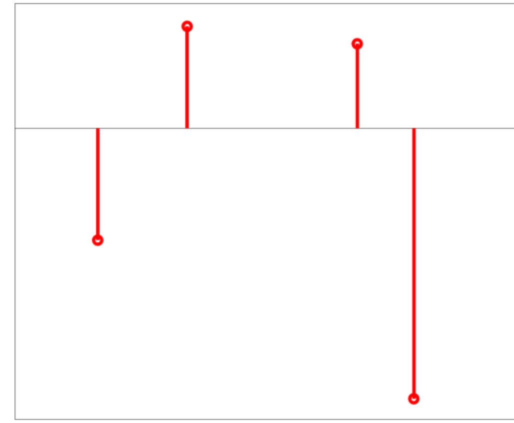
$$\Phi\mu = \int \varphi(x) d\mu(x)$$

- Fourier

$\varphi(x) = [e^{2i\pi x k}]_{|k| \leq f_c} \in \mathbb{C}^{2f_c+1}$

- Convolution

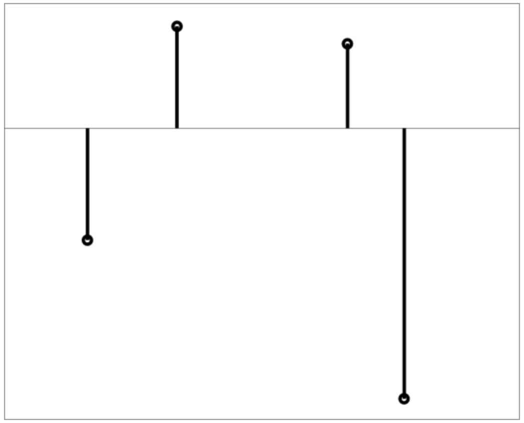
$\varphi(x) = f(x - \cdot) \in L^2(\mathcal{X})$



$\tilde{\mu}$

- **Signal**: Radon measure
- **Sparsity**: $\mu^\star = \sum_i a_i \delta_{x_i}$

Off-the-grid recovery: « super-resolution »



$\mu \in \mathcal{M}(\mathcal{X})$

$$\longrightarrow y = \Phi\mu + e \longrightarrow$$

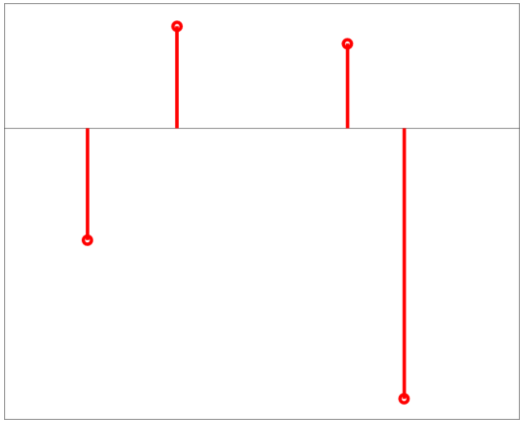
$$\Phi\mu = \int \varphi(x) d\mu(x)$$

- Fourier

$\varphi(x) = [e^{2i\pi xk}]_{|k| \leq f_c} \in \mathbb{C}^{2f_c+1}$

- Convolution

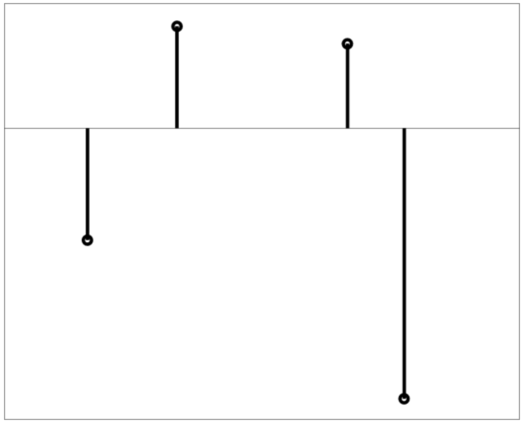
$\varphi(x) = f(x - \cdot) \in L^2(\mathcal{X})$



$\tilde{\mu}$

- **Signal**: Radon measure
- **Sparsity**: $\mu^\star = \sum_i a_i \delta_{x_i}$
- **Dimensionality reduction** (e.g. first Fourier coefficients)

Off-the-grid recovery: « super-resolution »



$\mu \in \mathcal{M}(\mathcal{X})$

$\longrightarrow y = \Phi\mu + e \longrightarrow$

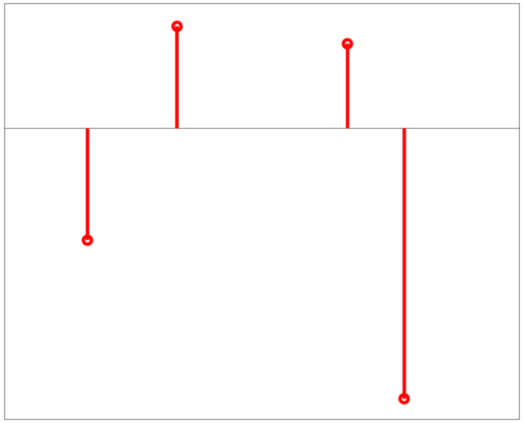
$\Phi\mu = \int \varphi(x) d\mu(x)$

- Fourier

$\varphi(x) = [e^{2i\pi xk}]_{|k|\leq f_c} \in \mathbb{C}^{2f_c+1}$

- Convolution

$\varphi(x) = f(x - \cdot) \in L^2(\mathcal{X})$



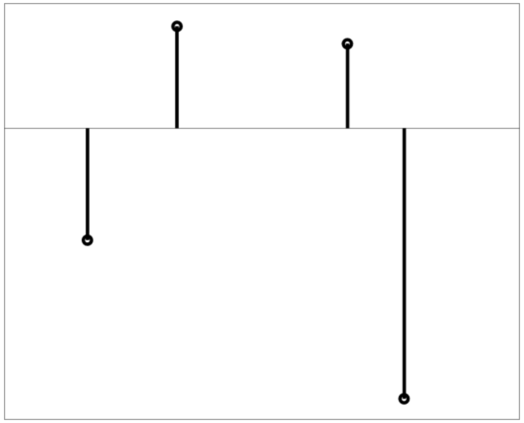
$\tilde{\mu}$

- **Signal**: Radon measure
- **Sparsity**: $\mu^\star = \sum_i a_i \delta_{x_i}$
- **Dimensionality reduction** (e.g. first Fourier coefficients)
- **Recovery**: convex relaxation?

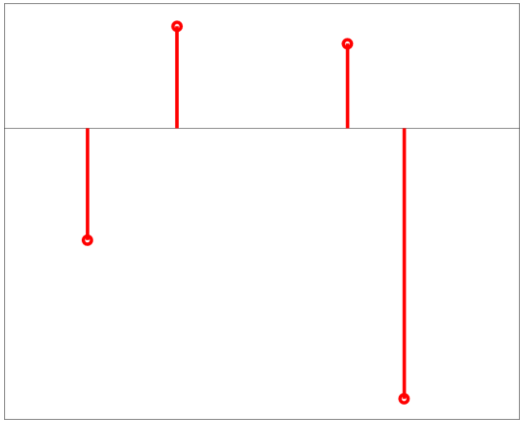
$$\min_{a,x} \|\Phi(\sum_i a_i \delta_{x_i}) - y\|_{\mathcal{H}}$$

See Keriven 2017, Gribonval 2017

Off-the-grid recovery: « super-resolution »



$\longrightarrow y = \Phi\mu + e \longrightarrow$



$\mu \in \mathcal{M}(\mathcal{X})$

$$\Phi\mu = \int \varphi(x) d\mu(x)$$

- Fourier

$\varphi(x) = [e^{2i\pi xk}]_{|k|\leq f_c} \in \mathbb{C}^{2f_c+1}$

- Convolution

$\varphi(x) = f(x - \cdot) \in L^2(\mathcal{X})$

$\tilde{\mu}$

- **Signal**: Radon measure
- **Sparsity**: $\mu^\star = \sum_i a_i \delta_{x_i}$
- **Dimensionality reduction** (e.g. first Fourier coefficients)
- **Recovery**: convex relaxation?

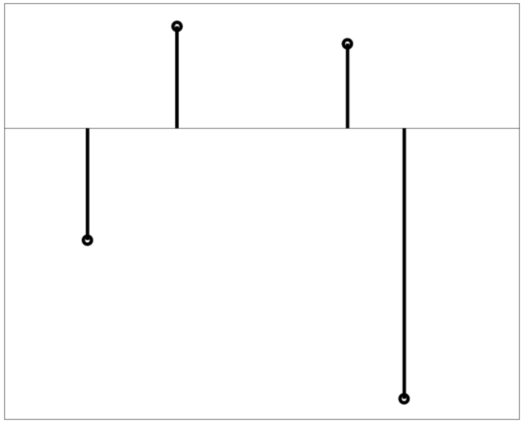
$\min_{a,x} \|\Phi(\sum_i a_i \delta_{x_i}) - y\|_{\mathcal{H}} \longrightarrow$

See Keriven 2017, Gribonval 2017

BLASSO ***[De Castro, Gamboa 2012]***

$$\min_\mu \frac{1}{2}\|\Phi\mu - y\|_{\mathcal{H}}^2 + \lambda|\mu|(\mathcal{X})$$

Off-the-grid recovery: « super-resolution »



$\mu \in \mathcal{M}(\mathcal{X})$

$\longrightarrow y = \Phi\mu + e \longrightarrow$

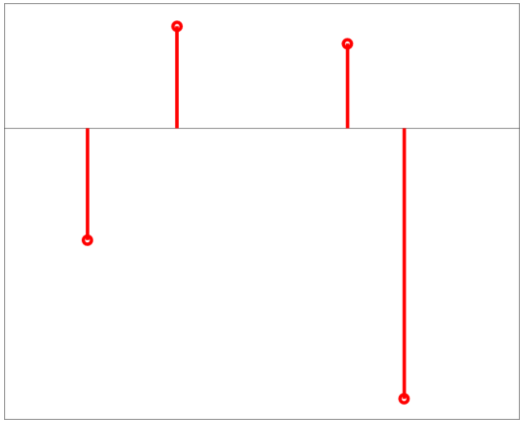
$\Phi\mu = \int \varphi(x) d\mu(x)$

- Fourier

$\varphi(x) = [e^{2i\pi x k}]_{|k| \leq f_c} \in \mathbb{C}^{2f_c+1}$

- Convolution

$\varphi(x) = f(x - \cdot) \in L^2(\mathcal{X})$



$\tilde{\mu}$

- **Signal**: Radon measure
- **Sparsity**: $\mu^\star = \sum_i a_i \delta_{x_i}$
- **Dimensionality reduction** (e.g. first Fourier coefficients)
- **Recovery**: convex relaxation?

$\min_{a,x} \|\Phi(\sum_i a_i \delta_{x_i}) - y\|_{\mathcal{H}} \longrightarrow$

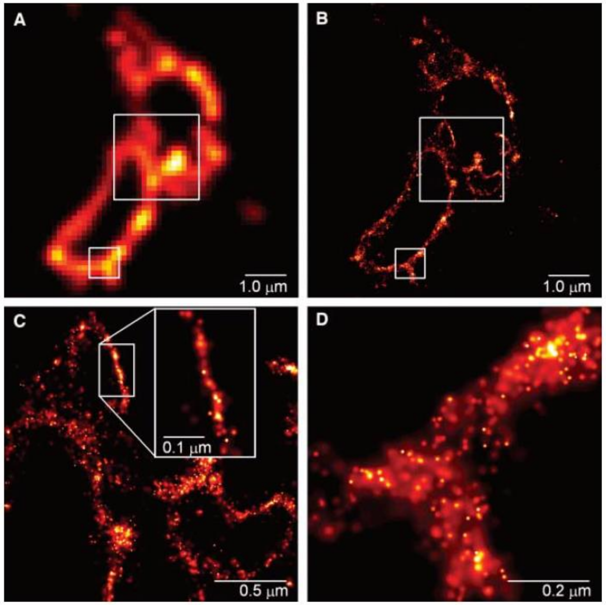
See Keriven 2017, Gribonval 2017

BLASSO ***[De Castro, Gamboa 2012]***

$$\min_\mu \frac{1}{2}\|\Phi\mu - y\|_{\mathcal{H}}^2 + \lambda|\mu|(\mathcal{X})$$

Other approaches: « Prony-like » ESPRIT, MUSIC… (but only 1d noiseless Fourier)

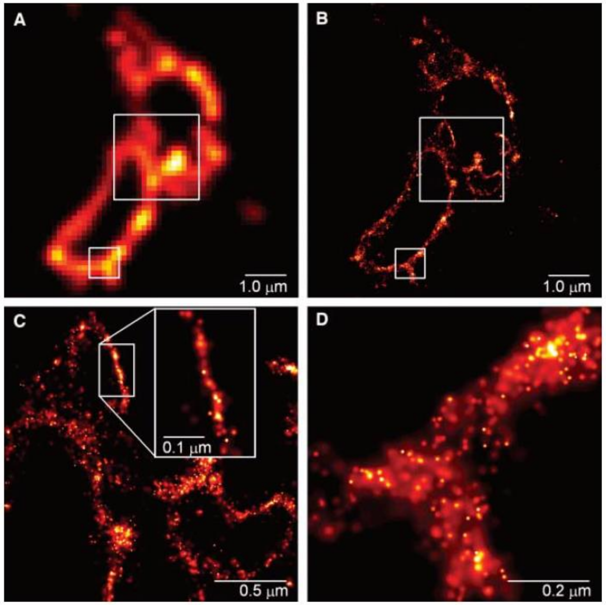
Example of applications



Fluorescence microscopy

[Betzig 2006]

Example of applications

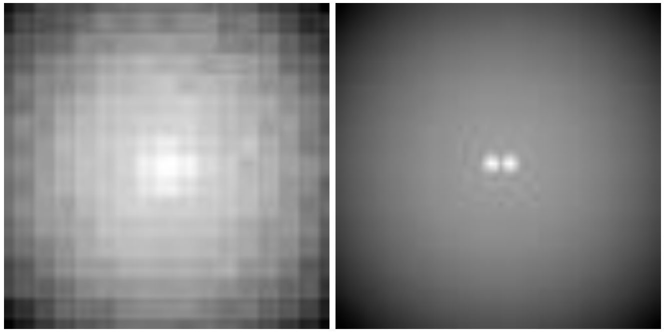


Fluorescence microscopy

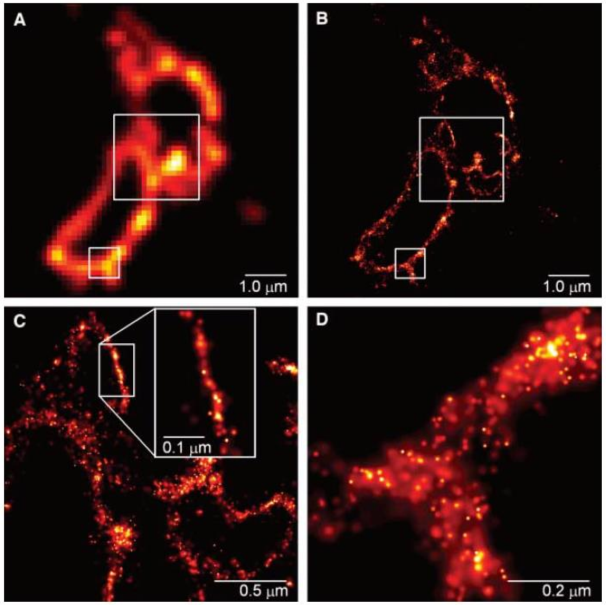
[Betzig 2006]

Astronomy

[Puschmann 2017]



Example of applications

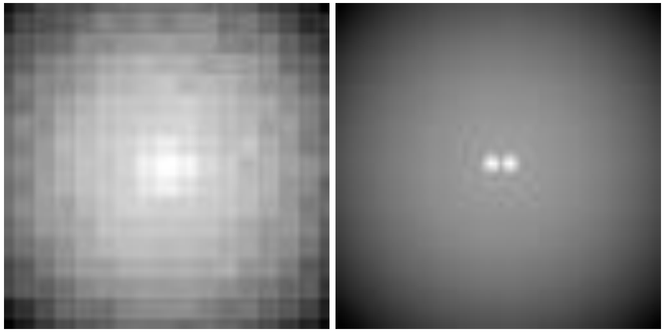


Fluorescence microscopy

[Betzig 2006]

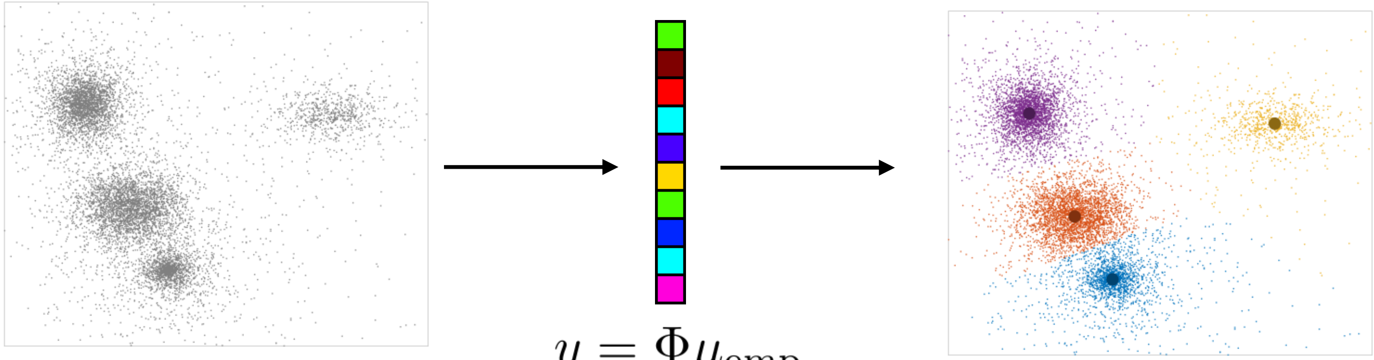
Astronomy

[Puschmann 2017]



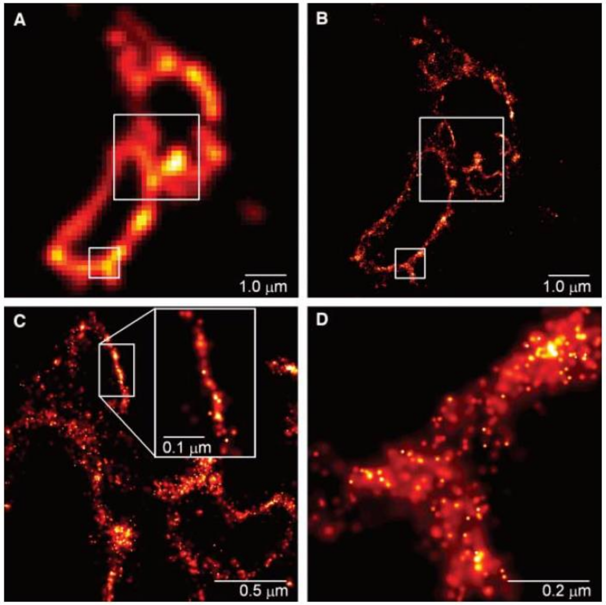
Compressive k-means (GMM…)

[Keriven 2017]



$$y = \Phi \mu_{\mathrm{emp.}}$$

Example of applications

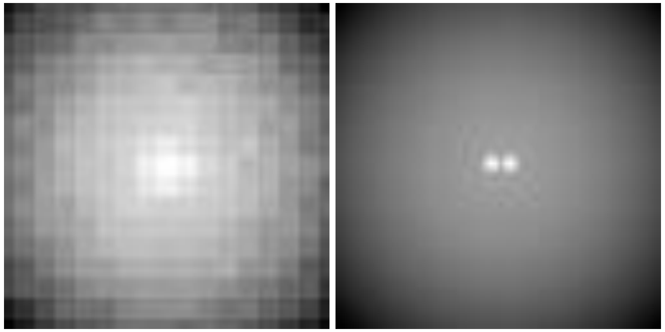


Fluorescence microscopy

[Betzig 2006]

Astronomy

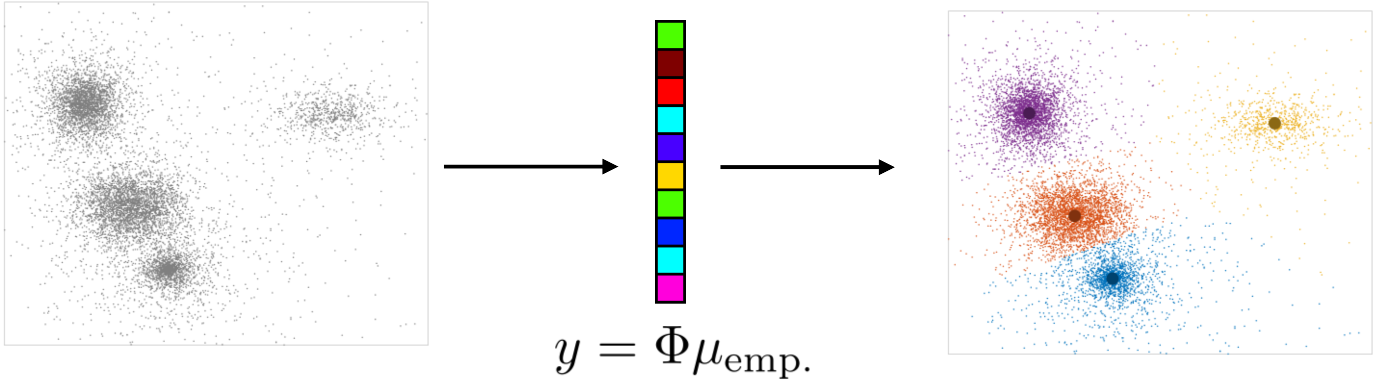
[Puschmann 2017]



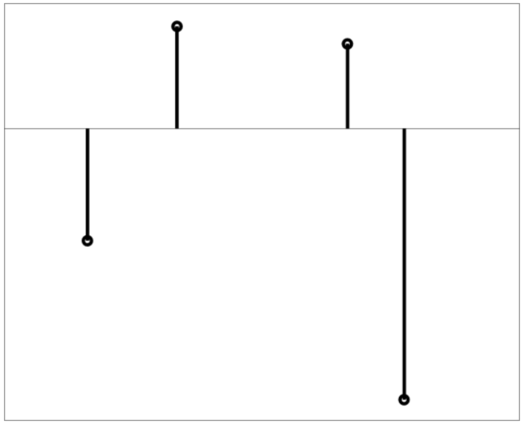
- Neuro-imaging with EEG *[Gramfort 2013]*
- 1-layer neural network *[Bach 2017]*
- Radar
- Geophysics
- …

Compressive k-means (GMM…)

[Keriven 2017]

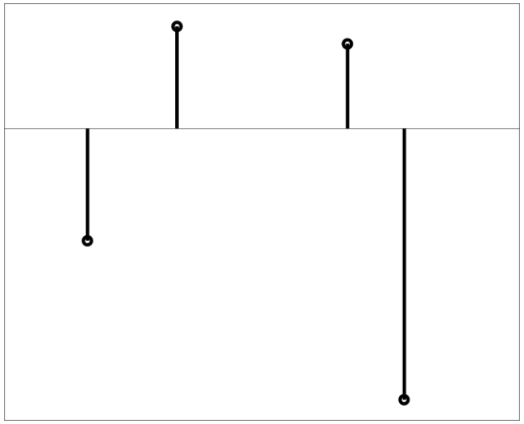


A seminal result *[Candès, Fernandez-Granda 2012]*



$$\mu \in \mathcal{M}(\mathbb{T})$$

A seminal result *[Candès, Fernandez-Granda 2012]*



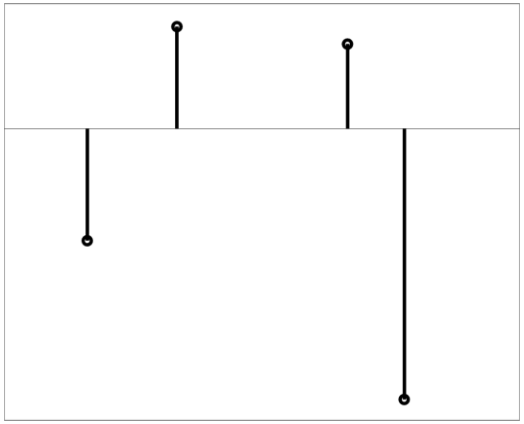
$\mu \in \mathcal{M}(\mathbb{T})$

→

First Fourier coefficients

$$[\int e^{2i\pi x k} d\mu(x)]_{|k| \leq f_c}$$

A seminal result *[Candès, Fernandez-Granda 2012]*

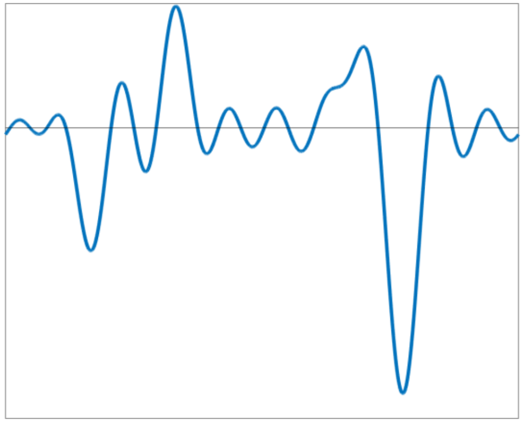


$\mu \in \mathcal{M}(\mathbb{T})$

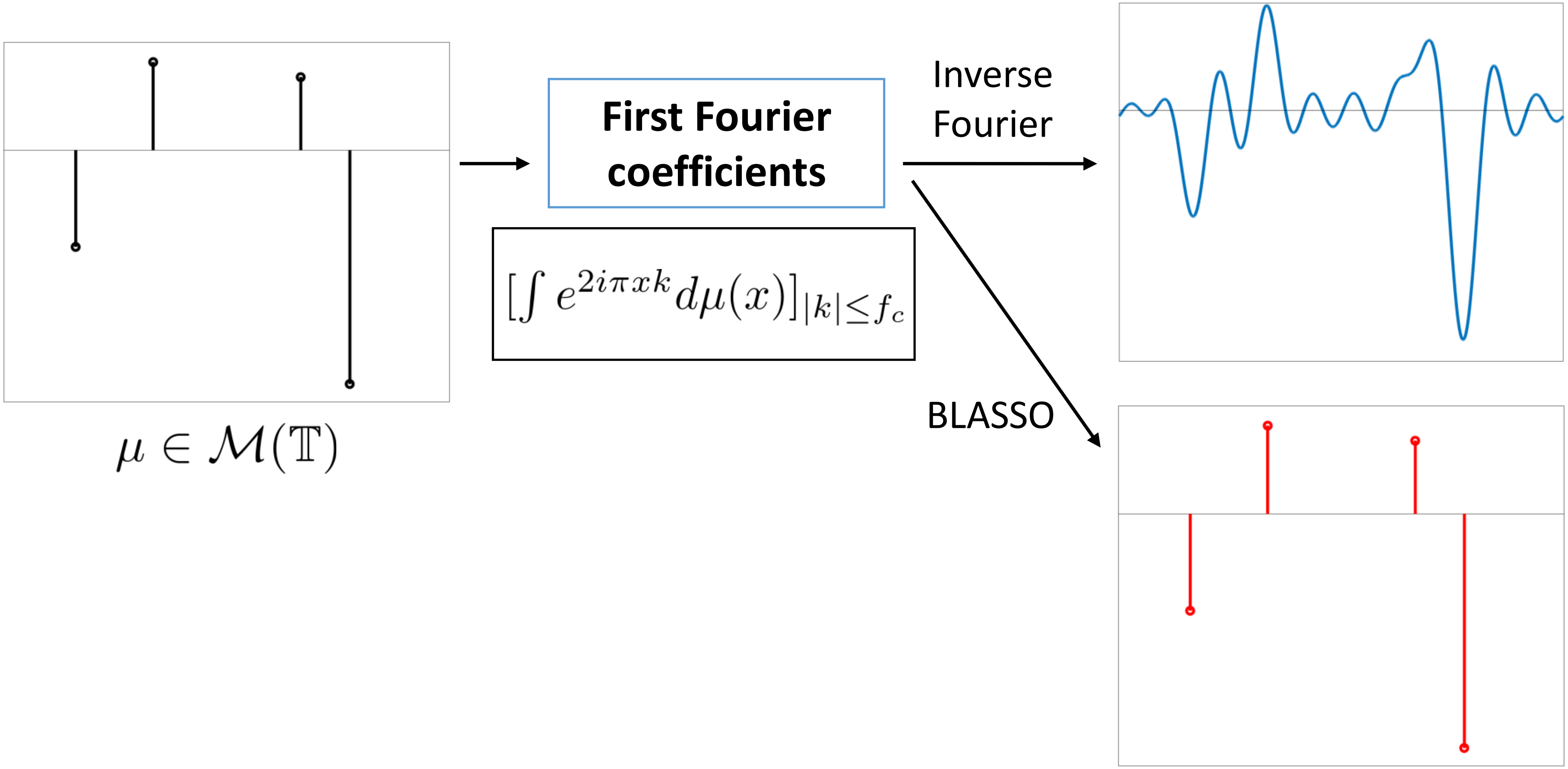
First Fourier coefficients

$$\left[\int e^{2i\pi x k} d\mu(x)\right]_{|k| \leq f_c}$$

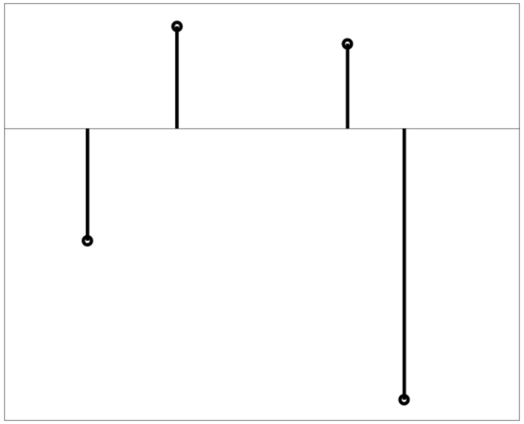
Inverse Fourier



A seminal result *[Candès, Fernandez-Granda 2012]*



A seminal result *[Candès, Fernandez-Granda 2012]*

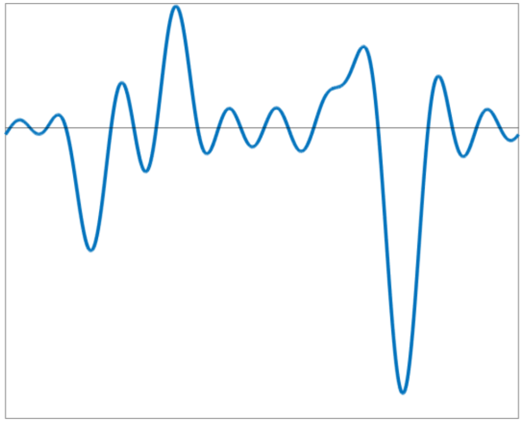


$\mu \in \mathcal{M}(\mathbb{T})$

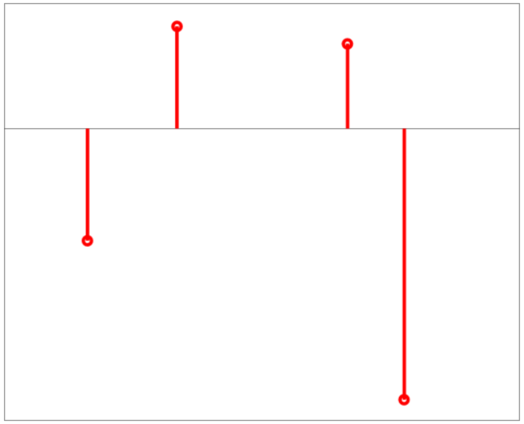
First Fourier coefficients

$[\int e^{2i\pi xk} d\mu(x)]_{|k|\leq f_c}$

Inverse Fourier

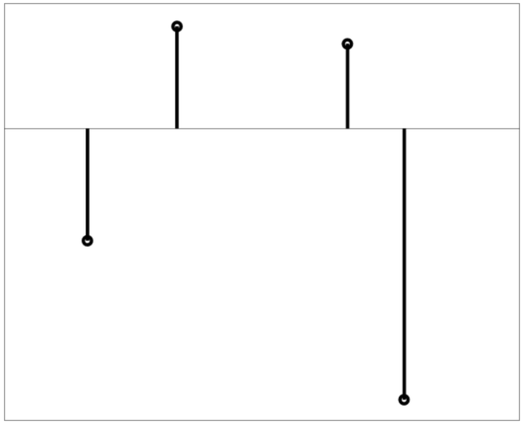


BLASSO



- Minimal separation $\Delta \geq \mathcal{O}(1/f_c)$

A seminal result *[Candès, Fernandez-Granda 2012]*

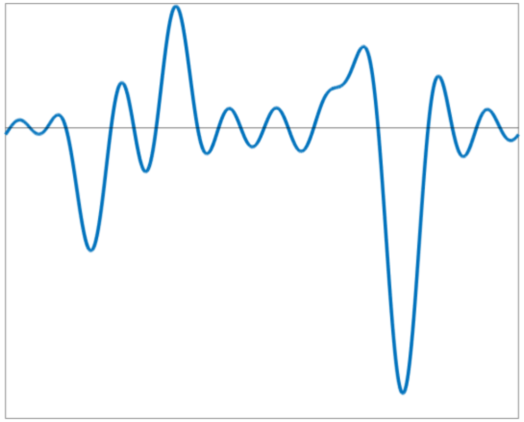


$\mu \in \mathcal{M}(\mathbb{T})$

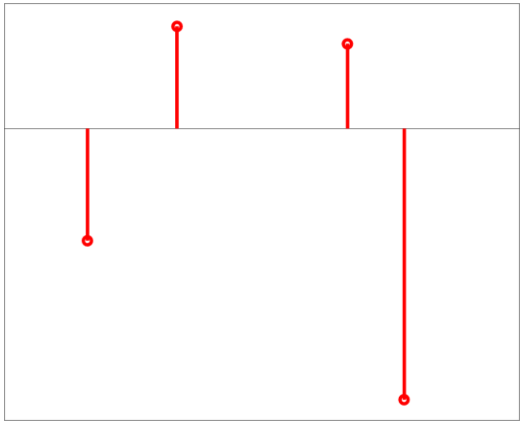
First Fourier coefficients

$$\left[\int e^{2i\pi x k} d\mu(x)\right]_{|k| \leq f_c}$$

Inverse Fourier

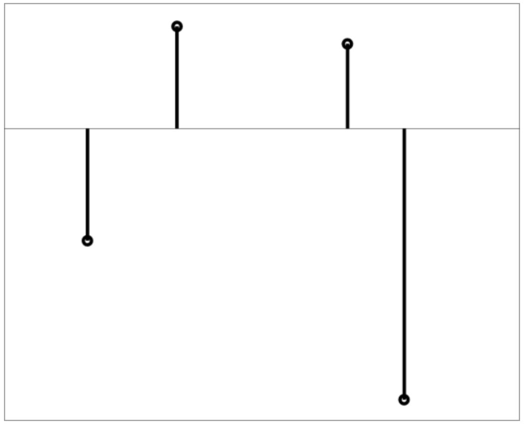


BLASSO



- Minimal separation $\Delta \geq \mathcal{O}(1/f_c)$
- **Regular Fourier** on the Torus

A seminal result *[Candès, Fernandez-Granda 2012]*

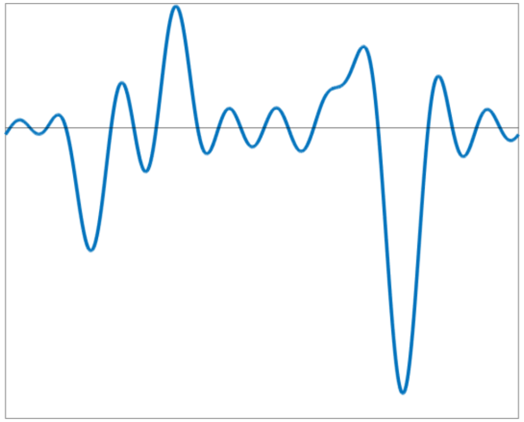


$\mu \in \mathcal{M}(\mathbb{T})$

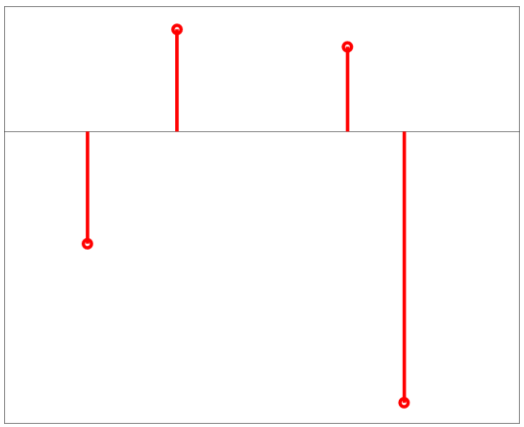
First Fourier coefficients

$[\int e^{2i\pi xk} d\mu(x)]_{|k|\leq f_c}$

Inverse Fourier

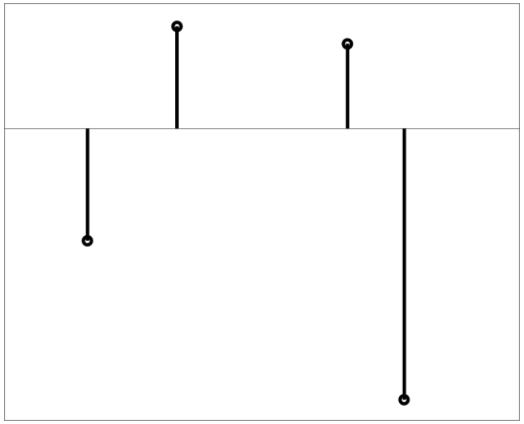


BLASSO



- Minimal separation $\Delta \geq \mathcal{O}(1/f_c)$
- **Regular Fourier** on the Torus
- **Reconstruction**: formulated as **SDP** (other: Frank-Wolfe, greedy, Prony-like…)

A seminal result *[Candès, Fernandez-Granda 2012]*

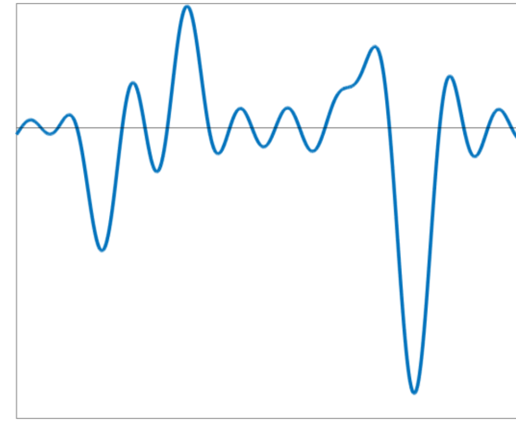


$\mu \in \mathcal{M}(\mathbb{T})$

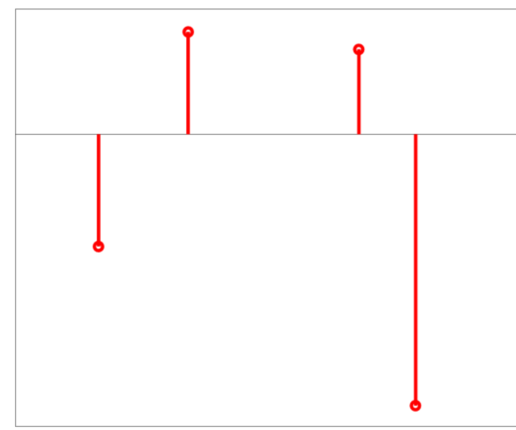
→ **First Fourier coefficients**

$$[\int e^{2i\pi xk} d\mu(x)]_{|k| \leq f_c}$$

Inverse Fourier →

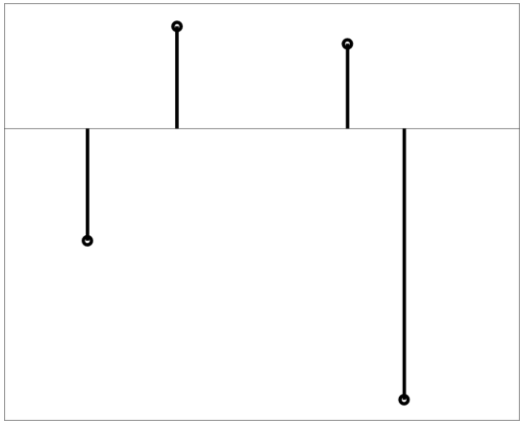


BLASSO ↘



- Minimal separation $\Delta \geq \mathcal{O}(1/f_c)$
- **Regular Fourier** on the Torus
- **Reconstruction**: formulated as **SDP** (other: Frank-Wolfe, greedy, Prony-like…)
- **Noise**: weak convergence (mass of $\tilde{\mu}$ ***concentrated*** around true Diracs)

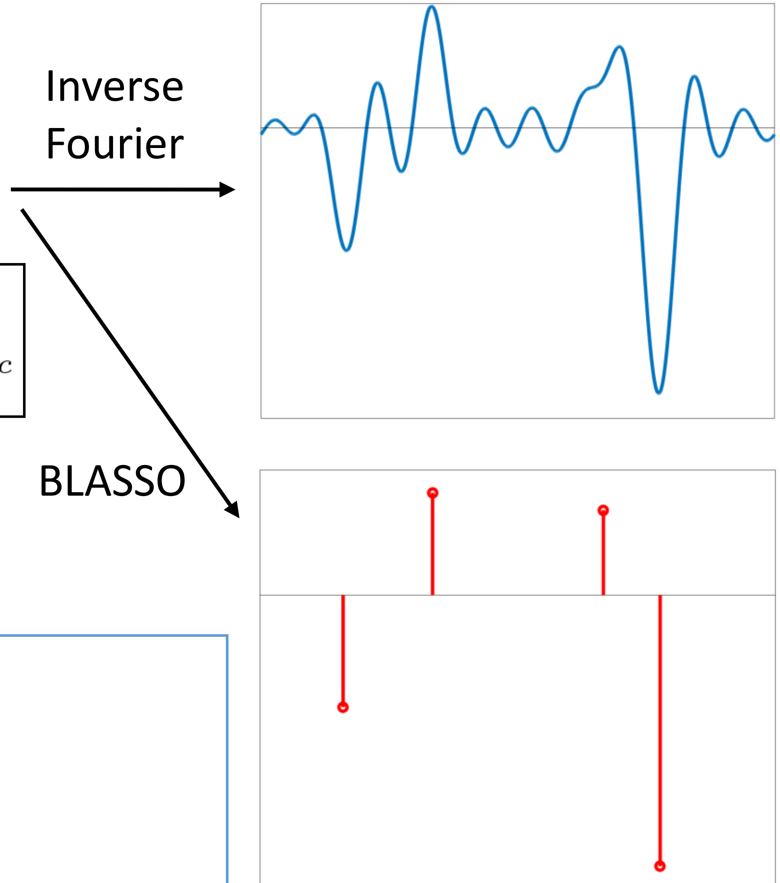
A seminal result *[Candès, Fernandez-Granda 2012]*



$\mu \in \mathcal{M}(\mathbb{T})$

First Fourier coefficients

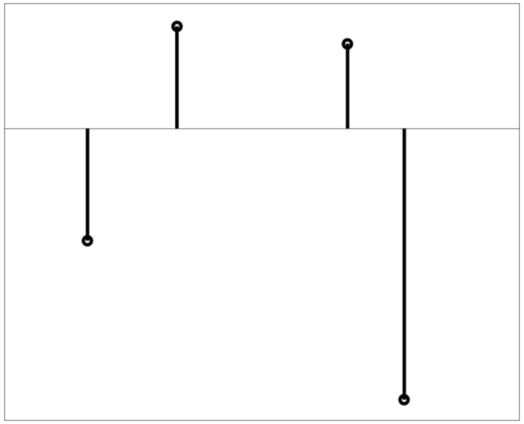
$$[\int e^{2i\pi xk} d\mu(x)]_{|k| \leq f_c}$$



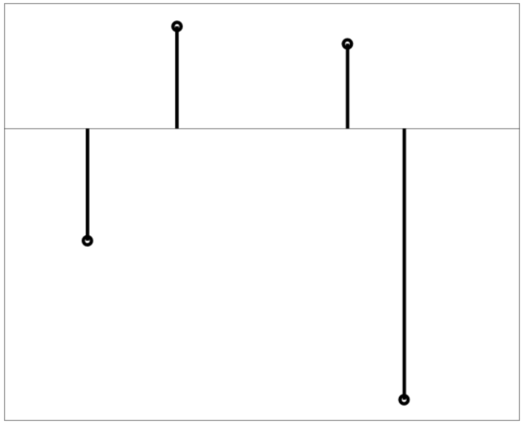
- Minimal separation $\Delta \geq \mathcal{O}(1/f_c)$
- **Regular Fourier** on the Torus
- **Reconstruction**: formulated as **SDP** (other: Frank-Wolfe, greedy, Prony-like…)
- **Noise**: weak convergence (mass of $\tilde{\mu}$ ***concentrated*** around true Diracs)

Randomness ? Number of measurements ?

Compressed sensing off-the-grid *[Tang, Recht 2013]*



Compressed sensing off-the-grid *[Tang, Recht 2013]*



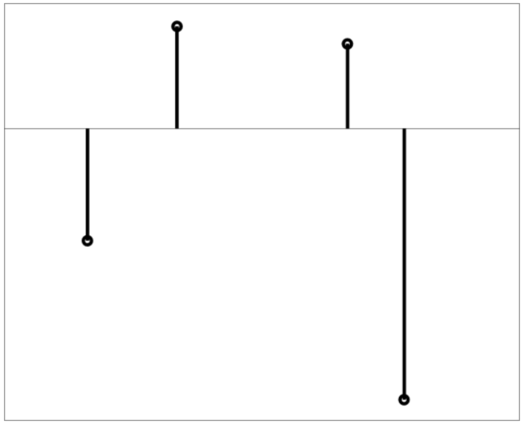
Random Fourier coefficients

$$\left[\int e^{2i\pi x k} d\mu(x)\right]_{k \in M}$$

$M \subset [-f_c, f_c]$ random

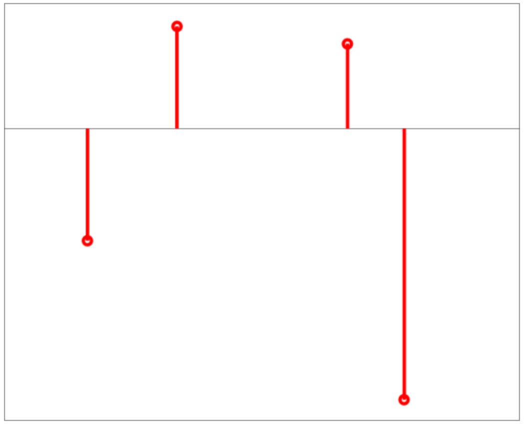
$|M| \geq \mathcal{O}(s \log(s) \log(f_c))$

Compressed sensing off-the-grid *[Tang, Recht 2013]*



Random Fourier coefficients

BLASSO

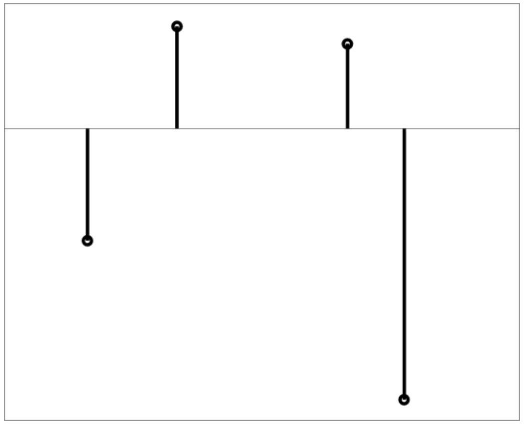


$$[\int e^{2i\pi xk} d\mu(x)]_{k \in M}$$

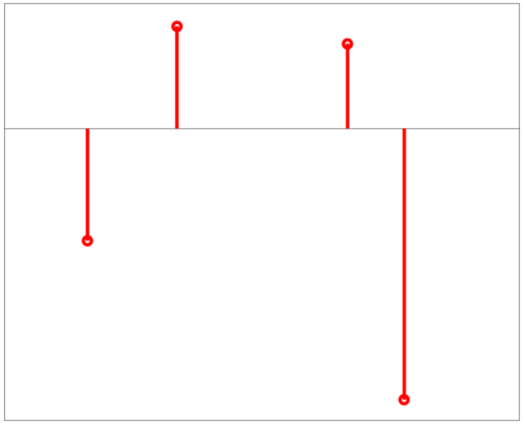
$M \subset [-f_c, f_c]$ random

$$|M| \geq \mathcal{O}(s \log(s) \log(f_c))$$

Compressed sensing off-the-grid *[Tang, Recht 2013]*



→ **Random Fourier coefficients** —BLASSO→



$[\int e^{2i\pi xk} d\mu(x)]_{k \in M}$

$M \subset [-f_c, f_c]$ random

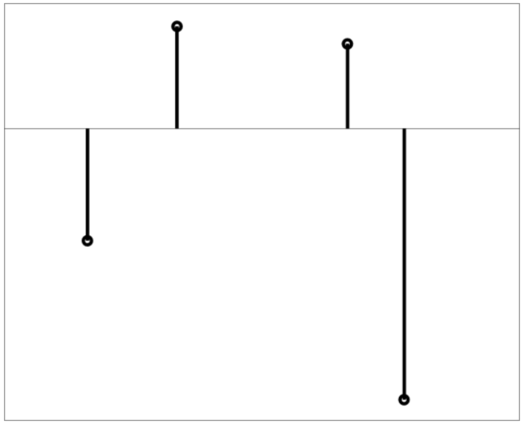
$|M| \geq \mathcal{O}(s \log(s) \log(f_c))$

- *As in compressive sensing*, **random Fourier sampling** is possible

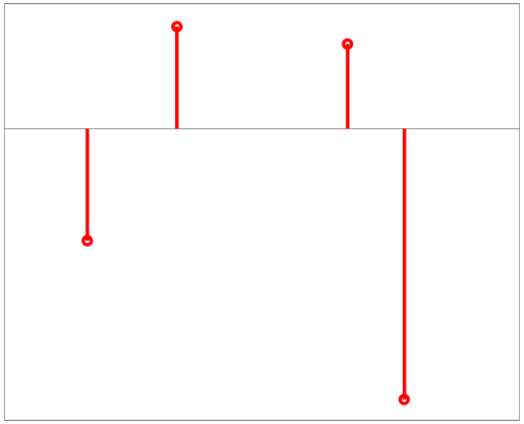
But:

- Limited to **1d Regular Fourier** (relies heavily on previous work by Candès)
- **Random signs assumption**

Compressed sensing off-the-grid *[Tang, Recht 2013]*



→ **Random Fourier coefficients** —BLASSO→



$[\int e^{2i\pi x k} d\mu(x)]_{k\in M}$

$M \subset [-f_c, f_c]$ random

$|M| \geq \mathcal{O}(s \log(s) \log(f_c))$

- *As in compressive sensing*, **random Fourier sampling** is possible

But:

- Limited to **1d Regular Fourier** (relies heavily on previous work by Candès)
- **Random signs assumption**

Questions:

- More general sampling scheme ?
- Multi-dimensional result ?
- **Get rid of random signs** ?

Outline

① Background on dual certificates

② Compressive off-the-grid recovery

③ Conclusion, outlooks

Dual certificates

Measurements

$$\Phi\mu = \int \varphi(x) d\mu(x)$$

$\varphi : \mathcal{X} \to \mathcal{H}$ Hilbert space

Dual certificates

Measurements

$$\Phi\mu = \int \varphi(x) d\mu(x)$$

$\varphi : \mathcal{X} \to \mathcal{H}$ Hilbert space

BLASSO

$$\min_\mu \frac{1}{2}\|\Phi\mu - y\|_{\mathcal{H}}^2 + \lambda|\mu|(\mathcal{X})$$

Dual certificates

Measurements

$$\Phi\mu = \int \varphi(x) d\mu(x)$$

$\varphi : \mathcal{X} \to \mathcal{H}$ Hilbert space

BLASSO

$$\min_\mu \frac{1}{2} \|\Phi\mu - y\|_{\mathcal{H}}^2 + \lambda |\mu|(\mathcal{X})$$

First-order conditions

μ_0 solution of BLASSO $\Leftrightarrow$ $\frac{1}{\lambda}\Phi^\star(\Phi\mu_0 - y) \in \partial|\mu_0|(\mathcal{X})$

Dual certificates

Measurements

$$\Phi\mu = \int \varphi(x) d\mu(x)$$

$\varphi : \mathcal{X} \to \mathcal{H}$ Hilbert space

BLASSO

$$\min_\mu \frac{1}{2}\|\Phi\mu - y\|_{\mathcal{H}}^2 + \lambda|\mu|(\mathcal{X})$$

First-order conditions

μ_0 solution of BLASSO $\iff \frac{1}{\lambda}\Phi^\star(\Phi\mu_0 - y) \in \partial|\mu_0|(\mathcal{X})$

Dual certificate **(noiseless case)**

μ_0 solution of $\min_{\Phi\mu = y} |\mu|(\mathcal{X})$ $\iff \mathrm{Im}(\Phi^\star) \cap \partial|\mu_0|(\mathcal{X}) \neq \emptyset$

What does it look like ?

What is a dual certificate ?

$$\eta \in \mathrm{Im}(\Phi^\star) \cap \partial|\mu_0|(\mathcal{X})$$

What does it look like ?

What is a dual certificate ?

$$\eta \in \mathrm{Im}(\Phi^\star) \cap \partial|\mu_0|(\mathcal{X})$$

$$\eta = \Phi^\star p = \langle p, \varphi(\cdot)\rangle_{\mathcal{H}} \in \mathcal{C}(\mathcal{X})$$

What does it look like ?

What is a dual certificate ?

$$\eta \in \mathrm{Im}(\Phi^\star) \cap \partial|\mu_0|(\mathcal{X})$$

$$\eta = \Phi^\star p = \langle p, \varphi(\cdot)\rangle_{\mathcal{H}} \in \mathcal{C}(\mathcal{X})$$

$$\|\eta\|_\infty \leq 1,\ \int \eta d\mu_0 = |\mu_0|(\mathcal{X})$$

What does it look like ?

What is a dual certificate ?

$$\eta \in \mathrm{Im}(\Phi^\star) \cap \partial|\mu_0|(\mathcal{X})$$

$$\eta = \Phi^\star p = \langle p, \varphi(\cdot)\rangle_{\mathcal{H}} \in \mathcal{C}(\mathcal{X})$$

$$\|\eta\|_\infty \leq 1,\ \int \eta d\mu_0 = |\mu_0|(\mathcal{X})$$

Case $\mu_0 = \sum_i a_i \pi_{x_i}$ **:**

What does it look like ?

What is a dual certificate ?

$$\eta \in \mathrm{Im}(\Phi^\star) \cap \partial|\mu_0|(\mathcal{X})$$

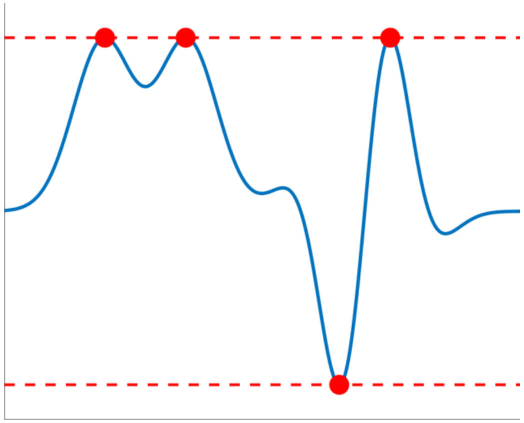
$$\eta = \Phi^\star p = \langle p, \varphi(\cdot)\rangle_{\mathcal{H}} \in \mathcal{C}(\mathcal{X})$$

$$\|\eta\|_\infty \leq 1, \ \int \eta d\mu_0 = |\mu_0|(\mathcal{X})$$

Case $\mu_0 = \sum_i a_i \pi_{x_i}$ **:**

$$\eta(x_i) = \mathrm{sign}(a_i)$$

$$\|\eta\|_\infty \leq 1$$



What does it look like ?

What is a dual certificate ?

$$\eta \in \mathrm{Im}(\Phi^\star) \cap \partial|\mu_0|(\mathcal{X})$$

$$\eta = \Phi^\star p = \langle p, \varphi(\cdot)\rangle_{\mathcal{H}} \in \mathcal{C}(\mathcal{X})$$

$$\|\eta\|_\infty \leq 1, \ \int \eta d\mu_0 = |\mu_0|(\mathcal{X})$$

Case $\mu_0 = \sum_i a_i \pi_{x_i}$:

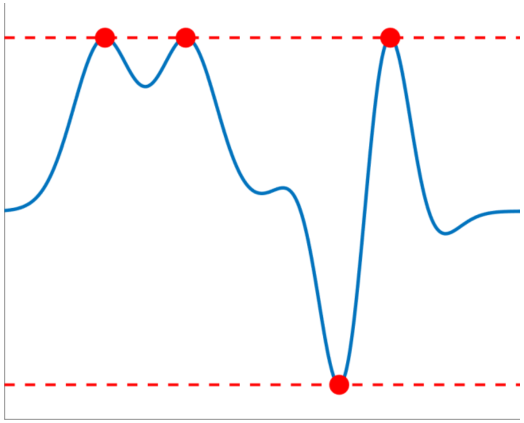
$$\eta(x_i) = \mathrm{sign}(a_i)$$

$$\|\eta\|_\infty \leq 1$$

Non-degenerate dual certif.

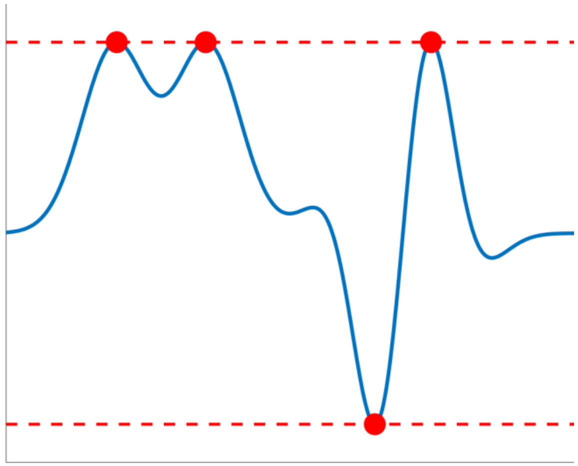
$$|\eta(x)| < 1$$

$$\mathrm{sign}(a_i)\nabla^2\eta(x_i) \prec 0$$

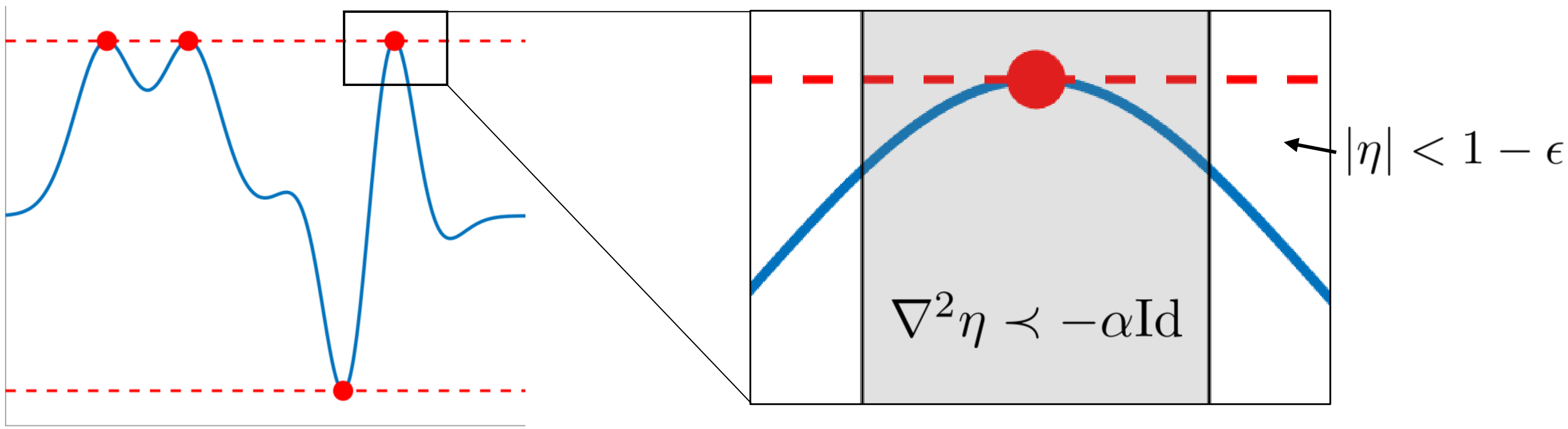


Ensures **uniqueness** and **robustness**…

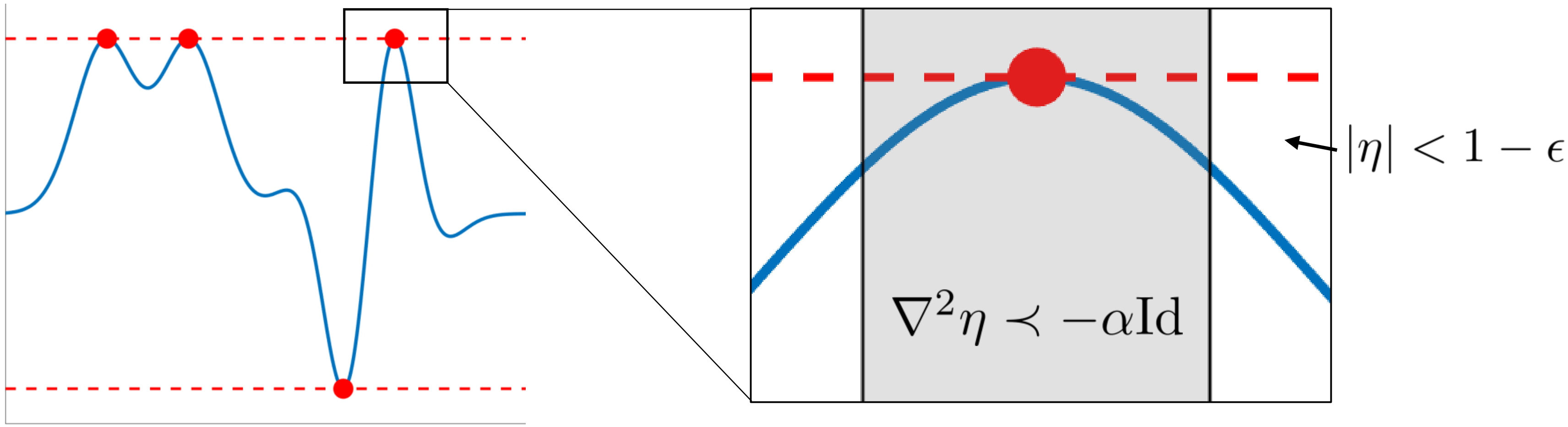
Recovery results



Recovery results

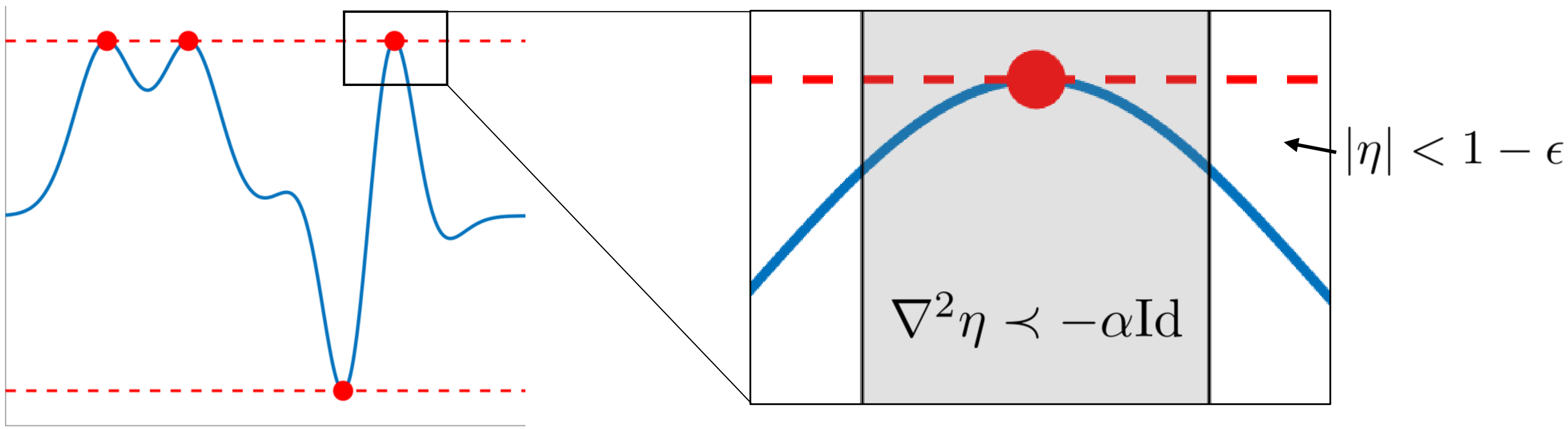


Recovery results



Theorem *(refinement of [Azaïs et al. 2015])*

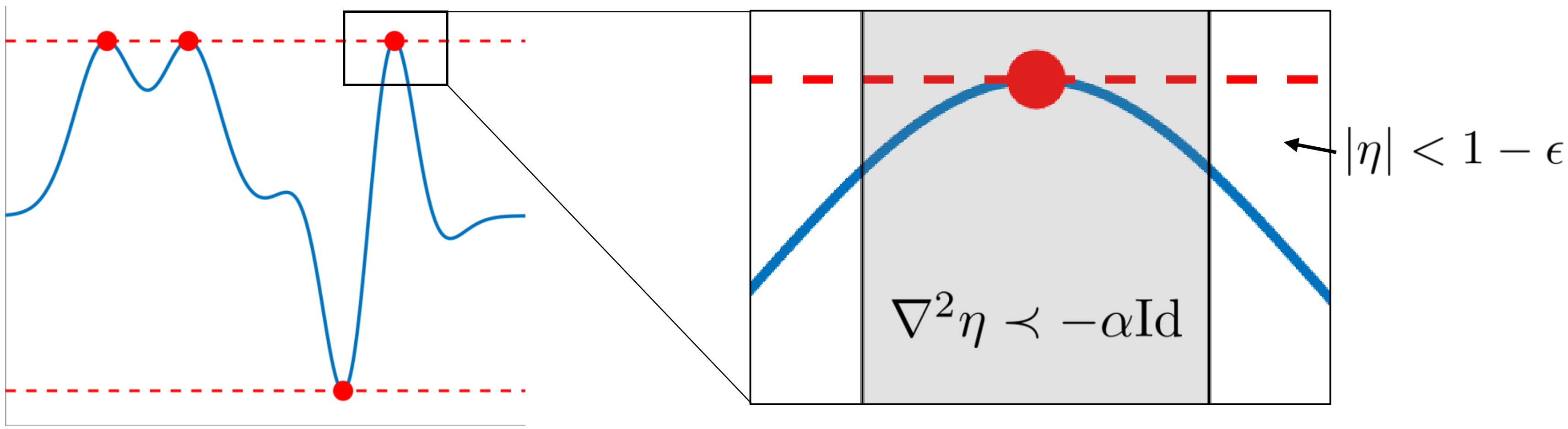
Recovery results



Theorem *(refinement of [Azaïs et al. 2015])*

Hyp: there exists a ND dual certif.

Recovery results



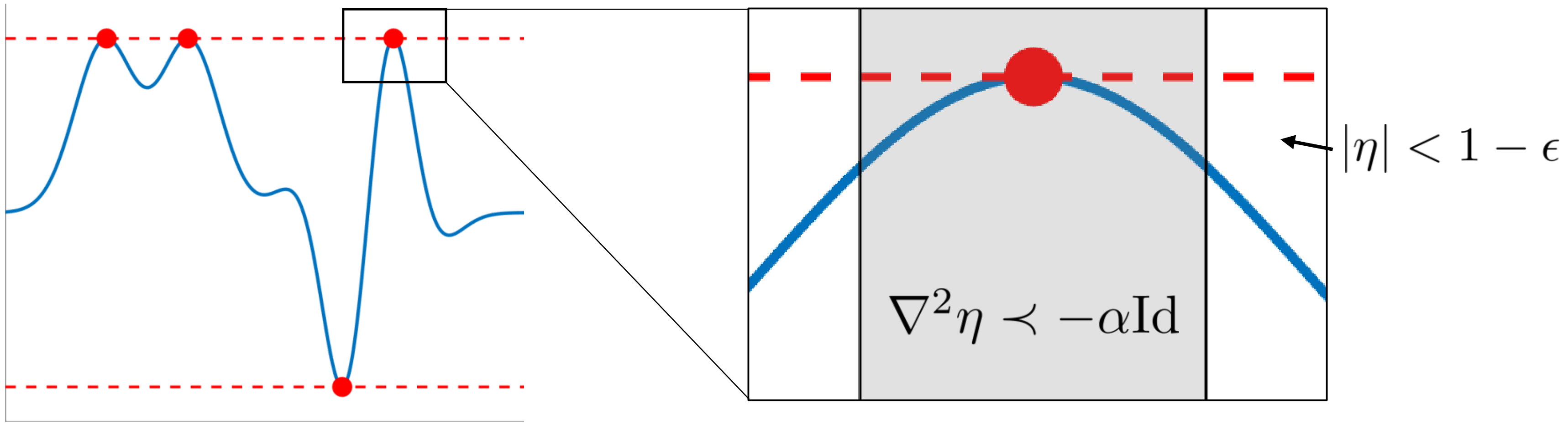
Theorem *(refinement of [Azaïs et al. 2015])*

Hyp: there exists a ND dual certif.

Result: *(wrt Bregman divergence)*

- $\tilde{\mu}$ **not necessarily sparse**, but
-

Recovery results



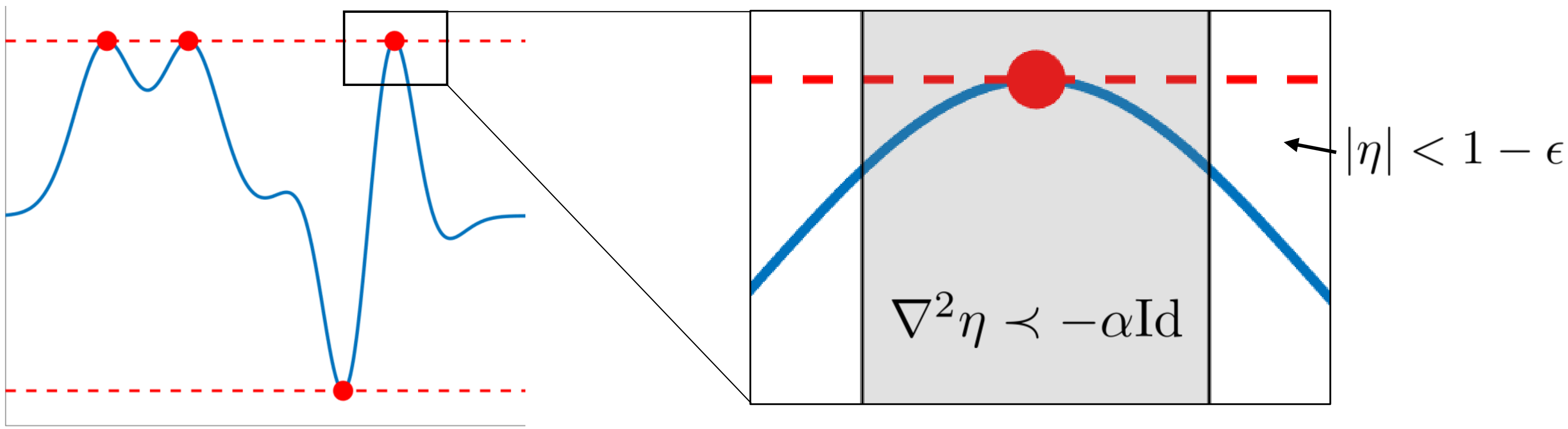
Theorem *(refinement of [Azaïs et al. 2015])*

Hyp: there exists a ND dual certif.

Result: *(wrt Bregman divergence)*

- $\tilde{\mu}$ **not necessarily sparse**, but
- Mass of $\tilde{\mu}$ **concentrated** around x_i

Recovery results



Theorem *(refinement of [Azaïs et al. 2015])*

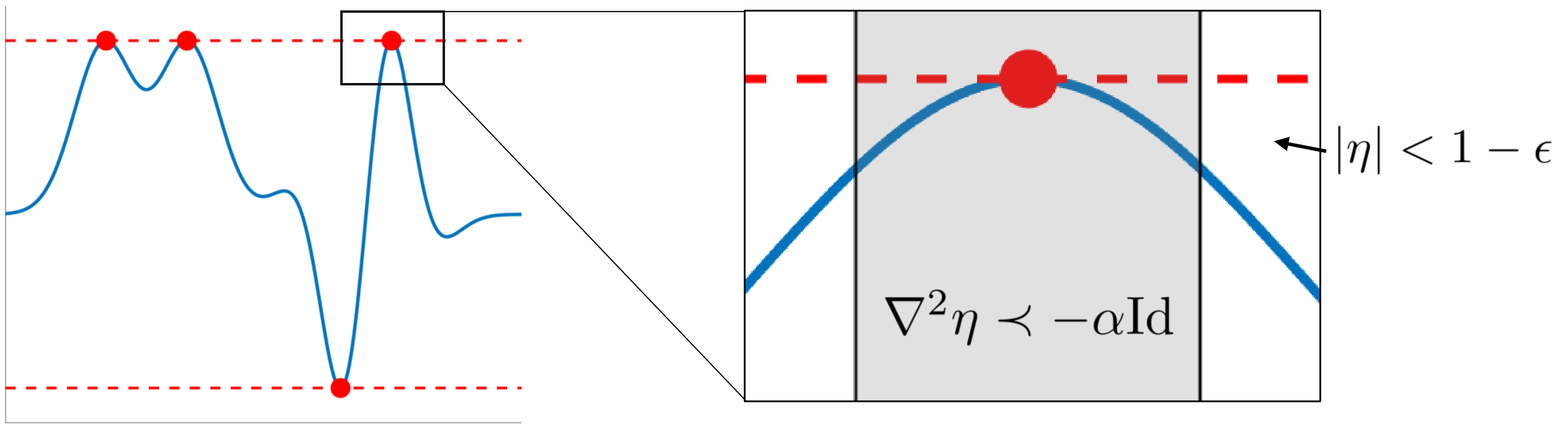
Hyp: there exists a ND dual certif.

Result: *(wrt Bregman divergence)*

- $\tilde{\mu}$ **not necessarily sparse**, but
- Mass of $\tilde{\mu}$ **concentrated** around x_i
- Concentration increases when:

$\alpha, \epsilon \nearrow$ noise $||e||, \lambda \searrow$

Recovery results



Theorem *(refinement of [Azaïs et al. 2015])*

Hyp: there exists a ND dual certif.

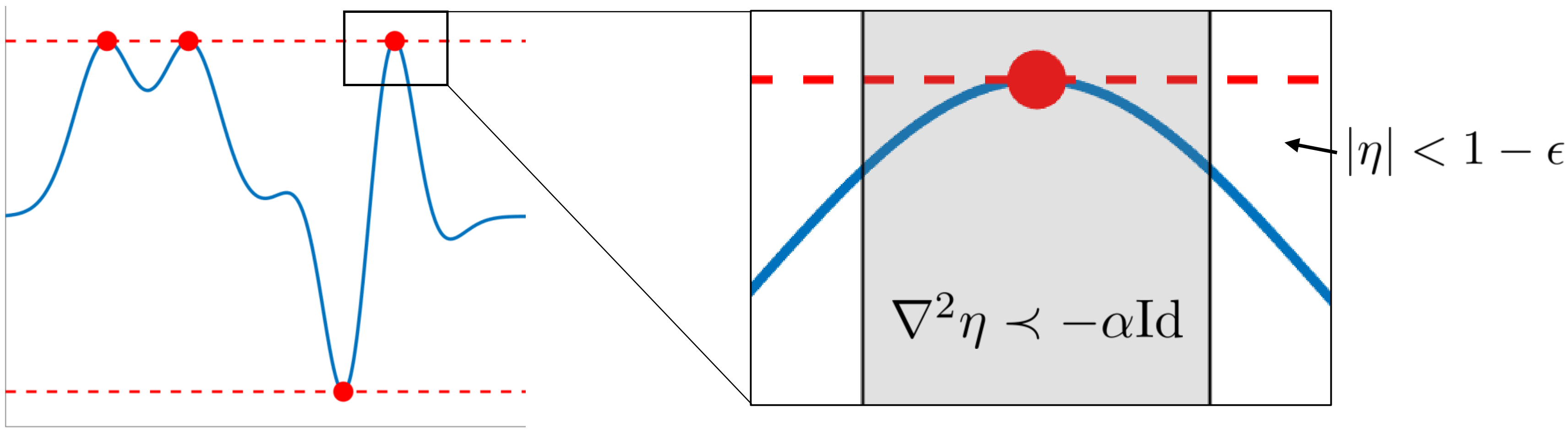
Result: *(wrt Bregman divergence)*

- $\tilde{\mu}$ **not necessarily sparse**, but
- Mass of $\tilde{\mu}$ **concentrated** around x_i
- Concentration increases when:

$\alpha, \epsilon \nearrow$ noise $||e||, \lambda \searrow$

Theorem *([Duval Peyré 2015])*

Recovery results



Theorem *(refinement of [Azaïs et al. 2015])*

Hyp: there exists a ND dual certif.

Result: *(wrt Bregman divergence)*

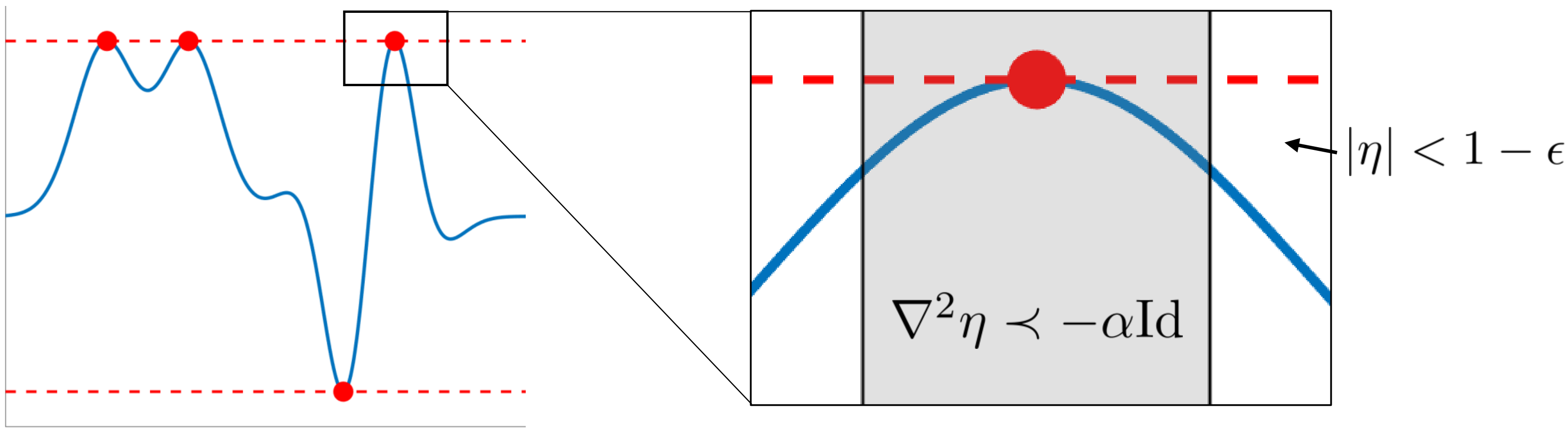
- $\tilde{\mu}$ **not necessarily sparse**, but
- Mass of $\tilde{\mu}$ **concentrated** around x_i
- Concentration increases when:

$\alpha, \epsilon \nearrow$ noise $\|e\|, \lambda \searrow$

Theorem *([Duval Peyré 2015])*

Hyp: the ***minimal norm certificate*** is non-degenerate

Recovery results



Theorem *(refinement of [Azaïs et al. 2015])*

Hyp: there exists a ND dual certif.

Result: *(wrt Bregman divergence)*

- $\tilde{\mu}$ **not necessarily sparse**, but
- Mass of $\tilde{\mu}$ **concentrated** around x_i
- Concentration increases when:

$\alpha, \epsilon \nearrow$ noise $\|e\|, \lambda \searrow$

Theorem *([Duval Peyré 2015])*

Hyp: the ***minimal norm certificate*** is non-degenerate

Result: *(in the small noise regime)*

- $\tilde{\mu}$ **is sparse, with the right number of components**
- $(\tilde{a}_i, \tilde{x}_i) \xrightarrow[\|e\| \to 0]{} (a_i, x_i)$

Outline

① Background on dual certificates

② Compressive off-the-grid recovery

③ Conclusion, outlooks

Goal

Goal: random sampling

$$\Phi\mu = \int \varphi(x) d\mu(x)$$

Low-dim, random

Goal

Goal: random sampling

$$\Phi\mu = \int \varphi(x) d\mu(x)$$

Low-dim, random

Strategy: Start with « high »-dimensional problem

$$\Phi_{\text{full}}\mu = \int \varphi_{\text{full}}(x) d\mu(x)$$

Ex: full Fourier (high-dim), full convolution (infinite-dim)

Goal

Goal: random sampling

$$\Phi\mu = \int \varphi(x) d\mu(x)$$

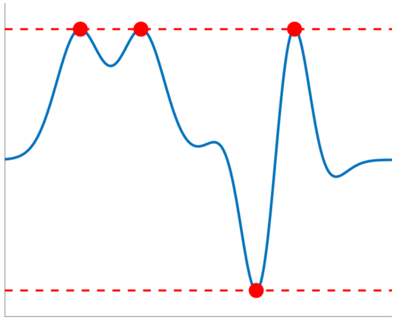
Low-dim, random

Strategy: Start with « high »-dimensional problem

$$\Phi_{\text{full}}\mu = \int \varphi_{\text{full}}(x) d\mu(x)$$

Ex: full Fourier (high-dim), full convolution (infinite-dim)

Step 1: Build ND certificate with ***full kernel***



Goal

Goal: random sampling

$$\Phi\mu = \int \varphi(x) d\mu(x)$$

Low-dim, random

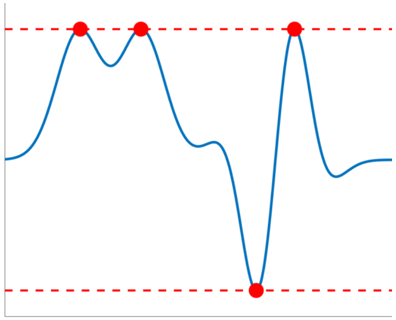
Strategy: Start with « high »-dimensional problem

$$\Phi_{\text{full}}\mu = \int \varphi_{\text{full}}(x) d\mu(x)$$

Ex: full Fourier (high-dim), full convolution (infinite-dim)

Step 1: Build ND certificate with ***full kernel***

$$\kappa_{\text{full}}(x, x') = \langle \varphi_{\text{full}}(x), \varphi_{\text{full}}(x') \rangle_{\mathcal{H}}$$



Goal

Goal: random sampling

$$\Phi\mu = \int \varphi(x) d\mu(x)$$

Low-dim, random

Strategy: Start with « high »-dimensional problem

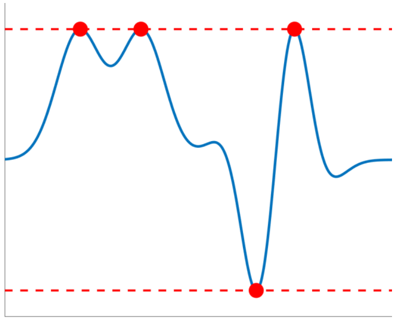
$$\Phi_{\text{full}}\mu = \int \varphi_{\text{full}}(x) d\mu(x)$$

Ex: full Fourier (high-dim), full convolution (infinite-dim)

Step 1: Build ND certificate with ***full kernel***

$$\kappa_{\text{full}}(x, x') = \langle \varphi_{\text{full}}(x), \varphi_{\text{full}}(x') \rangle_{\mathcal{H}}$$

$$\eta_{\text{full}} \in \text{Span}\left\{\kappa_{\text{full}}(x_i, \cdot), \partial_1 \kappa_{\text{full}}(x_i, \cdot)\right\} \subset \text{Im}(\Phi^\star_{\text{full}})$$



Goal

Goal: random sampling

$$\Phi\mu = \int \varphi(x)d\mu(x)$$

Low-dim, random

Strategy: Start with « high »-dimensional problem

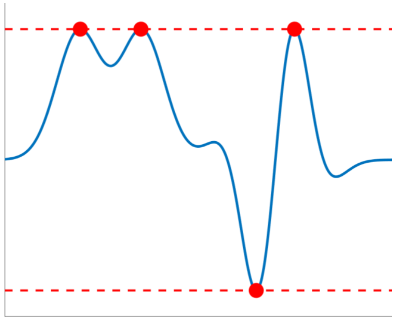
$$\Phi_{\text{full}}\mu = \int \varphi_{\text{full}}(x)d\mu(x)$$

Ex: full Fourier (high-dim), full convolution (infinite-dim)

Step 1: Build ND certificate with ***full kernel***

$$\kappa_{\text{full}}(x,x') = \langle \varphi_{\text{full}}(x), \varphi_{\text{full}}(x') \rangle_{\mathcal{H}}$$

$$\eta_{\text{full}} \in \operatorname{Span}\{\kappa_{\text{full}}(x_i,\cdot), \partial_1\kappa_{\text{full}}(x_i,\cdot)\} \subset \operatorname{Im}(\Phi_{\text{full}}^\star)$$



Step 2: Use ***Random Features*** on full kernel to define sampling

Goal

Goal: random sampling

$$\Phi\mu = \int \varphi(x) d\mu(x)$$

Low-dim, random

Strategy: Start with « high »-dimensional problem

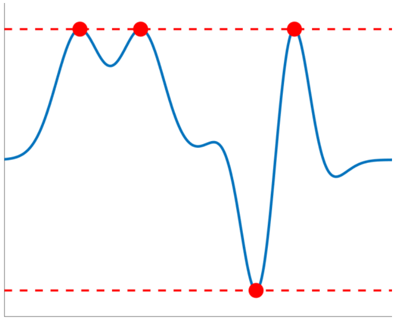
$$\Phi_{\text{full}}\mu = \int \varphi_{\text{full}}(x) d\mu(x)$$

Ex: full Fourier (high-dim), full convolution (infinite-dim)

Step 1: Build ND certificate with ***full kernel***

$$\kappa_{\text{full}}(x, x') = \langle \varphi_{\text{full}}(x), \varphi_{\text{full}}(x') \rangle_{\mathcal{H}}$$

$$\eta_{\text{full}} \in \text{Span}\left\{\kappa_{\text{full}}(x_i, \cdot), \partial_1 \kappa_{\text{full}}(x_i, \cdot)\right\} \subset \text{Im}(\Phi^\star_{\text{full}})$$



Step 2: Use ***Random Features*** on full kernel to define sampling

Assuming $\kappa_{\text{full}}(x, x') = \mathbb{E}_{\omega\sim\Lambda}\phi_\omega(x)\overline{\phi_\omega(x')}$

Define $\varphi(x) = \frac{1}{\sqrt{m}}[\phi_{\omega_j}(x)]_{j=1}^m$

Goal

Goal: random sampling

$$\Phi\mu = \int \varphi(x) d\mu(x)$$

Low-dim, random

Strategy: Start with « high »-dimensional problem

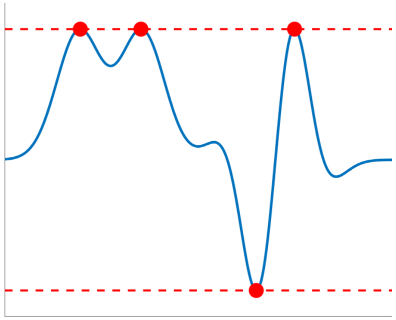
$$\Phi_{\text{full}}\mu = \int \varphi_{\text{full}}(x) d\mu(x)$$

Ex: full Fourier (high-dim), full convolution (infinite-dim)

Step 1: Build ND certificate with ***full kernel***

$$\kappa_{\text{full}}(x, x') = \langle \varphi_{\text{full}}(x), \varphi_{\text{full}}(x') \rangle_{\mathcal{H}}$$

$$\eta_{\text{full}} \in \text{Span}\left\{\kappa_{\text{full}}(x_i, \cdot), \partial_1 \kappa_{\text{full}}(x_i, \cdot)\right\} \subset \text{Im}(\Phi_{\text{full}}^\star)$$



Step 2: Use ***Random Features*** on full kernel to define sampling

Assuming $\kappa_{\text{full}}(x, x') = \mathbb{E}_{\omega \sim \Lambda} \phi_\omega(x) \overline{\phi_\omega(x')}$

Define $\varphi(x) = \frac{1}{\sqrt{m}} [\phi_{\omega_j}(x)]_{j=1}^m$

Then $\kappa(x, x') = \langle \varphi(x), \varphi(x') \rangle_{\mathbb{C}^m}$

Goal

Goal: random sampling

$$\Phi\mu = \int \varphi(x) d\mu(x)$$

Low-dim, random

Strategy: Start with « high »-dimensional problem

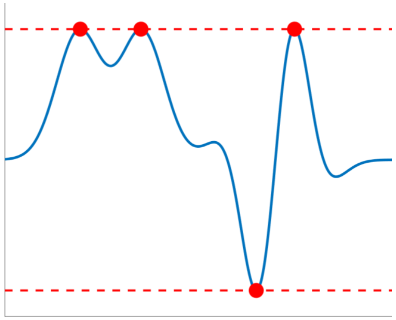
$$\Phi_{\text{full}}\mu = \int \varphi_{\text{full}}(x) d\mu(x)$$

Ex: full Fourier (high-dim), full convolution (infinite-dim)

Step 1: Build ND certificate with ***full kernel***

$$\kappa_{\text{full}}(x, x') = \langle \varphi_{\text{full}}(x), \varphi_{\text{full}}(x') \rangle_{\mathcal{H}}$$

$$\eta_{\text{full}} \in \text{Span}\left\{\kappa_{\text{full}}(x_i, \cdot), \partial_1 \kappa_{\text{full}}(x_i, \cdot)\right\} \subset \text{Im}(\Phi^\star_{\text{full}})$$



Step 2: Use ***Random Features*** on full kernel to define sampling

Assuming $\kappa_{\text{full}}(x, x') = \mathbb{E}_{\omega \sim \Lambda} \phi_\omega(x) \overline{\phi_\omega(x')}$

Define $\varphi(x) = \frac{1}{\sqrt{m}} [\phi_{\omega_j}(x)]_{j=1}^m$

Then $\kappa(x, x') = \langle \varphi(x), \varphi(x') \rangle_{\mathbb{C}^m}$

$\approx \kappa_{\text{full}}(x, x')$

Goal

Goal: random sampling

$$\Phi\mu = \int \varphi(x) d\mu(x)$$

Low-dim, random

Strategy: Start with « high »-dimensional problem

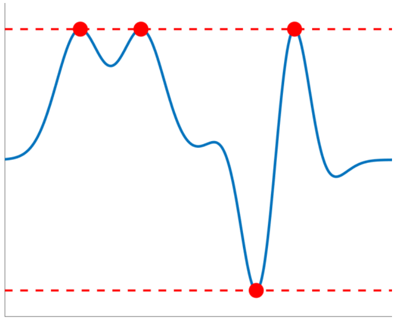
$$\Phi_{\text{full}}\mu = \int \varphi_{\text{full}}(x) d\mu(x)$$

Ex: full Fourier (high-dim), full convolution (infinite-dim)

Step 1: Build ND certificate with ***full kernel***

$$\kappa_{\text{full}}(x, x') = \langle \varphi_{\text{full}}(x), \varphi_{\text{full}}(x') \rangle_{\mathcal{H}}$$

$$\eta_{\text{full}} \in \text{Span}\left\{\kappa_{\text{full}}(x_i, \cdot), \partial_1 \kappa_{\text{full}}(x_i, \cdot)\right\} \subset \text{Im}(\Phi^\star_{\text{full}})$$



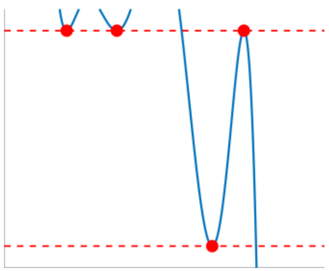
Step 2: Use ***Random Features*** on full kernel to define sampling

Assuming $\kappa_{\text{full}}(x, x') = \mathbb{E}_{\omega \sim \Lambda} \phi_\omega(x) \overline{\phi_\omega(x')}$

Define $\varphi(x) = \frac{1}{\sqrt{m}} [\phi_{\omega_j}(x)]_{j=1}^m$

Then $\kappa(x, x') = \langle \varphi(x), \varphi(x') \rangle_{\mathbb{C}^m}$

$\approx \kappa_{\text{full}}(x, x')$



m=10

Goal

Goal: random sampling

$$\Phi\mu = \int \varphi(x) d\mu(x)$$

Low-dim, random

Strategy: Start with « high »-dimensional problem

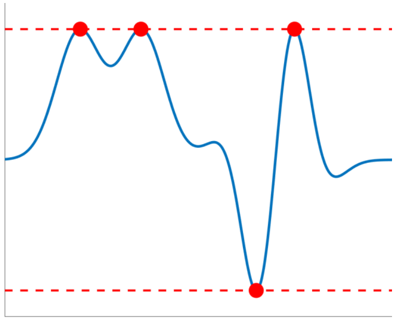
$$\Phi_{\text{full}}\mu = \int \varphi_{\text{full}}(x) d\mu(x)$$

Ex: full Fourier (high-dim), full convolution (infinite-dim)

Step 1: Build ND certificate with ***full kernel***

$$\kappa_{\text{full}}(x, x') = \langle \varphi_{\text{full}}(x), \varphi_{\text{full}}(x') \rangle_{\mathcal{H}}$$

$$\eta_{\text{full}} \in \text{Span}\left\{\kappa_{\text{full}}(x_i, \cdot), \partial_1 \kappa_{\text{full}}(x_i, \cdot)\right\} \subset \text{Im}(\Phi_{\text{full}}^\star)$$



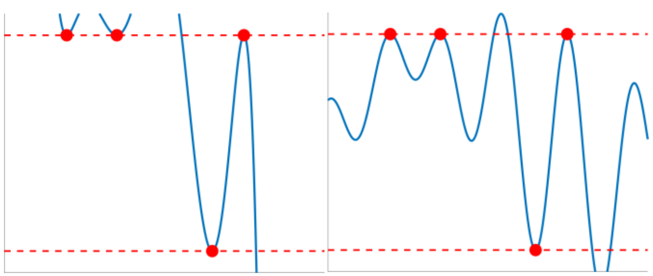
Step 2: Use ***Random Features*** on full kernel to define sampling

Assuming $\kappa_{\text{full}}(x, x') = \mathbb{E}_{\omega\sim\Lambda}\phi_\omega(x)\overline{\phi_\omega(x')}$

Define $\varphi(x) = \frac{1}{\sqrt{m}}[\phi_{\omega_j}(x)]_{j=1}^m$

Then $\kappa(x, x') = \langle \varphi(x), \varphi(x') \rangle_{\mathbb{C}^m}$

$\approx \kappa_{\text{full}}(x, x')$



m=10 *m=50*

Goal

Goal: random sampling

$$\Phi\mu = \int \varphi(x) d\mu(x)$$

Low-dim, random

Strategy: Start with « high »-dimensional problem

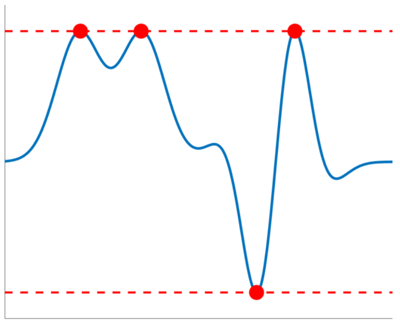
$$\Phi_{\text{full}}\mu = \int \varphi_{\text{full}}(x) d\mu(x)$$

Ex: full Fourier (high-dim), full convolution (infinite-dim)

Step 1: Build ND certificate with ***full kernel***

$$\kappa_{\text{full}}(x, x') = \langle \varphi_{\text{full}}(x), \varphi_{\text{full}}(x') \rangle_{\mathcal{H}}$$

$$\eta_{\text{full}} \in \text{Span}\left\{\kappa_{\text{full}}(x_i, \cdot), \partial_1 \kappa_{\text{full}}(x_i, \cdot)\right\} \subset \text{Im}(\Phi^\star_{\text{full}})$$



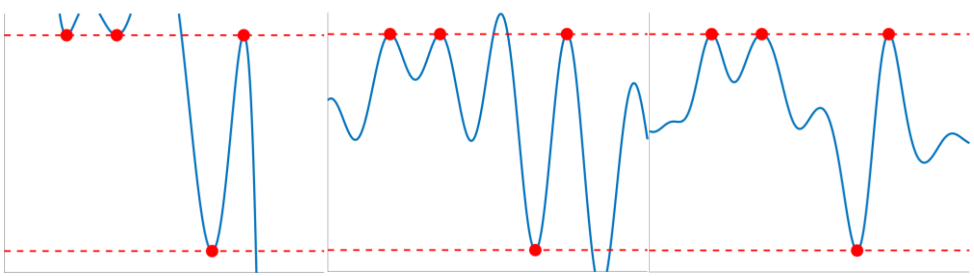
Step 2: Use ***Random Features*** on full kernel to define sampling

Assuming $\kappa_{\text{full}}(x, x') = \mathbb{E}_{\omega \sim \Lambda} \phi_\omega(x) \overline{\phi_\omega(x')}$

Define $\varphi(x) = \frac{1}{\sqrt{m}} [\phi_{\omega_j}(x)]_{j=1}^m$

Then $\kappa(x, x') = \langle \varphi(x), \varphi(x') \rangle_{\mathbb{C}^m}$

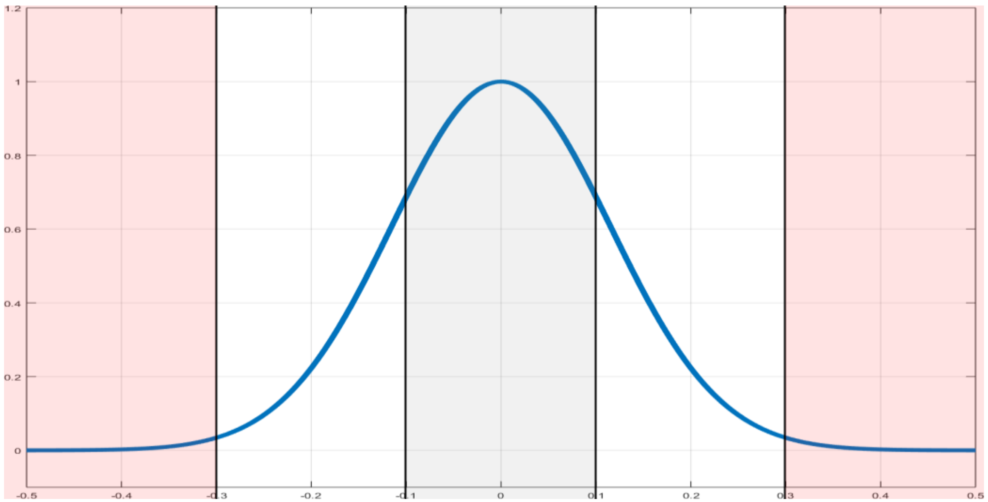
$\approx \kappa_{\text{full}}(x, x')$



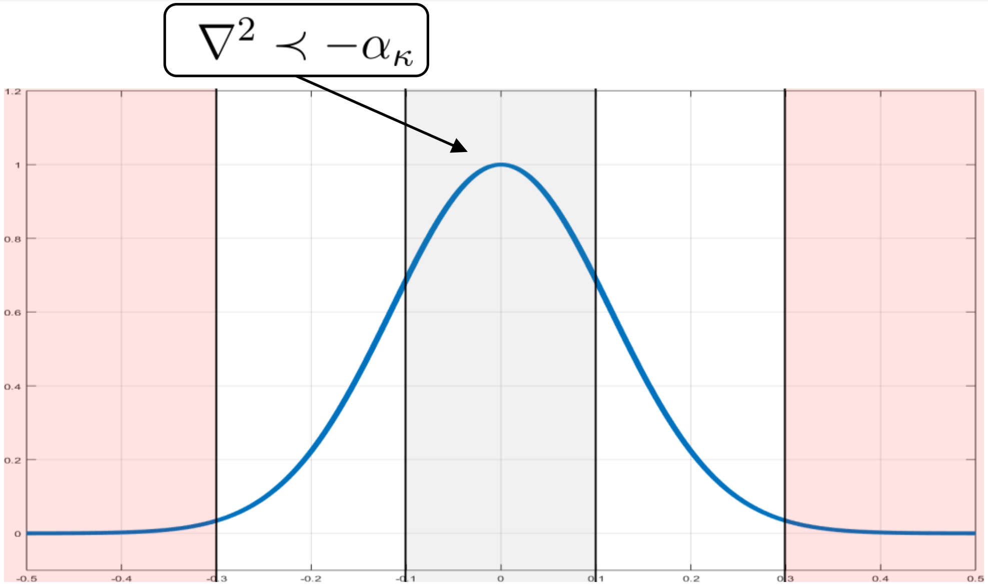
m=10 *m=50* *m=500*

Step 1: Acceptable full kernels

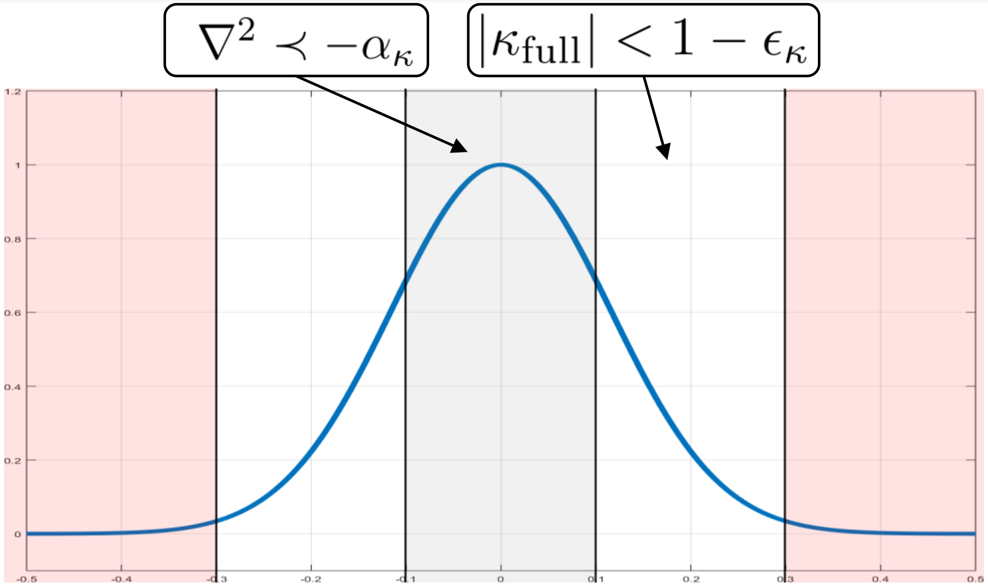
Step 1: Acceptable full kernels



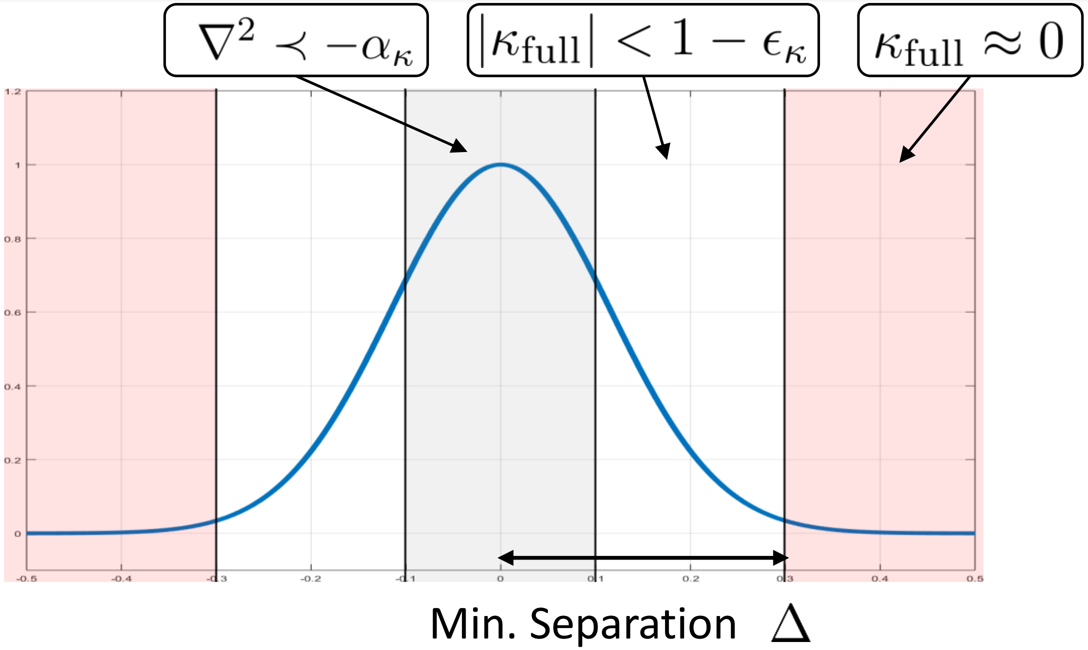
Step 1: Acceptable full kernels



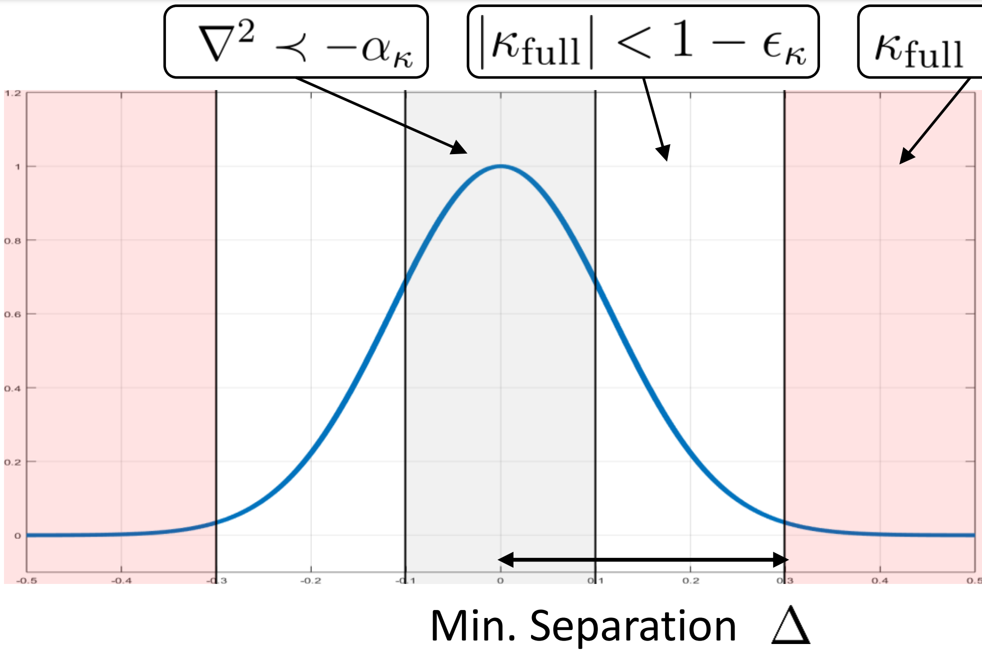
Step 1: Acceptable full kernels

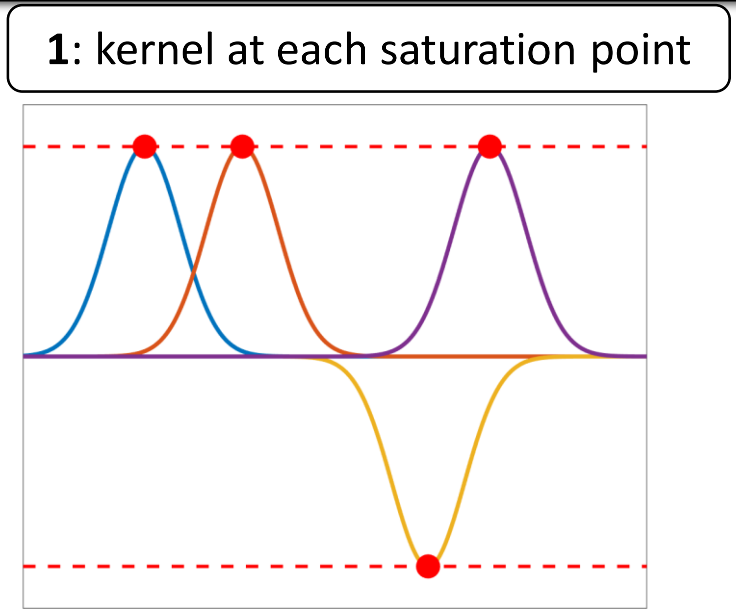


Step 1: Acceptable full kernels



Step 1: Acceptable full kernels



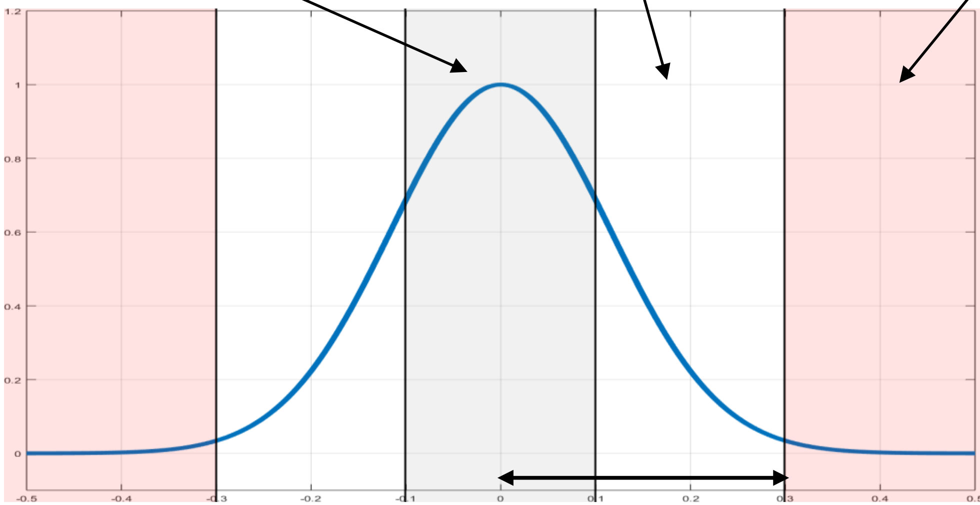


Step 1: Acceptable full kernels

$\nabla^2 \prec -\alpha_\kappa$

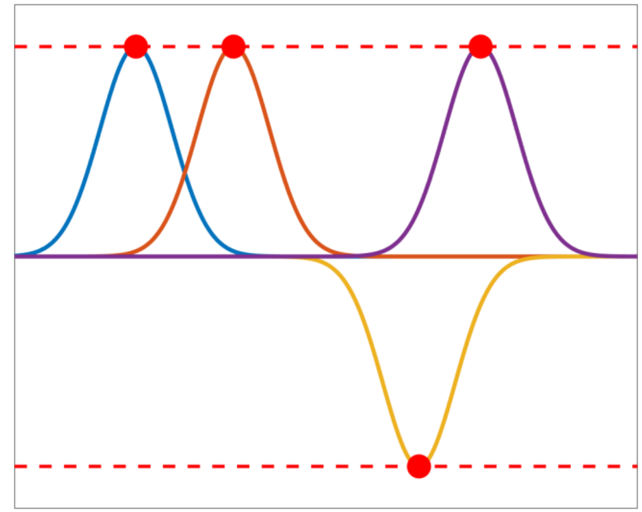
$|\kappa_{\text{full}}| < 1 - \epsilon_\kappa$

$\kappa_{\text{full}} \approx 0$

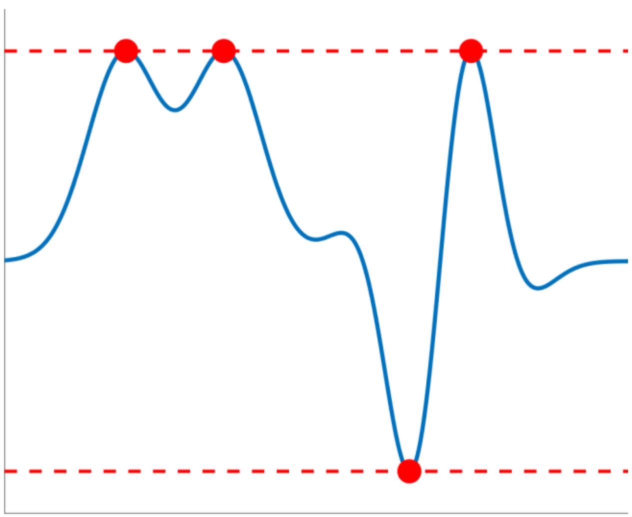


Min. Separation Δ

1: kernel at each saturation point

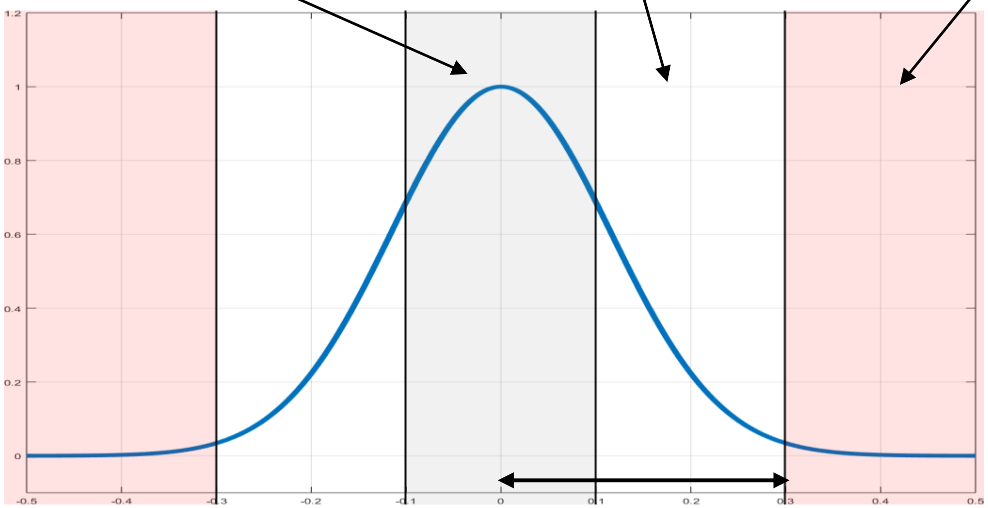


2: ***Small*** adjustements: minimal separation



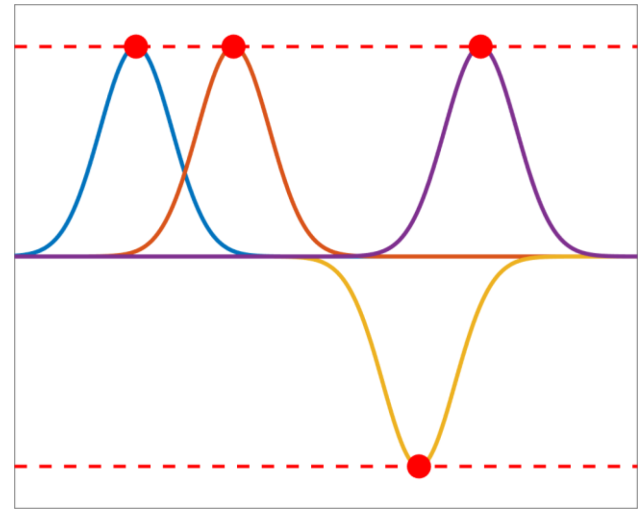
Step 1: Acceptable full kernels

$\nabla^2 \prec -\alpha_\kappa$ | $|\kappa_{\text{full}}| < 1 - \epsilon_\kappa$ | $\kappa_{\text{full}} \approx 0$

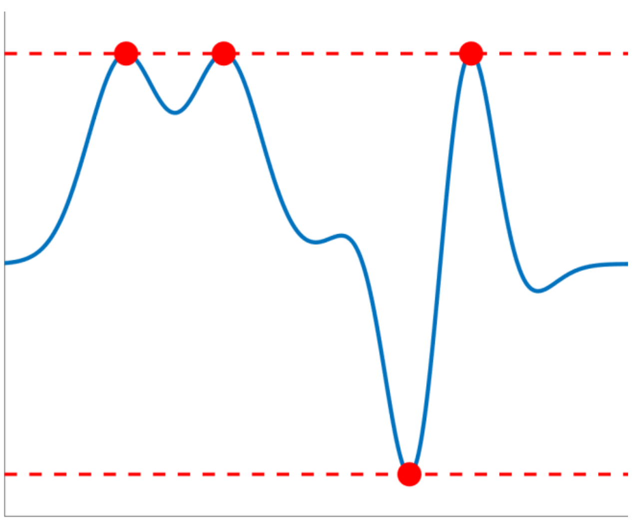


Min. Separation Δ

1: kernel at each saturation point



2: ***Small*** adjustements: minimal separation



- ***Multi-d square Féjer*** kernel (regular Fourier on Torus)

$$\Delta = \mathcal{O}(s^{\frac{1}{4}} d^{\frac{1}{2}} f_c^{-1})$$

- ***Multi-d Gaussian*** kernel

$$\Delta = \mathcal{O}(\sigma^{-1}\sqrt{\log(s) + \log(d)})$$

Step 2: sampling

Thm: Ideal scaling in sparsity: infinite-dim. golfing scheme

Step 2: sampling

Thm: Ideal scaling in sparsity: infinite-dim. golfing scheme

$$m \geq \mathcal{O}(s d^r \cdot \texttt{polylog}(s, d))$$

Depend on kernel

Step 2: sampling

Thm: Ideal scaling in sparsity: infinite-dim. golfing scheme

$$m \geq \mathcal{O}(sd^r \cdot \texttt{polylog}(s,d))$$

↖ *Depend on kernel*

- There exists a ND dual certificate (however ***not*** the minimal norm certificate)

Step 2: sampling

Thm: Ideal scaling in sparsity: infinite-dim. golfing scheme

$$m \geq \mathcal{O}(sd^r \cdot \texttt{polylog}(s,d))$$

↖ *Depend on kernel*

- There exists a ND dual certificate (however ***not*** the minimal norm certificate)
- No need for random signs

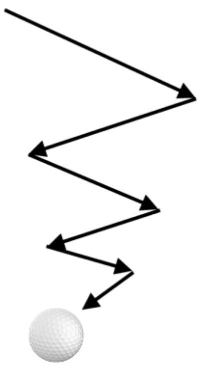
Step 2: sampling

Thm: Ideal scaling in sparsity: infinite-dim. golfing scheme

$$m \geq \mathcal{O}(sd^r \cdot \texttt{polylog}(s,d))$$

Depend on kernel

- There exists a ND dual certificate (however ***not*** the minimal norm certificate)
- No need for random signs
- Proof: ***golfing scheme***
  [Gross 2009, Candès Plan 2011]



$$m = m_1 + m_2 + m_3 + \dots$$

Step 2: sampling

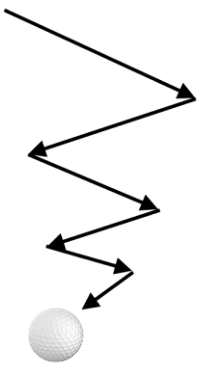
Thm: Ideal scaling in sparsity: infinite-dim. golfing scheme

$$m \geq \mathcal{O}(s d^r \cdot \mathtt{polylog}(s, d))$$

Depend on kernel

- There exists a ND dual certificate (however ***not*** the minimal norm certificate)
- No need for random signs
- Proof: ***golfing scheme***

  [Gross 2009, Candès Plan 2011]



$$m = m_1 + m_2 + m_3 + \dots$$

Thm: Minimal norm certificate (adaptation of *[Tang, Recht 2013]*)

$$m \geq \mathcal{O}(s d^r \cdot \mathtt{polylog}(s, d))$$

With random signs

$$m \geq \mathcal{O}(s^2 d^r \cdot \mathtt{polylog}(s, d))$$

Without random signs

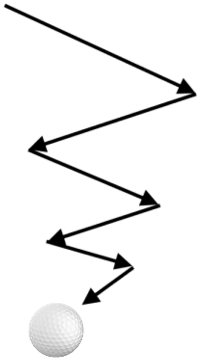
Step 2: sampling

Thm: Ideal scaling in sparsity: infinite-dim. golfing scheme

$$m \geq \mathcal{O}(s d^r \cdot \texttt{polylog}(s,d))$$

Depend on kernel

- There exists a ND dual certificate (however ***not*** the minimal norm certificate)
- No need for random signs
- Proof: ***golfing scheme***
  [Gross 2009, Candès Plan 2011]



$$m = m_1 + m_2 + m_3 + \ldots$$

Thm: Minimal norm certificate (adaptation of *[Tang, Recht 2013]*)

$$m \geq \mathcal{O}(s d^r \cdot \texttt{polylog}(s,d))$$

With random signs

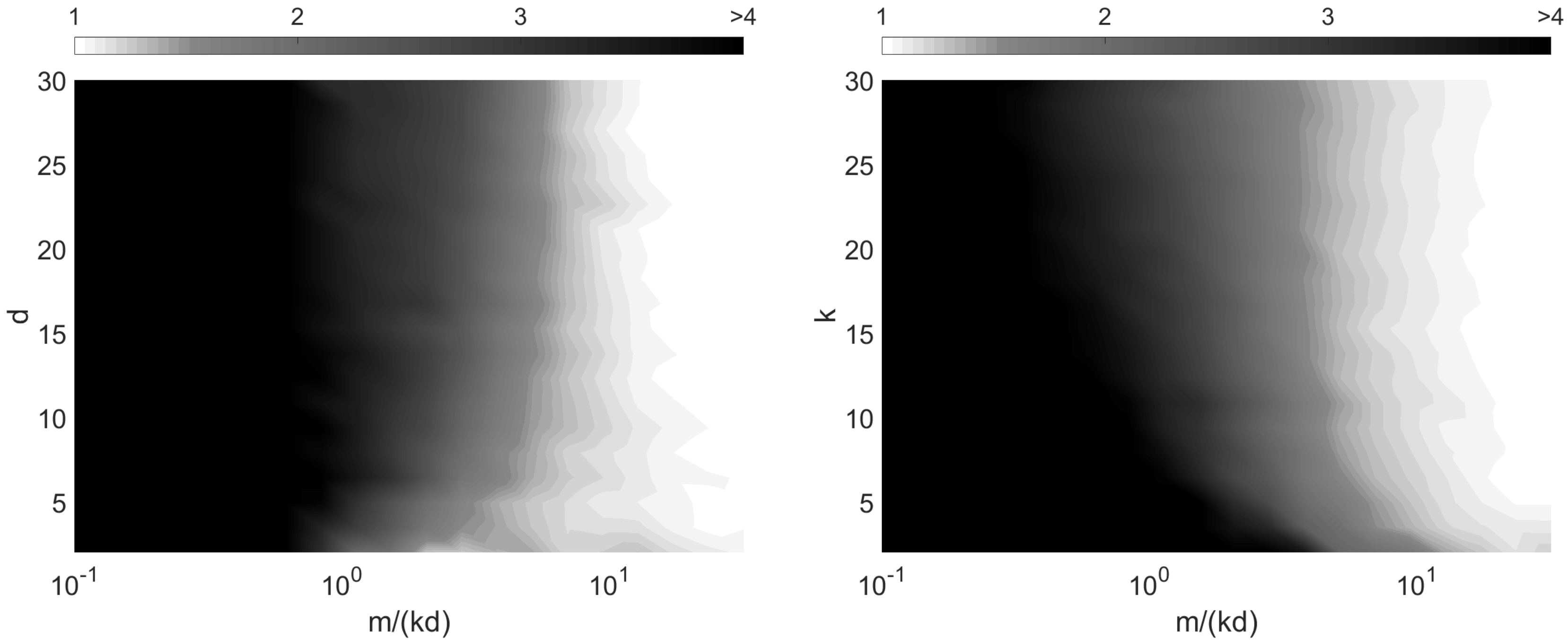
$$m \geq \mathcal{O}(s^2 d^r \cdot \texttt{polylog}(s,d))$$

Without random signs

Fun application: convex approach for automatic estimation of number of components in a GMM

Number of measurements in practice ?

Compressive k-means *[Keriven 2017]*



Relative number of measurements m/(sd)

Outline

① Background on dual certificates

② Compressive off-the-grid recovery

③ Conclusion, outlooks

Summary, outlooks

- **Summary**: generalization of existing results on ***super-resolution with random measurements*** (and minimal separation)
  - Beyond Fourier on the Torus (« acceptable » kernels)
  - Multi-d
  - No need for random signs for basic recovery result
  - Support recovery when random signs (or quadratic number of measurements)

Summary, outlooks

- **Summary**: generalization of existing results on ***super-resolution with random measurements*** (and minimal separation)
  - Beyond Fourier on the Torus (« acceptable » kernels)
  - Multi-d
  - No need for random signs for basic recovery result
  - Support recovery when random signs (or quadratic number of measurements)

- **Outlooks**
  - Other kernels, *very different from translation-invariant*
  - More quantified treatment of dimension
  - Other practical applications (eg 1-layer neural networks with continuum of neurons *[Bach 2017]*)

Thank you !

Poon, Keriven, Peyré. **A Dual Certificates Analysis of Compressive Off-the-Grid Recovery.** *Preprint arxiv:1802.08464*

Code: sketchml.gforge.inria.fr,
github: nkeriven

data-ens.github.io

Enter the data challenges!
Come to the colloquium!
Come to the Laplace seminars!