Fisher metric, support stability and optimal number of measurements in compressive off-the-grid recovery

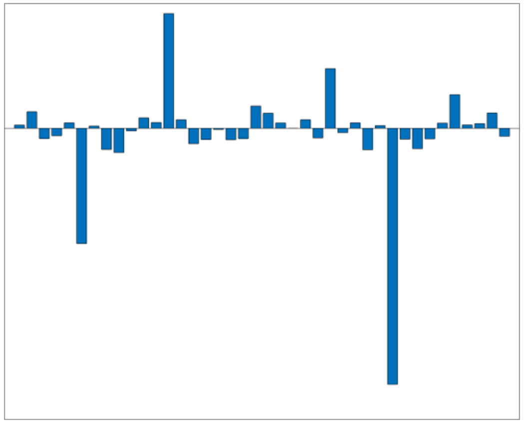
Nicolas Keriven

Ecole Normale Supérieure (Paris)

CFM-ENS chair in Data Science

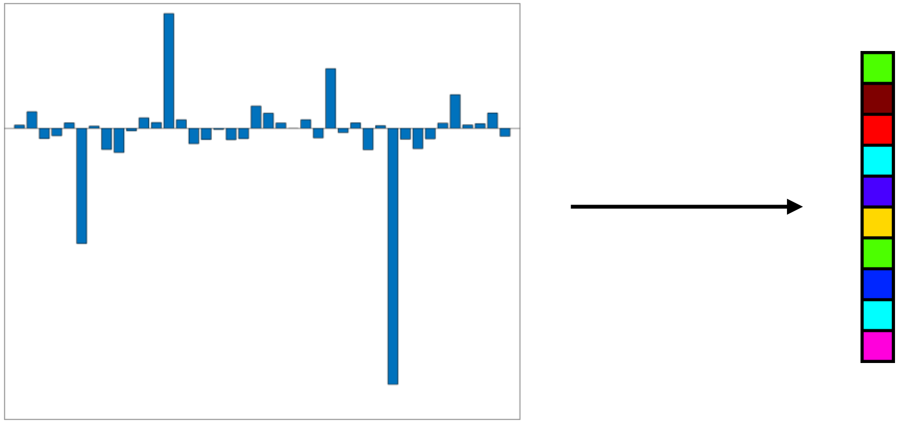
Joint work with **Clarice Poon** (Cambridge Uni.), **Gabriel Peyré** (ENS)

Discrete compressive sensing



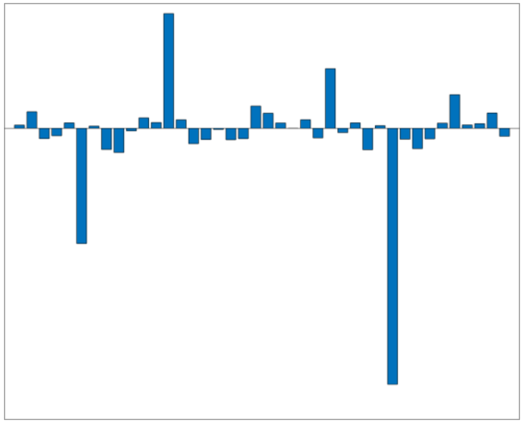
$$x \in \mathbb{R}^n$$

Discrete compressive sensing

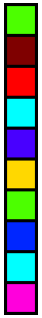


$$x \in \mathbb{R}^n \qquad\qquad y = Mx + w$$

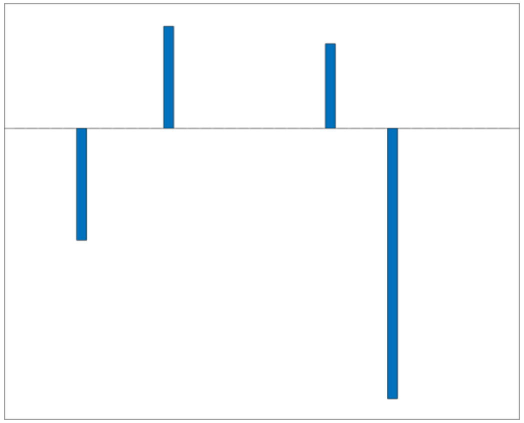
Discrete compressive sensing



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

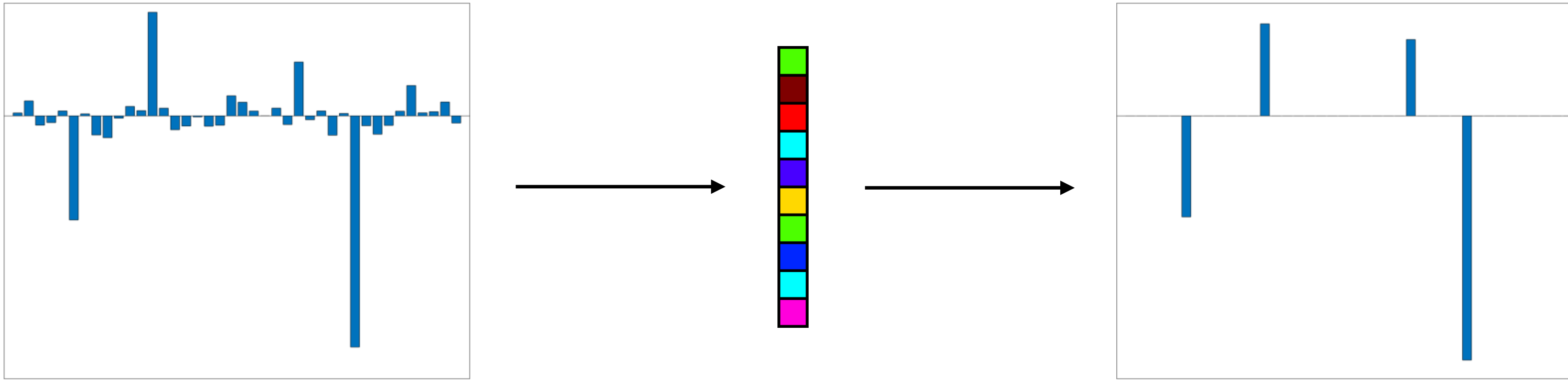


$y = Mx + w$



$\tilde{x}$

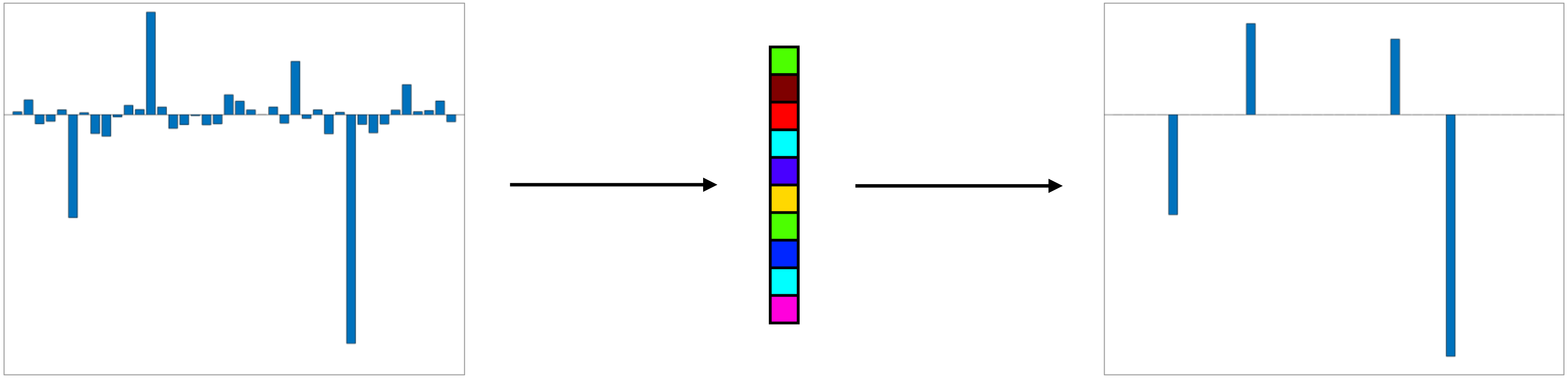
Discrete compressive sensing



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

- **Signal**: vector

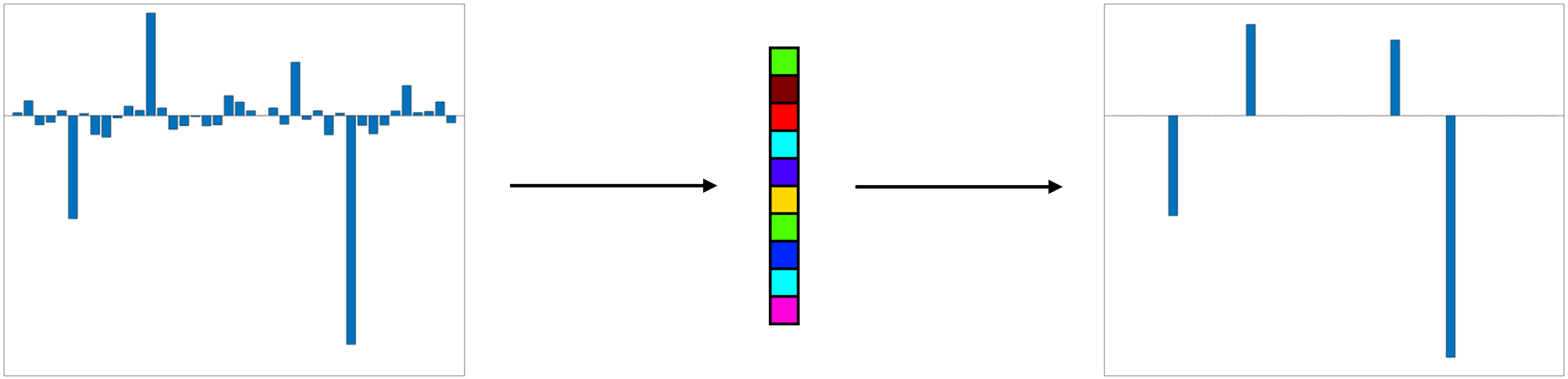
Discrete compressive sensing



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

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

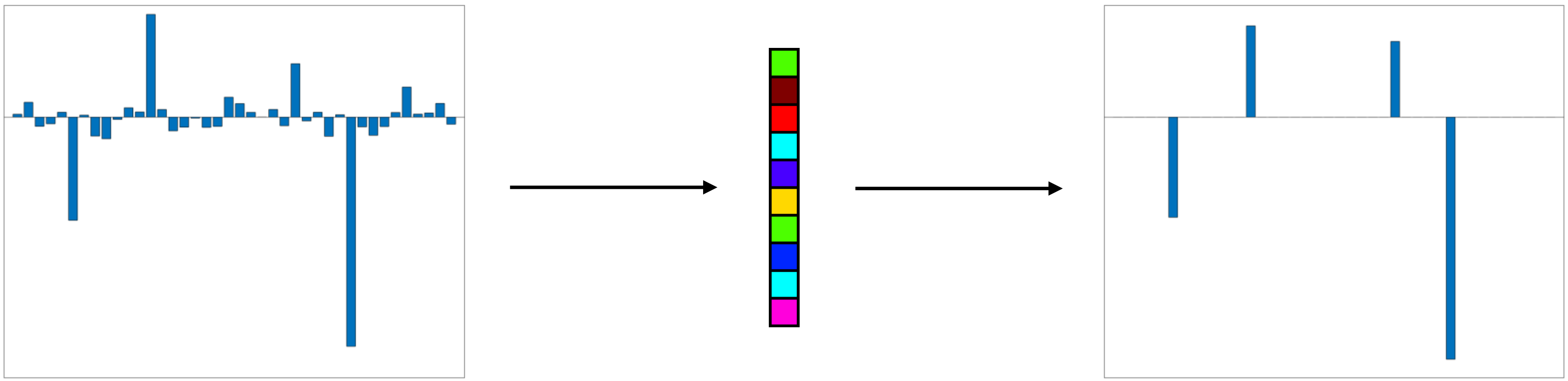
Discrete compressive sensing



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

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

Discrete compressive sensing



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

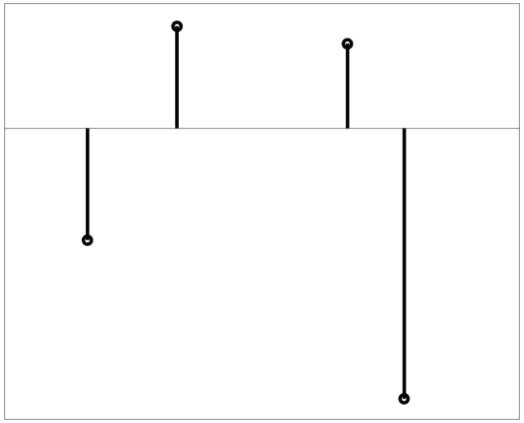
- **Signal**: vector
- **Sparsity**: few non-zeros coefficients
- **Dimensionality reduction** (often random matrix)
- **Recovery**: convex relaxation

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

LASSO

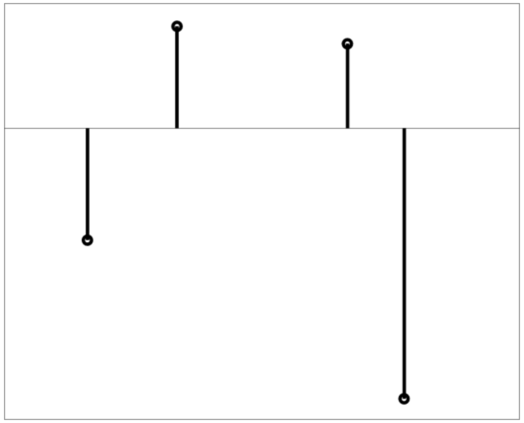
$$\min_x \frac{1}{2}\|Mx - y\|_2^2 + \lambda \|x\|_1$$

Continuous sparsity ?



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

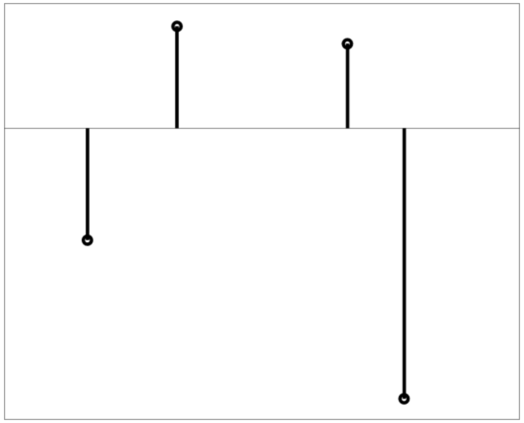
Continuous sparsity ?



$$y = \Phi\mu + w$$

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

Continuous sparsity ?



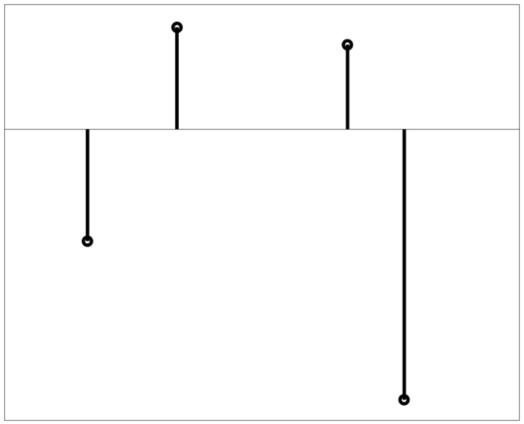
$\mu \in \mathcal{M}(\mathcal{X})$

$$y = \Phi\mu + w$$

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

- Fourier coeff
- Convolution

Continuous sparsity ?

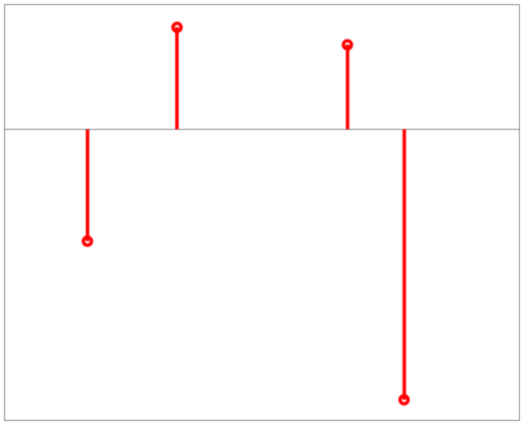


$\mu \in \mathcal{M}(\mathcal{X})$

$$y = \Phi\mu + w$$

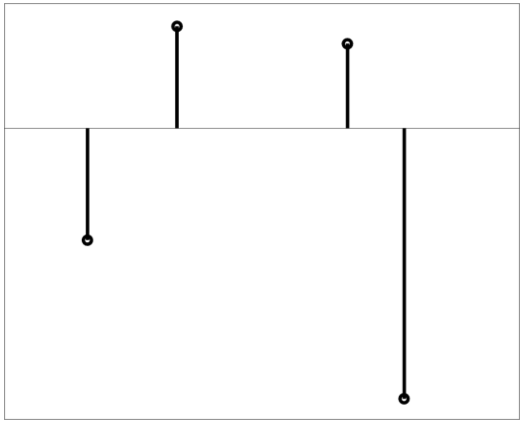
$$\Phi\mu = \int \varphi(x) d\mu(x)$$

- Fourier coeff
- Convolution



$\tilde{\mu}$

Continuous sparsity ?

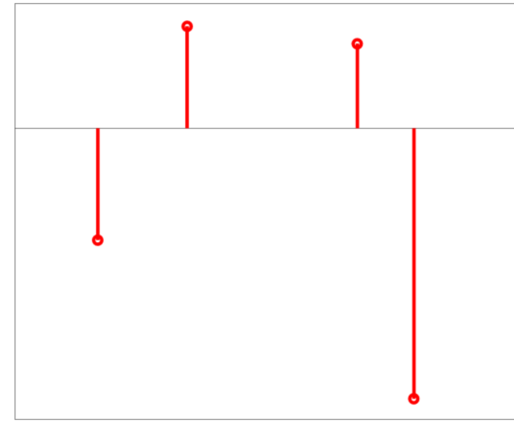


$\mu \in \mathcal{M}(\mathcal{X})$

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

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

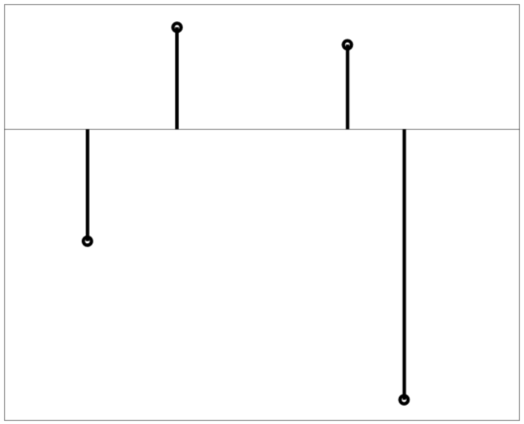
- Fourier coeff
- Convolution



$\tilde{\mu}$

- **Signal**: Radon measure

Continuous sparsity ?

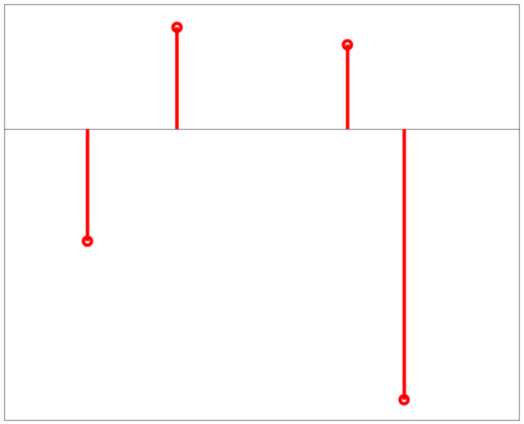


$\mu \in \mathcal{M}(\mathcal{X})$

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

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

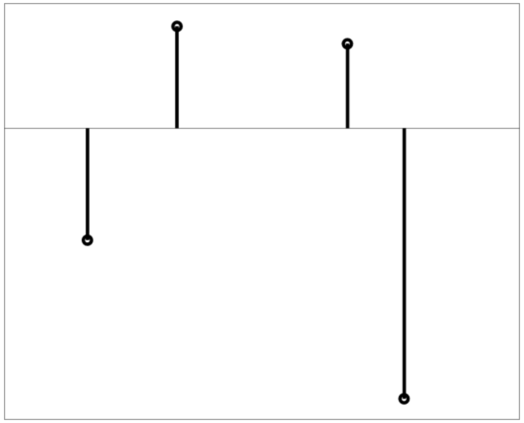
- Fourier coeff
- Convolution



$\tilde{\mu}$

- **Signal**: Radon measure
- **Sparsity**: $\mu_0 = \sum_i a_i \delta_{x_i}$

Continuous sparsity ?

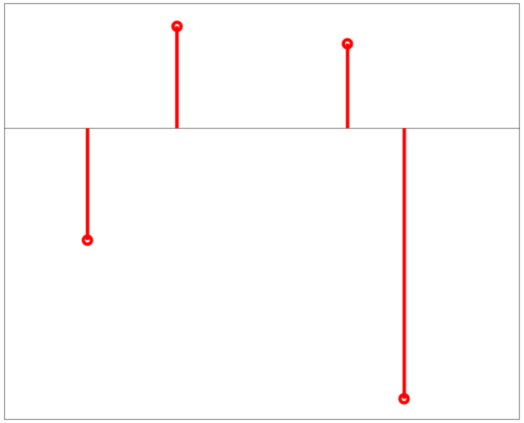


$\mu \in \mathcal{M}(\mathcal{X})$

$$y = \Phi\mu + w$$

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

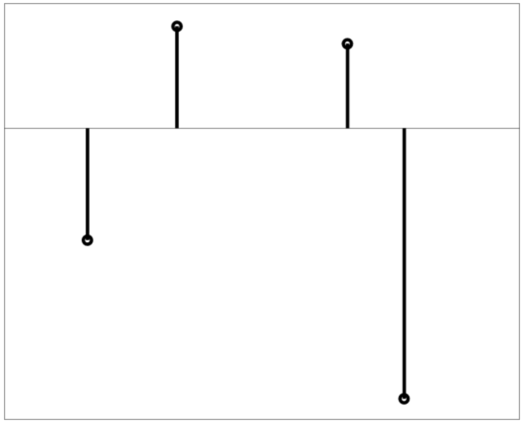
- Fourier coeff
- Convolution



$\tilde{\mu}$

- **Signal**: Radon measure
- **Sparsity**: $\mu_0 = \sum_i a_i \delta_{x_i}$
- **Dimensionality reduction** (e.g. first Fourier coefficients)

Continuous sparsity ?

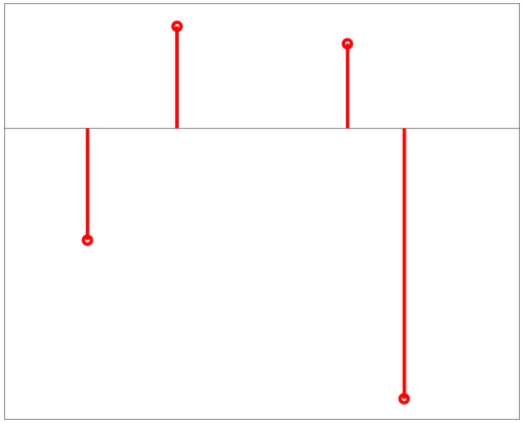


$\mu \in \mathcal{M}(\mathcal{X})$

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

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

- Fourier coeff
- Convolution



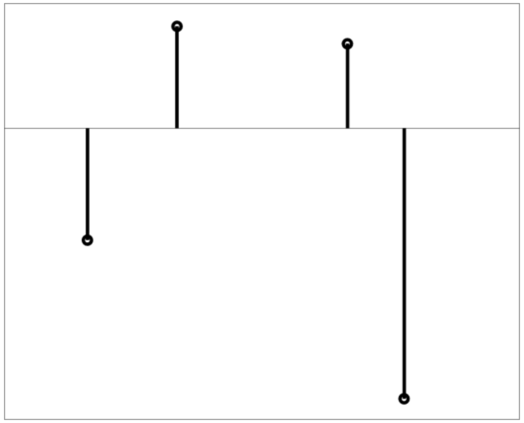
$\tilde{\mu}$

- **Signal**: Radon measure
- **Sparsity**: $\mu_0 = \sum_i a_i \delta_{x_i}$
- **Dimensionality reduction** (e.g. first Fourier coefficients)
- **Recovery**: convex relaxation?

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

See Keriven 2017, Gribonval 2017

Continuous sparsity ?

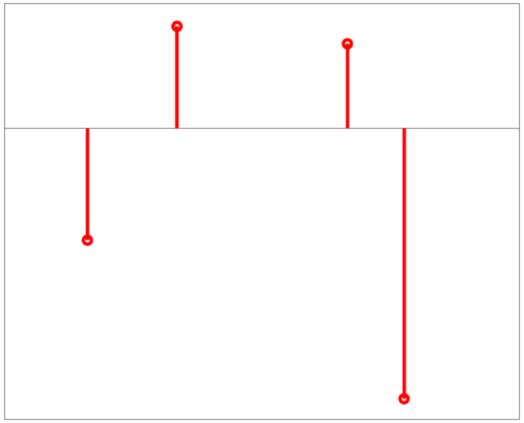


$\mu \in \mathcal{M}(\mathcal{X})$

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

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

- Fourier coeff
- Convolution



$\tilde{\mu}$

- **Signal**: Radon measure
- **Sparsity**: $\mu_0 = \sum_i a_i \delta_{x_i}$
- **Dimensionality reduction** (e.g. first Fourier coefficients)
- **Recovery**: convex relaxation?

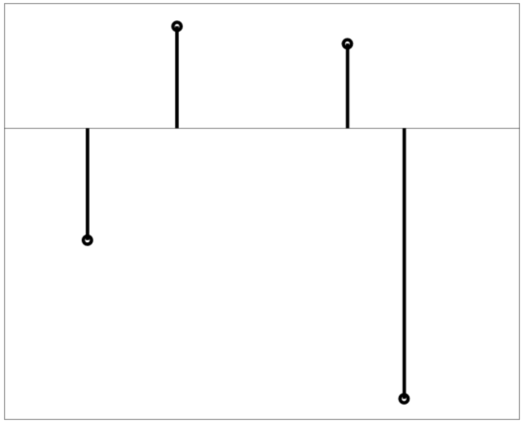
$$\min_{a,x} \|\Phi(\textstyle\sum_i a_i \delta_{x_i}) - y\| \longrightarrow$$

See Keriven 2017, Gribonval 2017

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

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

Continuous sparsity ?

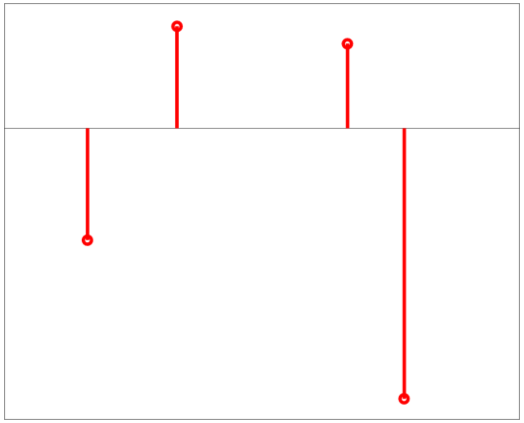


$\mu \in \mathcal{M}(\mathcal{X})$

$$y = \Phi\mu + w$$

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

- Fourier coeff
- Convolution



$\tilde{\mu}$

- **Signal**: Radon measure
- **Sparsity**: $\mu_0 = \sum_i a_i \delta_{x_i}$
- **Dimensionality reduction** (e.g. first Fourier coefficients)
- **Recovery**: convex relaxation?

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

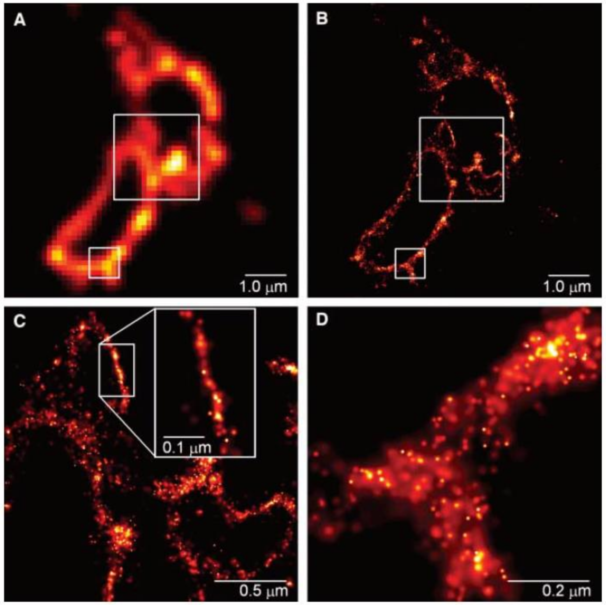
See Keriven 2017, Gribonval 2017

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

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

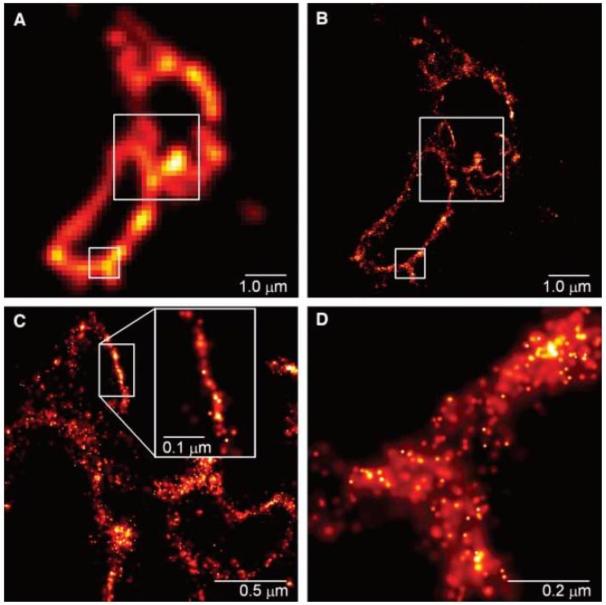
Other approaches: « Prony-like » ESPRIT, MUSIC... (but only 1d noiseless Fourier)

Example of applications



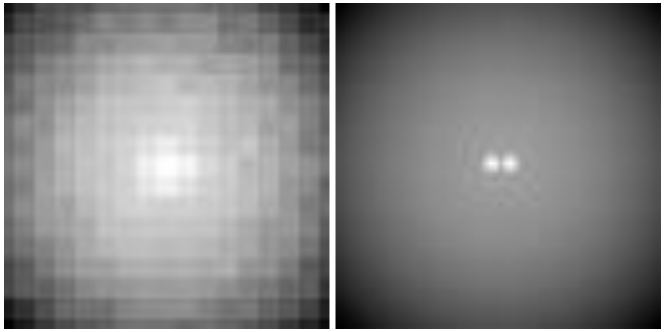
Fluorescence microscopy (3D)
PALM, STORM… *[Betzig 2006]*

Example of applications

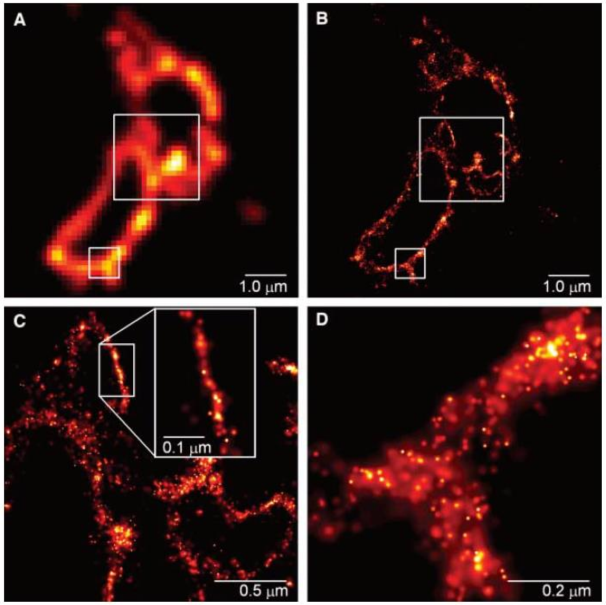


Fluorescence microscopy (3D)
PALM, STORM… *[Betzig 2006]*

Astronomy (2D)
[Puschmann 2017]

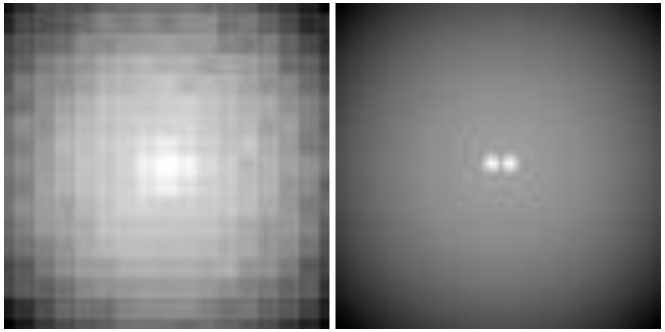


Example of applications

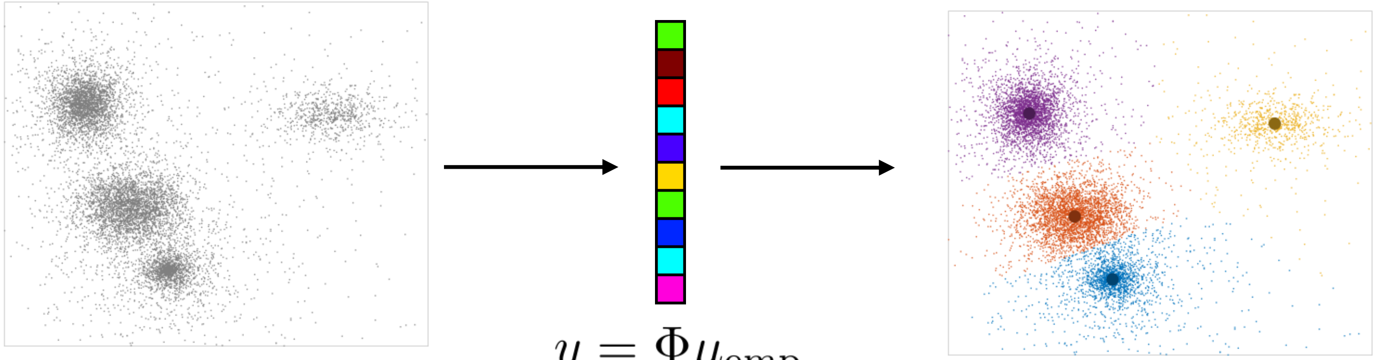


Fluorescence microscopy (3D)
PALM, STORM… *[Betzig 2006]*

Astronomy (2D)
[Puschmann 2017]

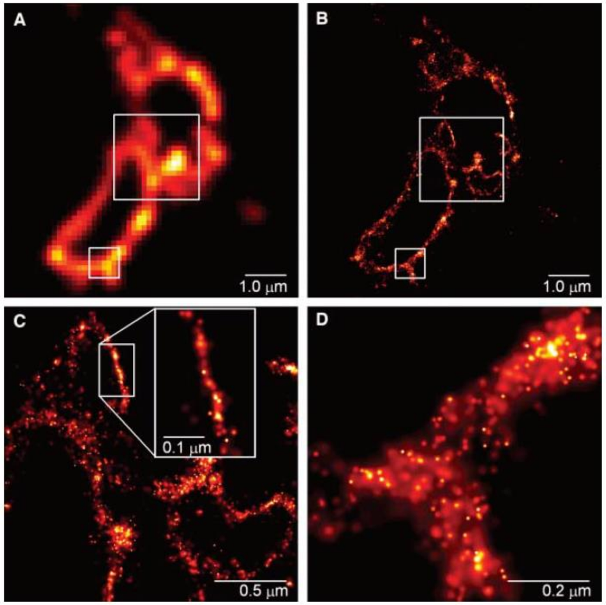


Compressive mixture model learning (many D)
[Keriven 2017]



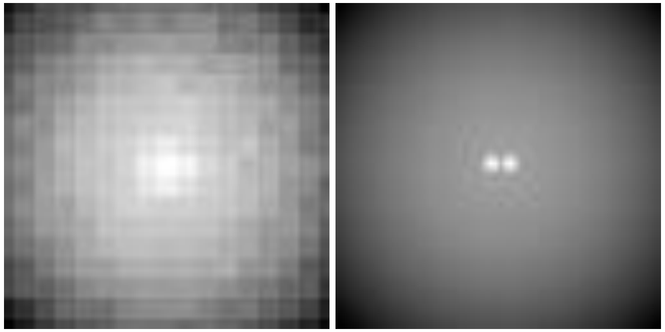
$$y = \Phi \mu_{\text{emp.}}$$

Example of applications



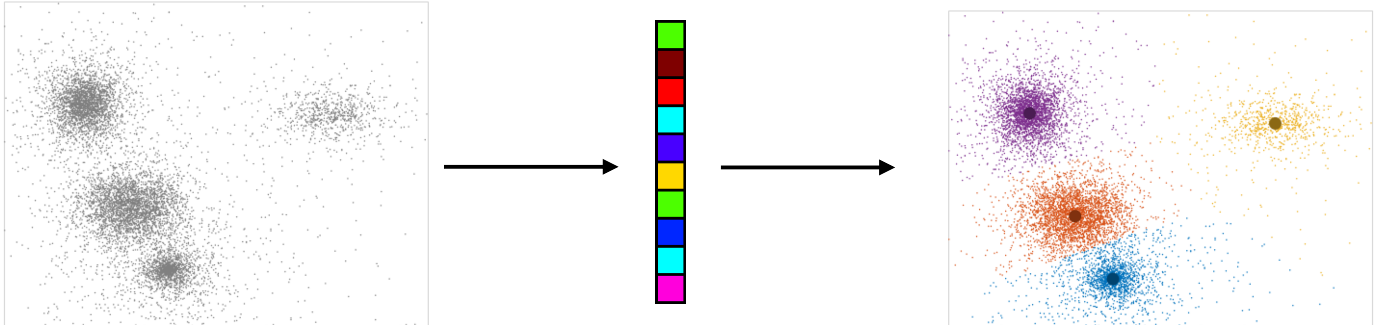
Fluorescence microscopy (3D)
PALM, STORM… *[Betzig 2006]*

Astronomy (2D)
[Puschmann 2017]



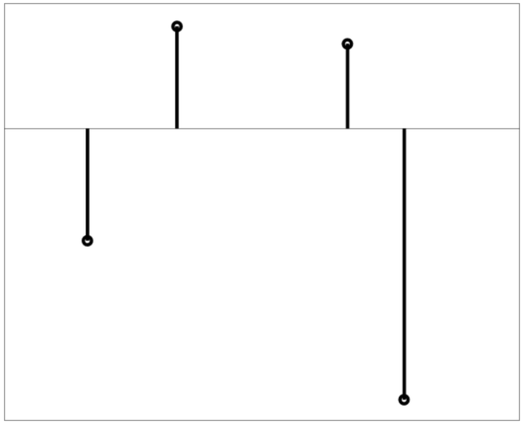
- Neuro-imaging with EEG (3D) *[Gramfort 2013]*
- 1-layer neural network (many D) *[Bach 2017]*
- Radar
- Geophysics
- …

Compressive mixture model learning (many D)
[Keriven 2017]



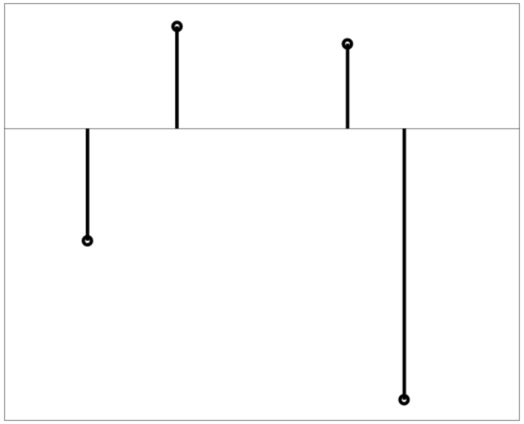
$$y = \Phi \mu_{\mathrm{emp}}.$$

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



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

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

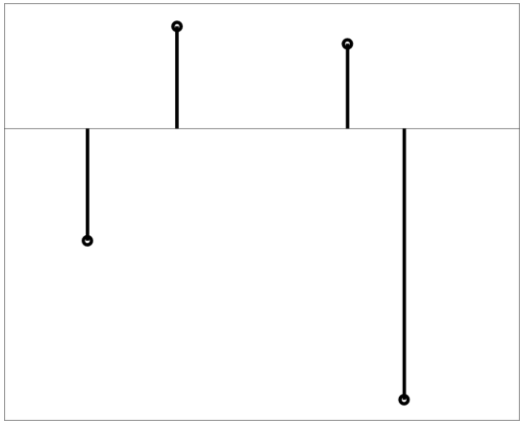


$\mu \in \mathcal{M}(\mathbb{T})$

$\longrightarrow$ **First Fourier coefficients**

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

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

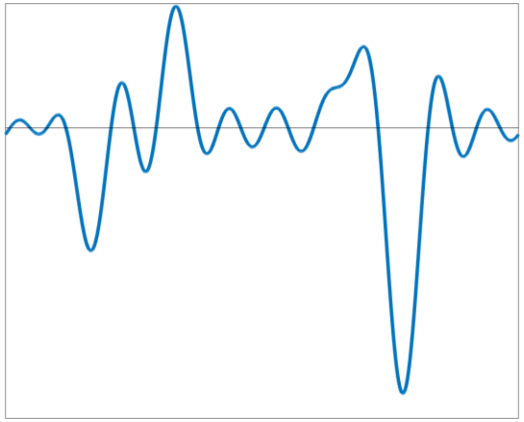


$\mu \in \mathcal{M}(\mathbb{T})$

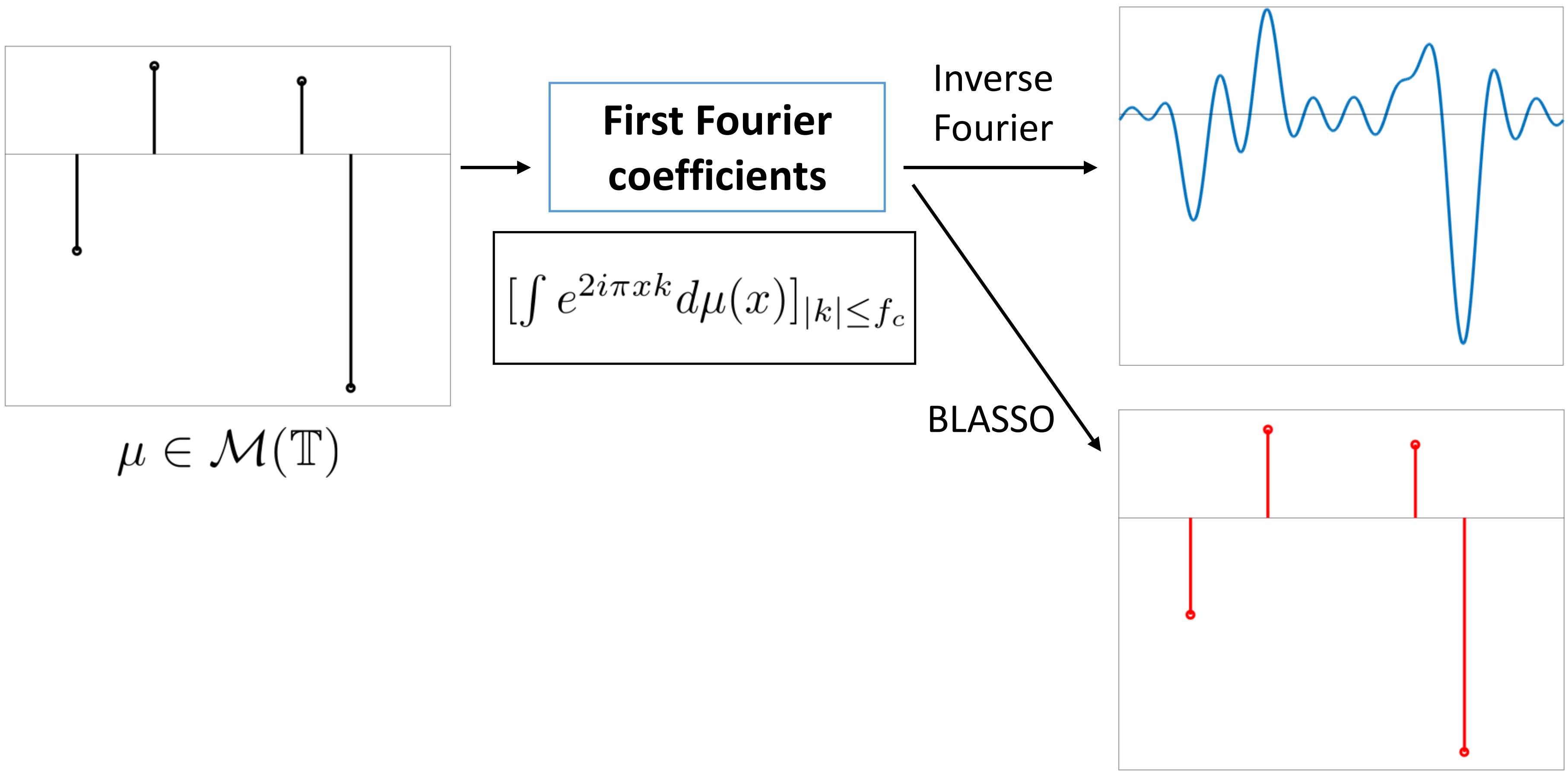
First Fourier coefficients

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

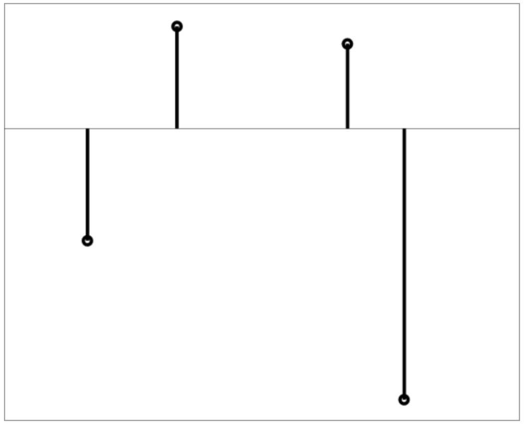
$\xrightarrow{\text{Inverse Fourier}}$



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



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

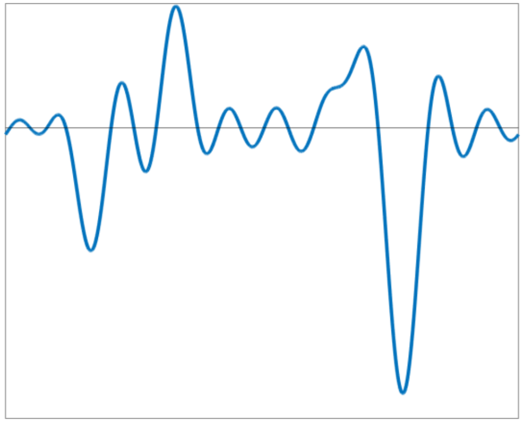


$\mu \in \mathcal{M}(\mathbb{T})$

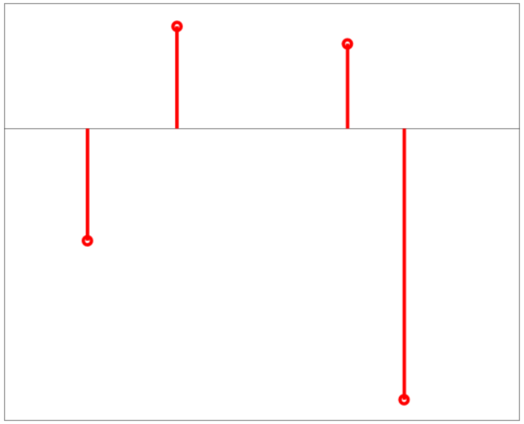
First Fourier coefficients

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

Inverse Fourier

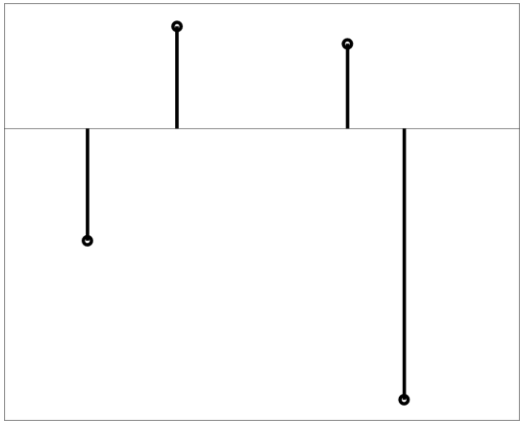


BLASSO



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

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

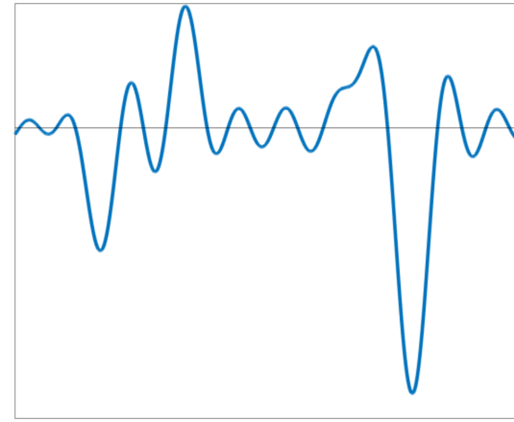


$\mu \in \mathcal{M}(\mathbb{T})$

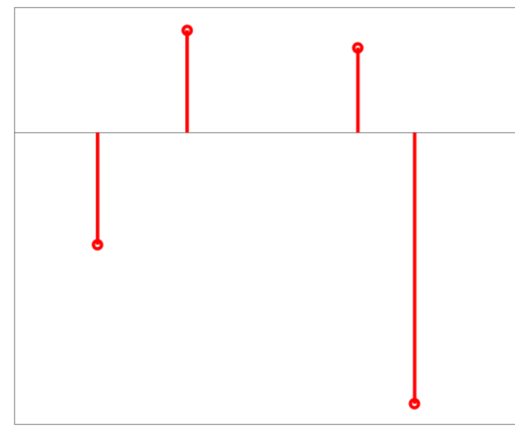
→ **First Fourier coefficients**

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

Inverse Fourier →

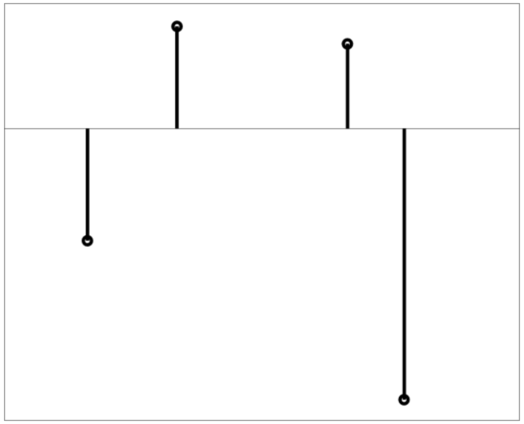


BLASSO ↘



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

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

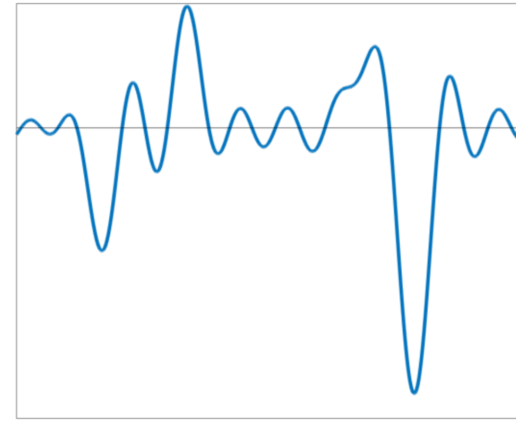


$\mu \in \mathcal{M}(\mathbb{T})$

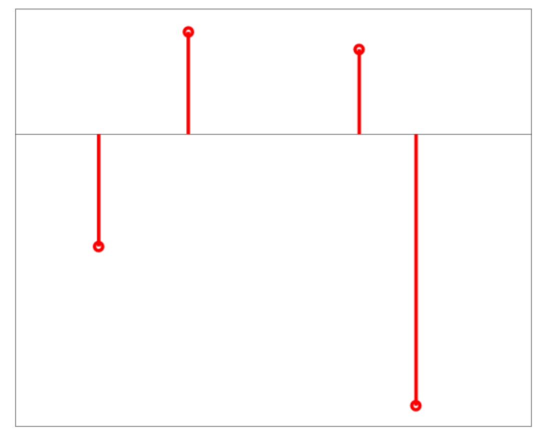
$\rightarrow$ **First Fourier coefficients**

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

Inverse Fourier $\rightarrow$

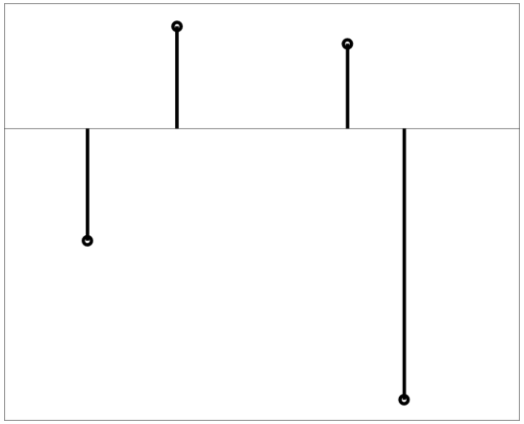


BLASSO $\rightarrow$



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

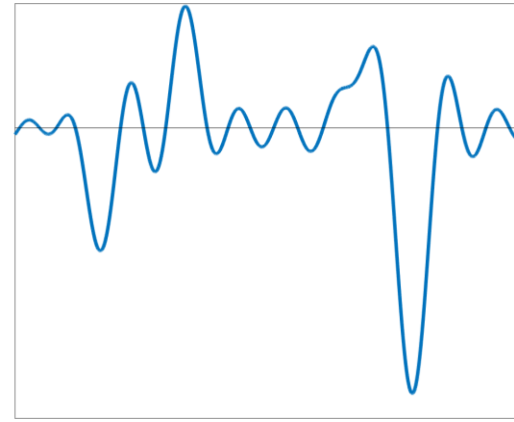


$\mu \in \mathcal{M}(\mathbb{T})$

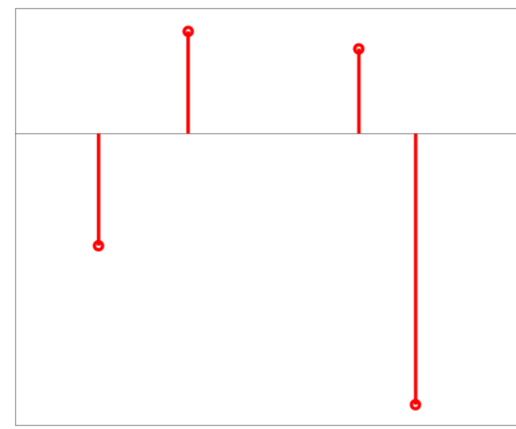
→ **First Fourier coefficients**

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

Inverse Fourier →



BLASSO →



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

Many extensions since…

Previous work, contribution

Relevant previous works:

- *[Tang, Recht 2013]*: $m \geq s \log(s) \log(f_c)$ **random** Fourier coefficients are sufficient

Previous work, contribution

Relevant previous works:

- *[Tang, Recht 2013]*: $m \geq s \log(s) \log(f_c)$ **random** Fourier coefficients are sufficient
  - **Random signs assumption**
  - 1D discrete Fourier

Previous work, contribution

Relevant previous works:

- *[Tang, Recht 2013]*: $m \geq s \log(s) \log(f_c)$ **random** Fourier coefficients are sufficient
  - **Random signs assumption**
  - 1D discrete Fourier
- *[Bendory et al. 2016]*: extension to other measurement operators

Previous work, contribution

Relevant previous works:

- *[Tang, Recht 2013]*: $m \geq s \log(s) \log(f_c)$ **random** Fourier coefficients are sufficient
  - **Random signs assumption**
  - 1D discrete Fourier
- *[Bendory et al. 2016]*: extension to other measurement operators
  - Minimal separation does not take into account **geometry** of the meas. operator

Previous work, contribution

Relevant previous works:

- *[Tang, Recht 2013]*: $m \geq s \log(s) \log(f_c)$ **random** Fourier coefficients are sufficient
  - **Random signs assumption**
  - 1D discrete Fourier
- *[Bendory et al. 2016]*: extension to other measurement operators
  - Minimal separation does not take into account **geometry** of the meas. operator
- *[Duval, Peyré 2015]*: In some cases, **support stability in the small noise regime**

Previous work, contribution

Relevant previous works:

- *[Tang, Recht 2013]*: $m \geq s\log(s)\log(f_c)$ **random** Fourier coefficients are sufficient
  - **Random signs assumption**
  - 1D discrete Fourier
- *[Bendory et al. 2016]*: extension to other measurement operators
  - Minimal separation does not take into account **geometry** of the meas. operator
- *[Duval, Peyré 2015]*: In some cases, **support stability in the small noise regime**
  - **Noise level** under which support stability is achieved?

Previous work, contribution

Relevant previous works:

- *[Tang, Recht 2013]*: $m \geq s \log(s) \log(f_c)$ **random** Fourier coefficients are sufficient
  - **Random signs assumption**
  - 1D discrete Fourier
- *[Bendory et al. 2016]*: extension to other measurement operators
  - Minimal separation does not take into account **geometry** of the meas. operator
- *[Duval, Peyré 2015]*: In some cases, **support stability in the small noise regime**
  - **Noise level** under which support stability is achieved?

Contributions:

- Generalize to many multi-d measurement operators, express the minimal separation as a **geometry-aware Fisher metric**

Previous work, contribution

Relevant previous works:

- *[Tang, Recht 2013]*: $m \geq s\log(s)\log(f_c)$ **random** Fourier coefficients are sufficient
  - **Random signs assumption**
  - 1D discrete Fourier
- *[Bendory et al. 2016]*: extension to other measurement operators
  - Minimal separation does not take into account **geometry** of the meas. operator
- *[Duval, Peyré 2015]*: In some cases, **support stability in the small noise regime**
  - **Noise level** under which support stability is achieved?

Contributions:

- Generalize to many multi-d measurement operators, express the minimal separation as a **geometry-aware Fisher metric**
- 1: **Remove the random sign assumption** *(weak convergence)*

Previous work, contribution

Relevant previous works:

- *[Tang, Recht 2013]*: $m \geq s \log(s) \log(f_c)$ **random** Fourier coefficients are sufficient
  - **Random signs assumption**
  - 1D discrete Fourier
- *[Bendory et al. 2016]*: extension to other measurement operators
  - Minimal separation does not take into account **geometry** of the meas. operator
- *[Duval, Peyré 2015]*: In some cases, **support stability in the small noise regime**
  - **Noise level** under which support stability is achieved?

Contributions:

- Generalize to many multi-d measurement operators, express the minimal separation as a **geometry-aware Fisher metric**
- 1: **Remove the random sign assumption** *(weak convergence)*
- 2: Prove **support stability when** $\|w\| \leq s^{-1}$ *(with random signs)*

Outline

1. Background on dual certificates
2. Minimal separation and Fisher metric
3. Main results, applications
4. Conclusion, outlooks

Dual certificates

Random linear operator:

$$\omega_1, \dots, \omega_m \overset{iid}{\sim} \Lambda$$

$$\Phi\mu = \frac{1}{\sqrt{m}} \left[\int \varphi_{\omega_k}(x) d\mu(x) \right]_{k=1}^{m}$$

Dual certificates

Random linear operator:

$$\omega_1, ..., \omega_m \overset{iid}{\sim} \Lambda$$

$$\Phi\mu = \frac{1}{\sqrt{m}} \left[\int \varphi_{\omega_k}(x) d\mu(x) \right]_{k=1}^{m}$$

Noisy measurement: $y = \Phi\mu_0 + w$

Dual certificates

Random linear operator:

$$\omega_1, \ldots, \omega_m \overset{iid}{\sim} \Lambda$$

$$\Phi\mu = \frac{1}{\sqrt{m}} \left[\int \varphi_{\omega_k}(x) d\mu(x) \right]_{k=1}^{m}$$

Noisy measurement: $y = \Phi\mu_0 + w$

The BLASSO problem:

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

Dual certificates

Random linear operator:

$$\omega_1, \dots, \omega_m \overset{iid}{\sim} \Lambda$$

$$\Phi\mu = \frac{1}{\sqrt{m}} \left[\int \varphi_{\omega_k}(x) d\mu(x) \right]_{k=1}^{m}$$

Noisy measurement: $y = \Phi\mu_0 + w$

The BLASSO problem:

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

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

Random linear operator:

$$\omega_1, \dots, \omega_m \overset{iid}{\sim} \Lambda$$

$$\Phi\mu = \frac{1}{\sqrt{m}} \Big[\int \varphi_{\omega_k}(x) d\mu(x) \Big]_{k=1}^m$$

Noisy measurement: $y = \Phi\mu_0 + w$

The BLASSO problem:

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

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 certificate **(noiseless case)**

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

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 ?

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

$$\boxed{\eta(x) = \sum_{k=1}^m h_k \varphi_{\omega_k}(x)}$$

What does it look like ?

What is a dual certificate ?

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

$$\boxed{\eta(x) = \sum_{k=1}^{m} h_k \varphi_{\omega_k}(x)}$$

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

What does it look like ?

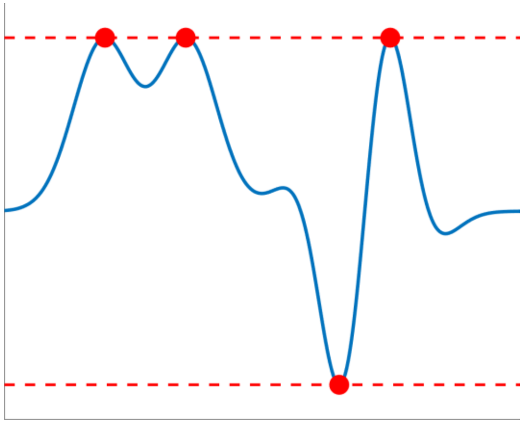
What is a dual certificate ?

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

$$\boxed{\eta(x) = \sum_{k=1}^{m} h_k \varphi_{\omega_k}(x)}$$

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

$$\boxed{\begin{gathered}\eta(x_i) = \mathrm{sign}(a_i) \\ \|\eta\|_\infty \leq 1\end{gathered}}$$



What does it look like ?

What is a dual certificate ?

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

$$\boxed{\eta(x) = \textstyle\sum_{k=1}^m h_k \varphi_{\omega_k}(x)}$$

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

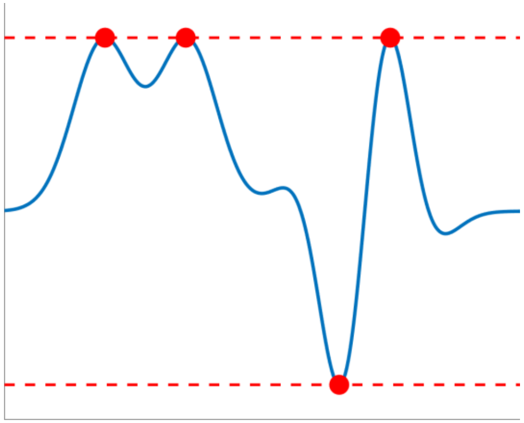
$$\eta(x_i) = \mathrm{sign}(a_i)$$

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

Non-degenerate dual certif.

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

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



Ensures **uniqueness** and **robustness**...

What does it look like ?

What is a dual certificate ?

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

$$\boxed{\eta(x) = \sum_{k=1}^m h_k \varphi_{\omega_k}(x)}$$

Intuitively:

Larger m -> easier interpolation

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

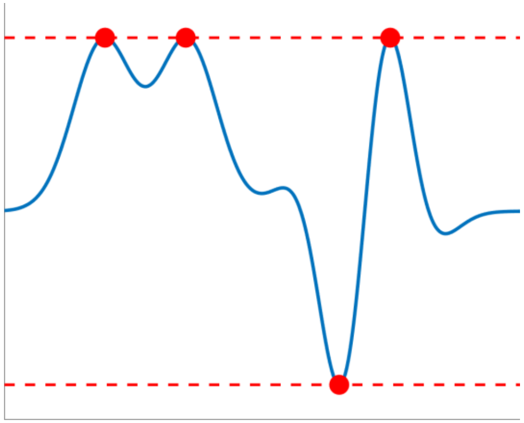
$$\eta(x_i) = \mathrm{sign}(a_i)$$

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

Non-degenerate dual certif.

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

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



Ensures **uniqueness** and **robustness**...

Proof strategy

Step 1: Study the limit case $m \to \infty$ to derive an appropriate notion of minimal separation

Proof strategy

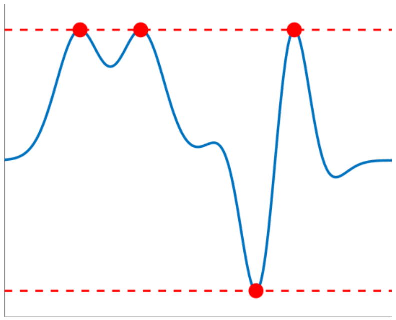
Step 1: Study the limit case $m \to \infty$ to derive an appropriate notion of minimal separation

Step 2: bound the deviation for finite number of measurements

Proof strategy

Step 1: Study the limit case $m \to \infty$ to derive an appropriate notion of minimal separation

Step 2: bound the deviation for finite number of measurements

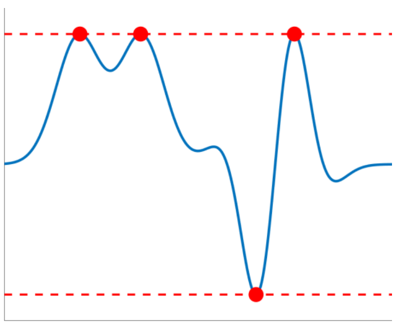


$m \to \infty$

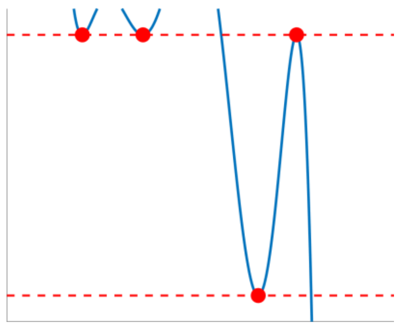
Proof strategy

Step 1: Study the limit case $m \to \infty$ to derive an appropriate notion of minimal separation

Step 2: bound the deviation for finite number of measurements



$m \to \infty$

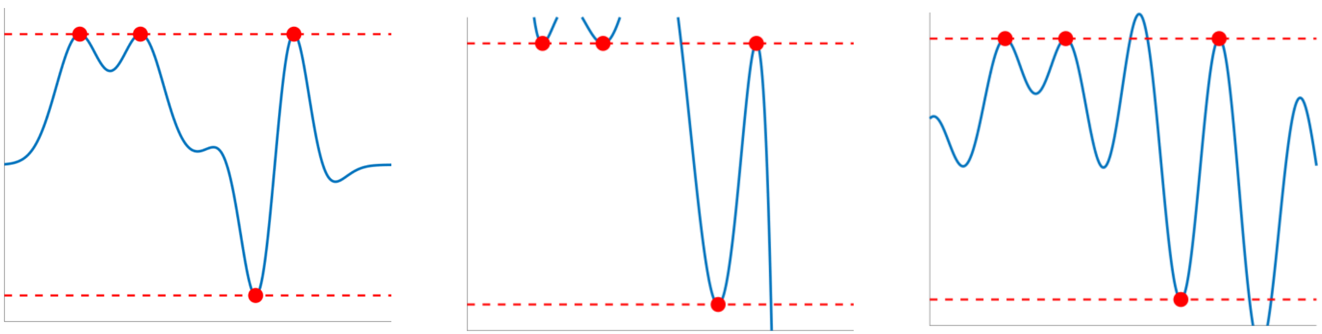


m=10

Proof strategy

Step 1: Study the limit case $m \to \infty$ to derive an appropriate notion of minimal separation

Step 2: bound the deviation for finite number of measurements

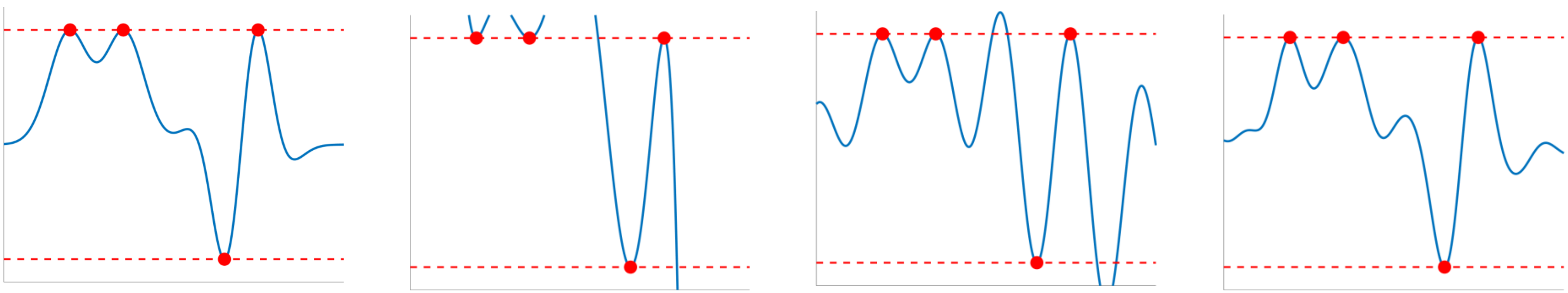


$m \to \infty$ *m=10* *m=50*

Proof strategy

Step 1: Study the limit case $m \to \infty$ to derive an appropriate notion of minimal separation

Step 2: bound the deviation for finite number of measurements

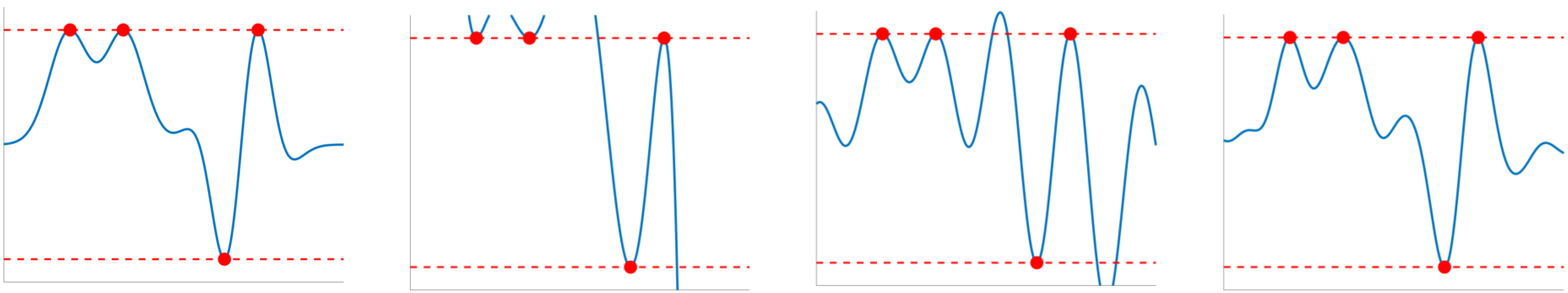


$m \to \infty$ | *m=10* | *m=50* | *m=500*

Proof strategy

Step 1: Study the limit case $m \to \infty$ to derive an appropriate notion of minimal separation

Step 2: bound the deviation for finite number of measurements



$m \to \infty$ *m=10* *m=50* *m=500*

Step 3: recovery

- Adaptation of *[Azaïs 2015]* for weak convergence
- Quantitative Implicit Function Theorem *[Denoyelle 2015]* for support stability

Outline

1. Background on dual certificates
2. Minimal separation and Fisher metric
3. Main results, applications
4. Conclusion, outlooks

Limit covariance kernels

How to construct a certificate ?

Study **limit covariance kernel** when $m \to \infty$:

$$\kappa(x, x') = \mathbb{E}_\omega \varphi_\omega(x) \varphi_\omega(x')$$

Limit covariance kernels

How to construct a certificate ?

Study **limit covariance kernel** when $m \to \infty$:

$$\kappa(x, x') = \mathbb{E}_\omega \varphi_\omega(x)\varphi_\omega(x')$$

Sub-sampled version:

$$\kappa(x, x') = \frac{1}{m} \sum_{k=1}^{m} \varphi_{\omega_k}(x)\varphi_{\omega_k}(x')$$

Limit covariance kernels

How to construct a certificate ?

Study **limit covariance kernel** when $m \to \infty$:

$$\kappa(x, x') = \mathbb{E}_\omega \varphi_\omega(x) \varphi_\omega(x')$$

Sub-sampled version:

$$\kappa(x, x') = \frac{1}{m} \sum_{k=1}^{m} \varphi_{\omega_k}(x) \varphi_{\omega_k}(x')$$

Strategy under minimal separation

Limit covariance kernels

How to construct a certificate ?

Study **limit covariance kernel** when $m \to \infty$:

$$\kappa(x, x') = \mathbb{E}_\omega \varphi_\omega(x) \varphi_\omega(x')$$

Sub-sampled version:

$$\kappa(x, x') = \frac{1}{m} \sum_{k=1}^{m} \varphi_{\omega_k}(x) \varphi_{\omega_k}(x')$$

Strategy under minimal separation

$$\eta \in \mathrm{Span}\{\kappa(x_i, .), \partial\kappa(x_i, .)\} \subset \mathrm{Im}(\Phi^\star)$$

Limit covariance kernels

How to construct a certificate ?

Study **limit covariance kernel** when $m \to \infty$:

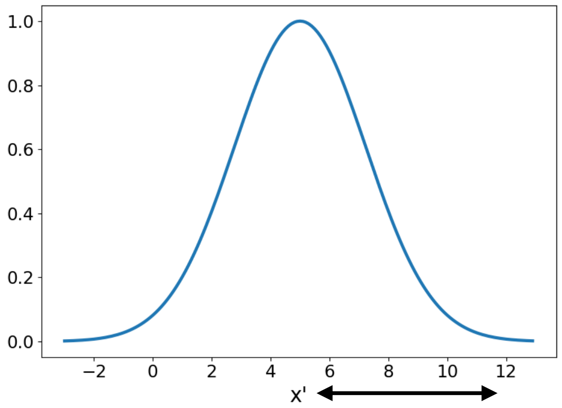
$$\kappa(x, x') = \mathbb{E}_\omega \varphi_\omega(x)\varphi_\omega(x')$$

Sub-sampled version:

$$\kappa(x, x') = \frac{1}{m}\sum_{k=1}^m \varphi_{\omega_k}(x)\varphi_{\omega_k}(x')$$

Strategy under minimal separation

$$\eta \in \mathrm{Span}\{\kappa(x_i, .), \partial\kappa(x_i, .)\} \subset \mathrm{Im}(\Phi^\star)$$



Min. Separation Δ

Limit covariance kernels

How to construct a certificate ?

Study **limit covariance kernel** when $m \to \infty$:

$$\kappa(x, x') = \mathbb{E}_\omega \varphi_\omega(x)\varphi_\omega(x')$$

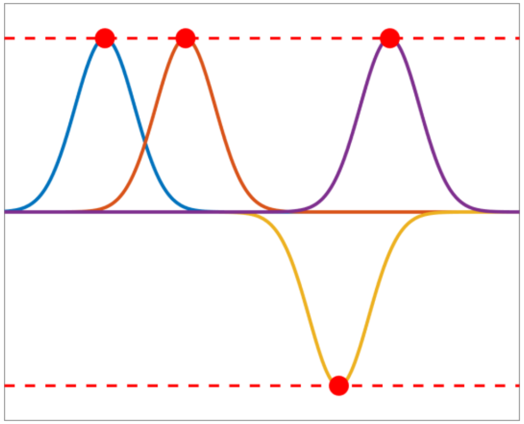
Sub-sampled version:

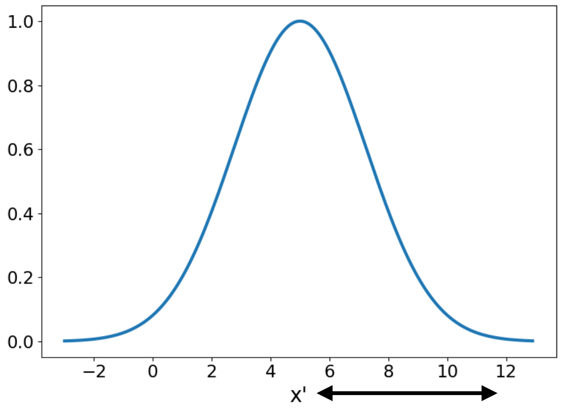
$$\kappa(x, x') = \frac{1}{m}\sum_{k=1}^{m} \varphi_{\omega_k}(x)\varphi_{\omega_k}(x')$$

Strategy under minimal separation

$$\eta \in \mathrm{Span}\{\kappa(x_i, .), \partial\kappa(x_i, .)\} \subset \mathrm{Im}(\Phi^\star)$$

1: kernel at each saturation point





Min. Separation Δ

Limit covariance kernels

How to construct a certificate ?

Study **limit covariance kernel** when $m \to \infty$:

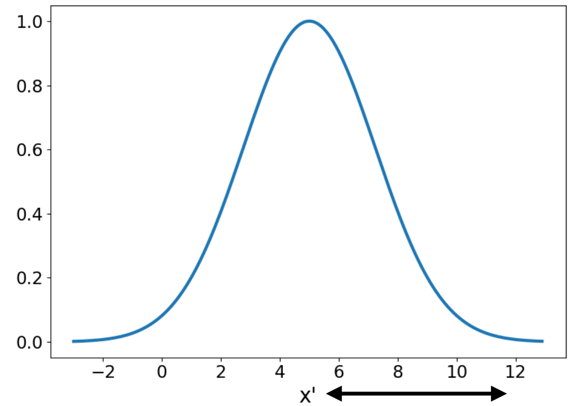
$$\kappa(x, x') = \mathbb{E}_\omega \varphi_\omega(x) \varphi_\omega(x')$$

Sub-sampled version:

$$\kappa(x, x') = \frac{1}{m} \sum_{k=1}^{m} \varphi_{\omega_k}(x) \varphi_{\omega_k}(x')$$

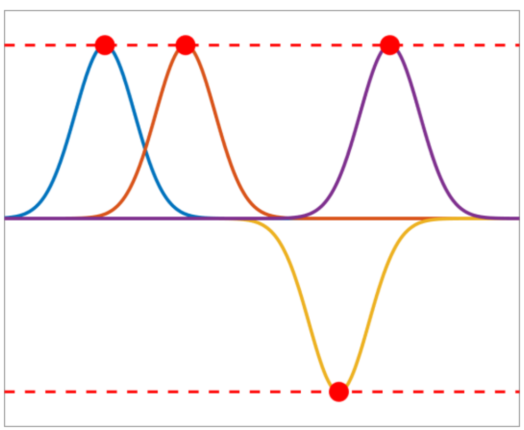
Strategy under minimal separation

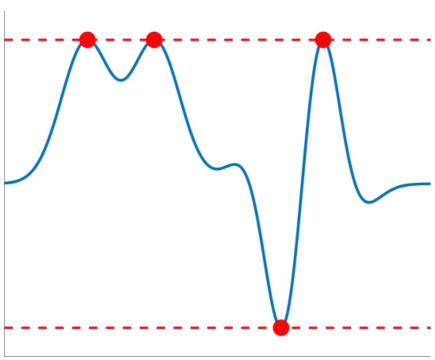
$$\eta \in \mathrm{Span}\{\kappa(x_i, .), \partial\kappa(x_i, .)\} \subset \mathrm{Im}(\Phi^\star)$$



Min. Separation Δ

1: kernel at each saturation point





2: ***Small*** adjustements (minimal separation)

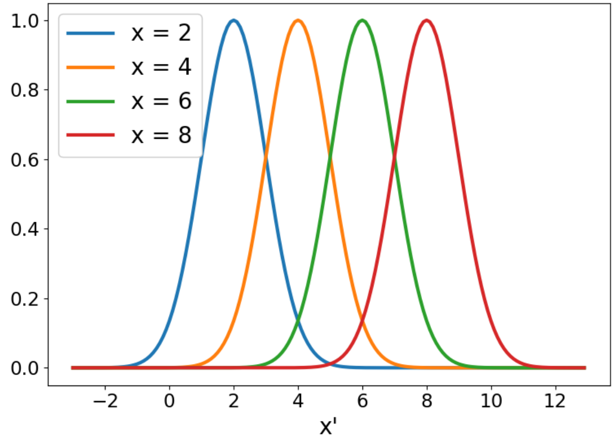
Minimal separation and Fisher metric

Which metric for separation ?

Minimal separation and Fisher metric

Which metric for separation ?

Classical case: translation-invariant kernel

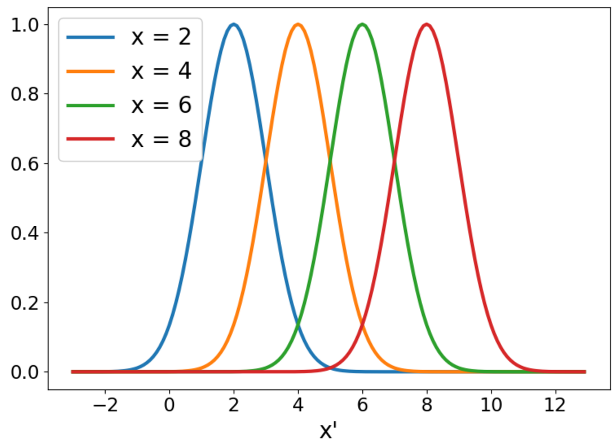


$$\kappa(x, x') = \kappa(x - x')$$

Minimal separation and Fisher metric

Which metric for separation ?

Classical case: translation-invariant kernel



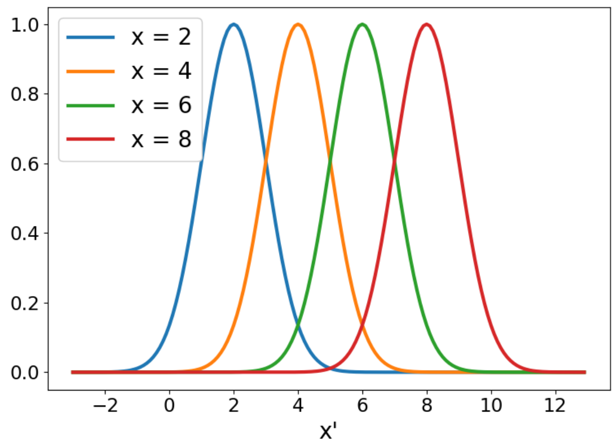
$$\kappa(x, x') = \kappa(x - x')$$

$\|x - x'\|$ natural

Minimal separation and Fisher metric

Which metric for separation ?

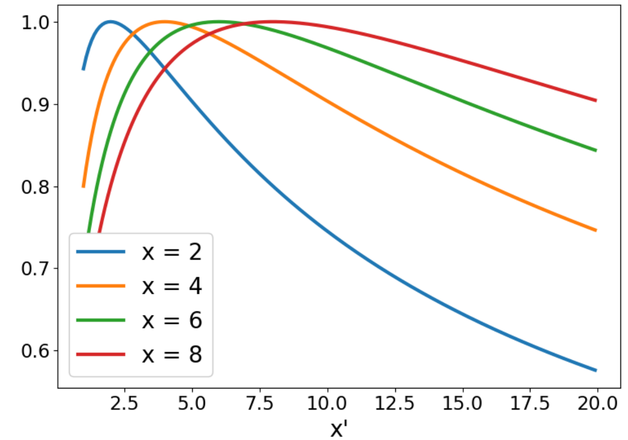
Classical case: translation-invariant kernel



$$\kappa(x, x') = \kappa(x - x')$$

$\boxed{\|x - x'\| \quad \text{natural}}$

Non translation-inv. ?

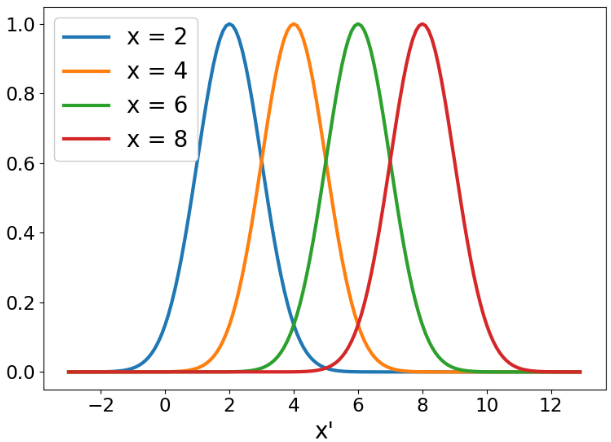


Kernel for microscopy

Minimal separation and Fisher metric

Which metric for separation ?

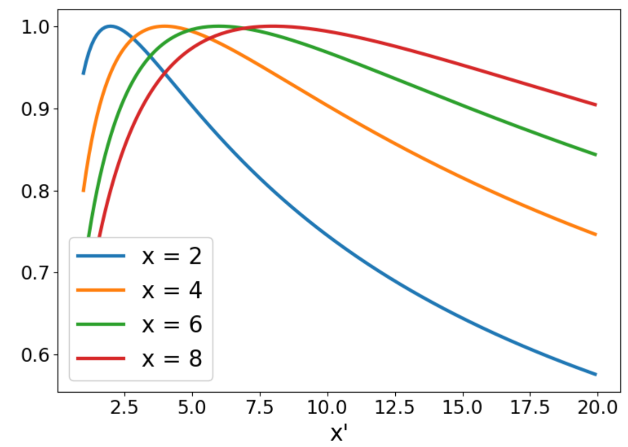
Classical case: translation-invariant kernel



$\kappa(x,x') = \kappa(x-x')$

$\|x-x'\|$ natural

Non translation-inv. ?



Kernel for microscopy

Riemannian metric associated to a kernel *[Amari 99]*:

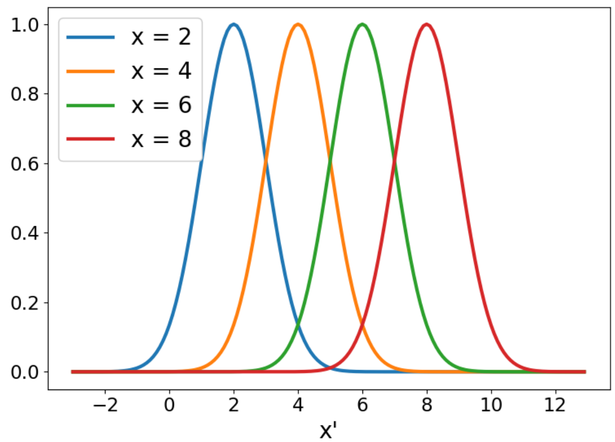
$H_x = \nabla_1 \nabla_2 \kappa(x,x)$: metric tensor

$d_H(x,x')$: geodesic distance

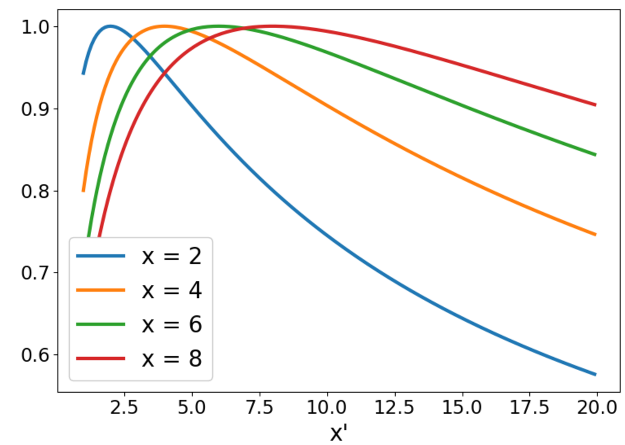
Minimal separation and Fisher metric

Which metric for separation ?

Classical case: translation-invariant kernel



Non translation-inv. ?



$\kappa(x,x') = \kappa(x - x')$ $\boxed{\|x - x'\| \text{ natural}}$

Kernel for microscopy

Riemannian metric associated to a kernel *[Amari 99]*:

$H_x = \nabla_1 \nabla_2 \kappa(x,x)$: metric tensor $d_H(x,x')$: geodesic distance

Thm: under some hypothesis, for $d_H(x_i,x_j) \geq \Delta$, there exists non-degenerate η

Examples

Kernel

Features

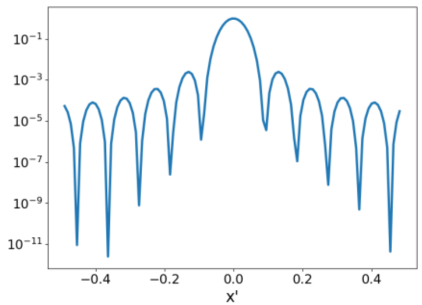
Fisher metric and minimal separation

Examples

Kernel

Discrete Fourier on Torus:

Féjer kernel



Features

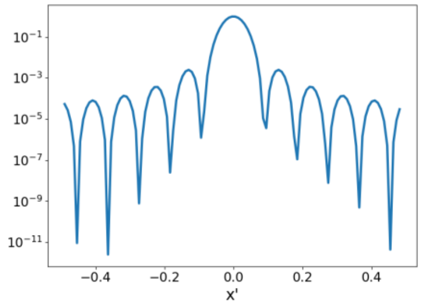
Fisher metric and minimal separation

Examples

Kernel

Discrete Fourier on Torus:

Féjer kernel



Features

$$\varphi_\omega(x) = e^{2\pi i \omega^\top x}$$

$$\Lambda \propto \prod_{j=1}^{d} g_j(\omega_j)$$

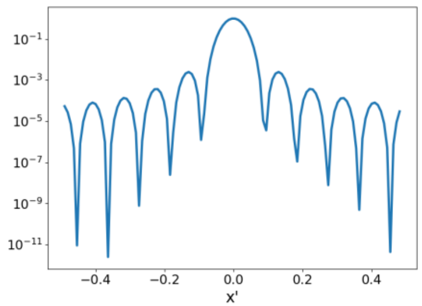
Fisher metric and minimal separation

Examples

Kernel

Discrete Fourier on Torus:

Féjer kernel



Features

$$\varphi_\omega(x) = e^{2\pi i \omega^\top x}$$

$$\Lambda \propto \prod_{j=1}^{d} g_j(\omega_j)$$

Fisher metric and minimal separation

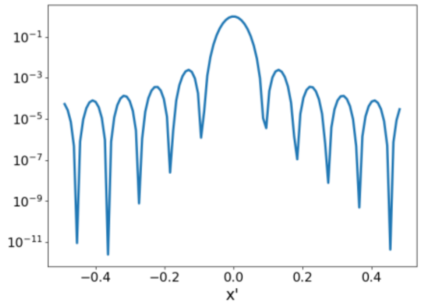
$$d_H(x, x') \propto \|x - x'\|_2$$

$$\Delta = \sqrt{d\sqrt{s}}/f_c$$

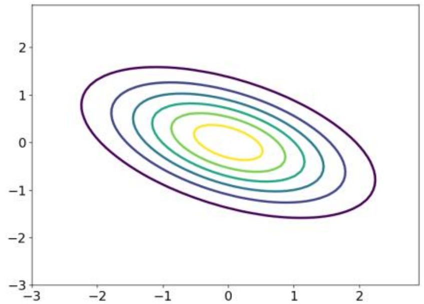
Examples

Kernel

Discrete Fourier on Torus:
Féjer kernel



Continuous Gaussian Fourier:
Gaussian kernel



Features

$$\varphi_\omega(x) = e^{2\pi i \omega^\top x}$$

$$\Lambda \propto \prod_{j=1}^{d} g_j(\omega_j)$$

Fisher metric and minimal separation

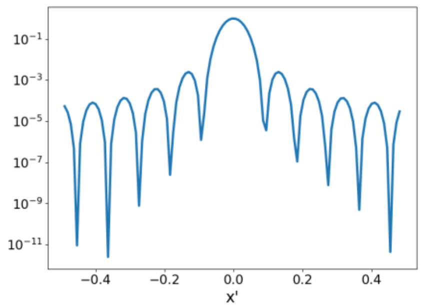
$$d_H(x, x') \propto \|x - x'\|_2$$

$$\Delta = \sqrt{d\sqrt{s}}/f_c$$

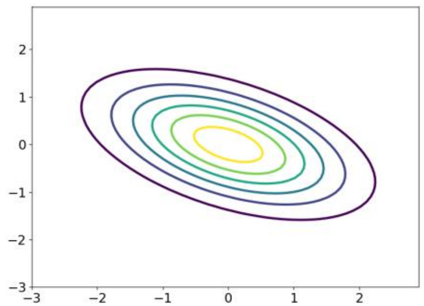
Examples

Kernel

Discrete Fourier on Torus:
Féjer kernel



Continuous Gaussian Fourier:
Gaussian kernel



Features

$$\varphi_\omega(x) = e^{2\pi i \omega^\top x}$$

$$\Lambda \propto \prod_{j=1}^d g_j(\omega_j)$$

$$\varphi_\omega(x) = e^{i \omega^\top x}$$

$$\Lambda = \mathcal{N}(0, \Sigma^{-1})$$

Fisher metric and minimal separation

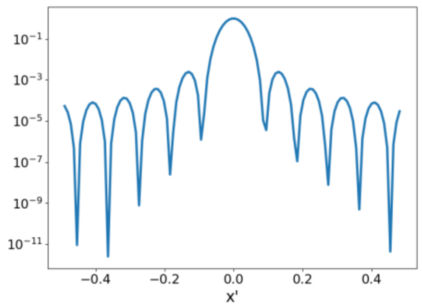
$$d_H(x, x') \propto \|x - x'\|_2$$

$$\Delta = \sqrt{d\sqrt{s}}/f_c$$

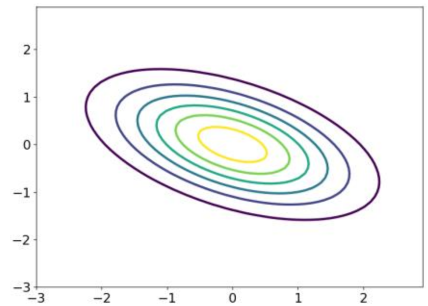
Examples

Kernel

Discrete Fourier on Torus:
Féjer kernel



Continuous Gaussian Fourier:
Gaussian kernel



Features

$$\varphi_\omega(x) = e^{2\pi i \omega^\top x} \qquad \varphi_\omega(x) = e^{i\omega^\top x}$$

$$\Lambda \propto \prod_{j=1}^{d} g_j(\omega_j) \qquad \Lambda = \mathcal{N}(0, \Sigma^{-1})$$

Fisher metric and minimal separation

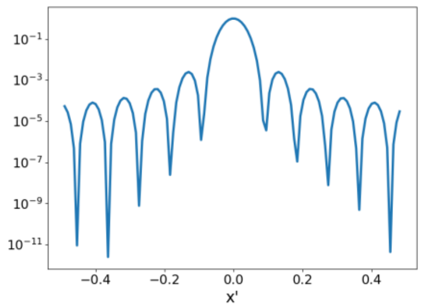
$$d_H(x, x') \propto \|x - x'\|_2 \qquad d_H(x, x') = \|x - x'\|_{\Sigma^{-1}}$$

$$\Delta = \sqrt{d\sqrt{s}}/f_c \qquad \Delta = \sqrt{\log(s)}$$

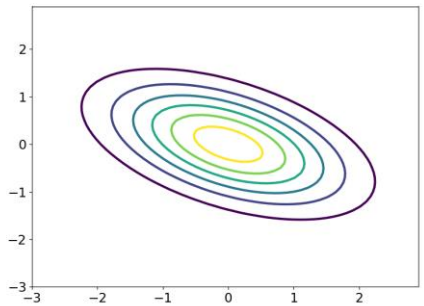
Examples

Kernel

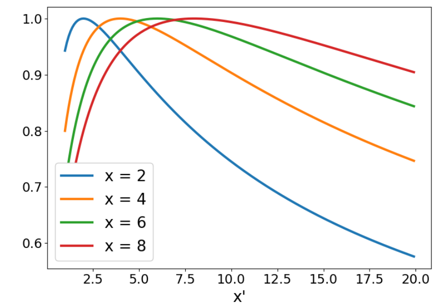
Discrete Fourier on Torus: *Féjer kernel*



Continuous Gaussian Fourier: *Gaussian kernel*



Microscopy (Laplace transform):



Features

$$\varphi_\omega(x) = e^{2\pi i \omega^\top x}$$

$$\varphi_\omega(x) = e^{i \omega^\top x}$$

$$\Lambda \propto \prod_{j=1}^{d} g_j(\omega_j)$$

$$\Lambda = \mathcal{N}(0, \Sigma^{-1})$$

Fisher metric and minimal separation

$$d_H(x, x') \propto \|x - x'\|_2$$

$$d_H(x, x') = \|x - x'\|_{\Sigma^{-1}}$$

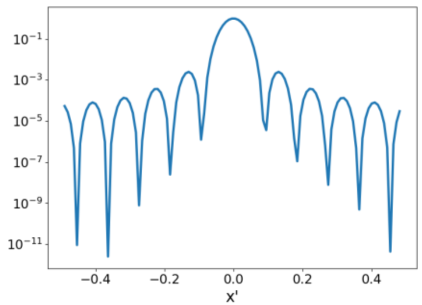
$$\Delta = \sqrt{d\sqrt{s}}/f_c$$

$$\Delta = \sqrt{\log(s)}$$

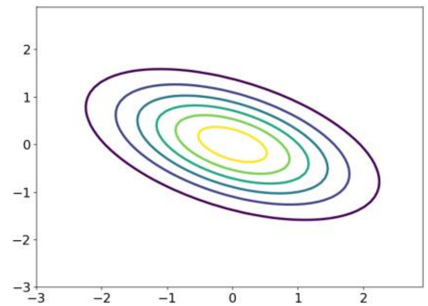
Examples

Kernel

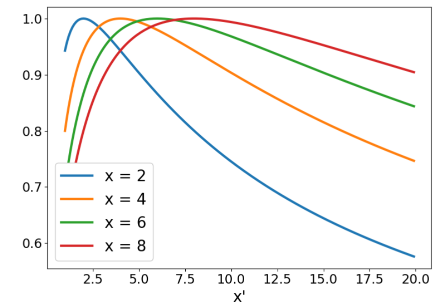
Discrete Fourier on Torus: *Féjer kernel*



Continuous Gaussian Fourier: *Gaussian kernel*



Microscopy (Laplace transform):



Features

$$\varphi_\omega(x) = e^{2\pi i \omega^\top x}$$

$$\Lambda \propto \prod_{j=1}^d g_j(\omega_j)$$

$$\varphi_\omega(x) = e^{i\omega^\top x}$$

$$\Lambda = \mathcal{N}(0, \Sigma^{-1})$$

$$\varphi_\omega(x) = \prod_{j=1}^d \frac{x_i+\alpha_i}{\alpha_i} e^{-\omega^\top x}$$

$$\Lambda \propto e^{-\alpha^\top \omega}$$

Fisher metric and minimal separation

$$d_H(x,x') \propto \|x-x'\|_2$$

$$\Delta = \sqrt{d\sqrt{s}}/f_c$$

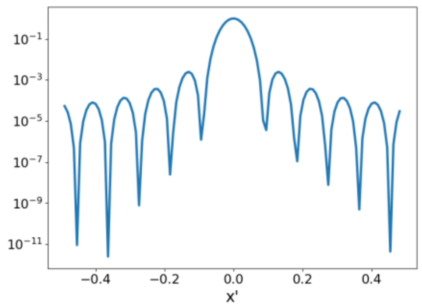
$$d_H(x,x') = \|x-x'\|_{\Sigma^{-1}}$$

$$\Delta = \sqrt{\log(s)}$$

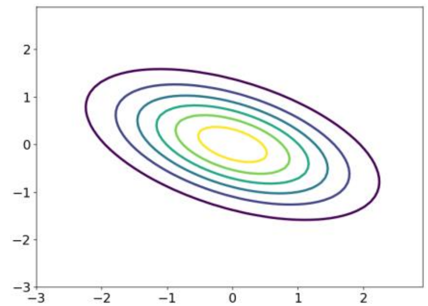
Examples

Kernel

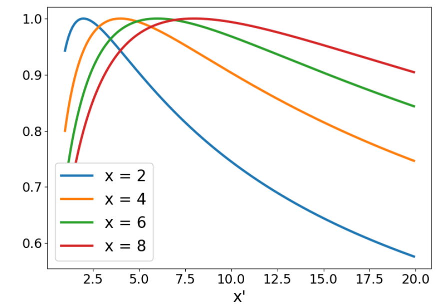
Discrete Fourier on Torus: *Féjer kernel*



Continuous Gaussian Fourier: *Gaussian kernel*



Microscopy (Laplace transform):



Features

$$\varphi_\omega(x) = e^{2\pi i \omega^\top x}$$

$$\Lambda \propto \prod_{j=1}^d g_j(\omega_j)$$

$$\varphi_\omega(x) = e^{i\omega^\top x}$$

$$\Lambda = \mathcal{N}(0, \Sigma^{-1})$$

$$\varphi_\omega(x) = \prod_{j=1}^d \frac{x_i+\alpha_i}{\alpha_i} e^{-\omega^\top x}$$

$$\Lambda \propto e^{-\alpha^\top \omega}$$

Fisher metric and minimal separation

$$d_H(x,x') \propto \|x-x'\|_2$$

$$\Delta = \sqrt{d\sqrt{s}}/f_c$$

$$d_H(x,x') = \|x-x'\|_{\Sigma^{-1}}$$

$$\Delta = \sqrt{\log(s)}$$

$$d_H(x,x') = \sqrt{\sum_j \left|\log\left(\frac{x_i+\alpha_i}{x'_i+\alpha_i}\right)\right|^2}$$

$$\Delta = d + \log(ds)$$

Outline

1. Background on dual certificates
2. Minimal separation and Fisher metric
3. Main results, applications
4. Conclusion, outlooks

Main results

Thm: Eliminating the random signs

Main results

Thm: Eliminating the random signs

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

Depend on kernel

Main results

Thm: Eliminating the random signs

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

Depend on kernel

- The recovered measure **concentrate** around true Diracs

Main results

Thm: Eliminating the random signs $$m \geq \mathcal{O}(sd^r \cdot \texttt{polylog}(s,d))$$

Depend on kernel

- The recovered measure **concentrate** around true Diracs
- Proof: ***golfing scheme***
  [Gross 2009, Candès Plan 2011]

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

Main results

Thm: Eliminating the random signs

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

Depend on kernel

- The recovered measure **concentrate** around true Diracs
- Proof: ***golfing scheme***
  [Gross 2009, Candès Plan 2011]

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

Thm: Support stability

Main results

Thm: Eliminating the random signs

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

Depend on kernel

- The recovered measure **concentrate** around true Diracs
- Proof: ***golfing scheme***
  [Gross 2009, Candès Plan 2011]

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

Thm: Support stability

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

With random signs

$$m \geq \mathcal{O}(s^{\frac{3}{2}} d^r \cdot \texttt{polylog}(s,d))$$

Without random signs

Main results

Thm: Eliminating the random signs

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

Depend on kernel

- The recovered measure **concentrate** around true Diracs
- Proof: ***golfing scheme***
  [Gross 2009, Candès Plan 2011]

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

Thm: Support stability

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

With random signs

$$m \geq \mathcal{O}(s^{\frac{3}{2}} d^r \cdot \texttt{polylog}(s, d))$$

Without random signs

- **Quantified small noise** : if $\lambda, \|w\| \leq \frac{\min_i |a_i|}{s d^q}$, then:

Main results

Thm: Eliminating the random signs

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

↖*Depend on kernel*

- The recovered measure **concentrate** around true Diracs
- Proof: ***golfing scheme***
  [Gross 2009, Candès Plan 2011]

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

Thm: Support stability

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

With random signs

$$m \geq \mathcal{O}(s^{\frac{3}{2}} d^r \cdot \texttt{polylog}(s,d))$$

Without random signs

- **Quantified small noise** : if $\lambda, \|w\| \leq \frac{\min_i |a_i|}{s d^q}$, then:
- The recovered measure is formed of exactly s Diracs

$$\sqrt{\sum_i |\tilde{a}_i - a_i|^2 + d_H(\tilde{x}_i, x_i)^2} \lesssim \frac{\sqrt{s}}{\min_i |a_i|}(\|w\| + \lambda)$$

Applications

- Féjer kernel (discrete Fourier): $m \geq sd^3, \, \|w\| \leq 1/(sd^3)$

Applications

- Féjer kernel (discrete Fourier): $m \geq sd^3,\ \|w\| \leq 1/(sd^3)$
- Microscopy with Laplace transform: $m \geq sd^7,\ \|w\| \leq 1/(sd^5)$

Applications

- Féjer kernel (discrete Fourier): $m \geq sd^3, \; \|w\| \leq 1/(sd^3)$
- Microscopy with Laplace transform: $m \geq sd^7, \; \|w\| \leq 1/(sd^5)$
- *Application*: Gaussian Mixture Model learning

Applications

- Féjer kernel (discrete Fourier): $m \geq sd^3, \; \|w\| \leq 1/(sd^3)$
- Microscopy with Laplace transform: $m \geq sd^7, \; \|w\| \leq 1/(sd^5)$
- *Application*: Gaussian Mixture Model learning

  - Given $z_1, \ldots, z_n \overset{iid}{\sim} \sum_{i=1}^{s} a_i \mathcal{N}(x_i, \Sigma)$

Applications

- Féjer kernel (discrete Fourier): $m \geq sd^3, \; \|w\| \leq 1/(sd^3)$

- Microscopy with Laplace transform: $m \geq sd^7, \; \|w\| \leq 1/(sd^5)$

- *Application*: Gaussian Mixture Model learning

  - Given $z_1, \ldots, z_n \overset{iid}{\sim} \sum_{i=1}^{s} a_i \mathcal{N}(x_i, \Sigma)$
  - Compute $y = \frac{1}{\sqrt{mn}} \left[\sum_{j=1}^{n} e^{i z_j^\top \omega_k} \right]_{k=1}^{m}$ with Gaussian $\omega_k \sim \mathcal{N}(0, \Sigma^{-1}/d)$

  (streaming, distributed)

Applications

- Féjer kernel (discrete Fourier): $m \geq sd^3, \; \|w\| \leq 1/(sd^3)$

- Microscopy with Laplace transform: $m \geq sd^7, \; \|w\| \leq 1/(sd^5)$

- *Application*: Gaussian Mixture Model learning

  - Given $z_1, \ldots, z_n \overset{iid}{\sim} \sum_{i=1}^{s} a_i \mathcal{N}(x_i, \Sigma)$
  - Compute $y = \frac{1}{\sqrt{mn}} \left[\sum_{j=1}^{n} e^{i z_j^\top \omega_k} \right]_{k=1}^{m}$ with Gaussian $\omega_k \sim \mathcal{N}(0, \Sigma^{-1}/d)$

    (streaming, distributed)
  - Solve the BLASSO with φ_ω as characteristic function of a Gaussian

Applications

- Féjer kernel (discrete Fourier): $m \geq sd^3, \; \|w\| \leq 1/(sd^3)$
- Microscopy with Laplace transform: $m \geq sd^7, \; \|w\| \leq 1/(sd^5)$
- *Application*: Gaussian Mixture Model learning

> - Given $z_1, \ldots, z_n \overset{iid}{\sim} \sum_{i=1}^{s} a_i \mathcal{N}(x_i, \Sigma)$
> - Compute $y = \frac{1}{\sqrt{mn}} \left[\sum_{j=1}^{n} e^{i z_j^\top \omega_k} \right]_{k=1}^{m}$ with Gaussian $\omega_k \sim \mathcal{N}(0, \Sigma^{-1}/d)$
>
>   *(streaming, distributed)*
> - Solve the BLASSO with φ_ω as characteristic function of a Gaussian

Then: if $m \geq s^{3/2} d^3, \quad n \geq s^2 d^6, \quad \|x_i - x_j\|_{\Sigma^{-1}} \geq \sqrt{d \log(s)}$

The BLASSO yields exactly s Diracs: *non-asymptotic* model selection !

Outline

1. Background on dual certificates
2. Minimal separation and Fisher metric
3. Main results, applications
4. Conclusion, outlooks

Summary, outlooks

- **Summary**: generalization of existing results on ***super-resolution with random measurements*** (and minimal separation)

  - Introduction of the ***kernel Fisher metric*** to measure minimal separation
    - Application in particular to a non-translation-invariant example for microscopy

  - **No need of random signs** for weak convergence (golfing scheme)
  - **Quantitative support stability**

Summary, outlooks

- **Summary**: generalization of existing results on ***super-resolution with random measurements*** (and minimal separation)

  - Introduction of the ***kernel Fisher metric*** to measure minimal separation
    - Application in particular to a non-translation-invariant example for microscopy

  - **No need of random signs** for weak convergence (golfing scheme)
  - **Quantitative support stability**

- **Outlooks**
  - **Implication of support stability for algorithms ?** (active field)
  - Better characterization of the « universality » of the geodesic distance
  - 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*

Poon, Keriven, Peyré. **Support Localization and the Fisher Metric for off-the-grid Sparse Regularization.** *Preprint arxiv:1810.03340*

data-ens.github.io

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