Off-the-Grid Sparse Estimation

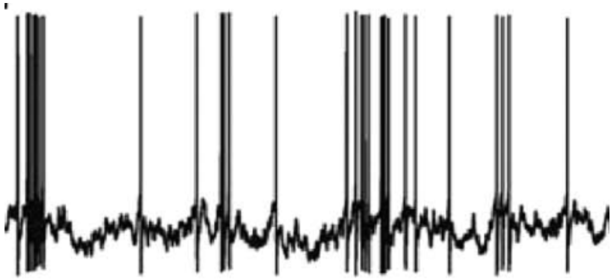
Gabriel Peyré

Joint work with
Vincent Duval, Quentin Denoyelle
Clarice Poon, Nicolas Keriven

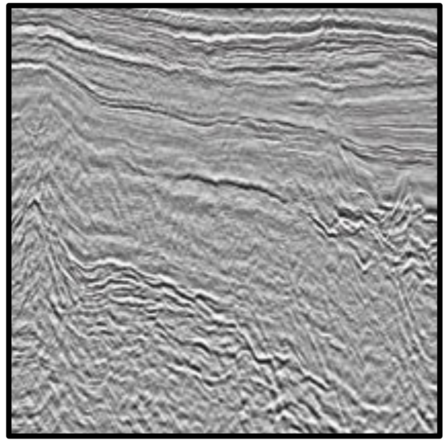
www.numerical-tours.com

Sparse Estimation

Recover pointwise sources from noisy low-resolution observations.

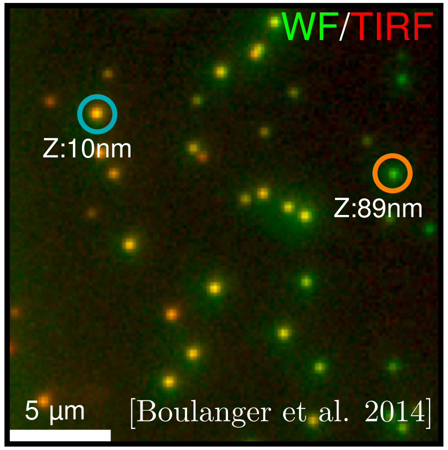


Neural spikes (1D)



Seismic imaging (1.5D)

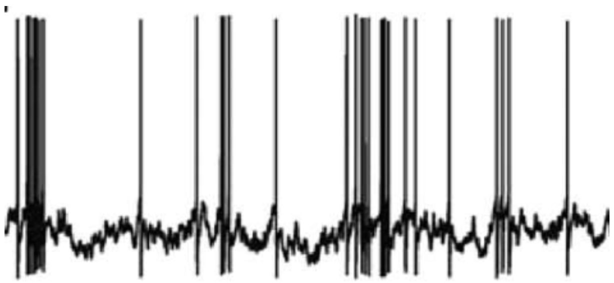
Astrophysics (2D)



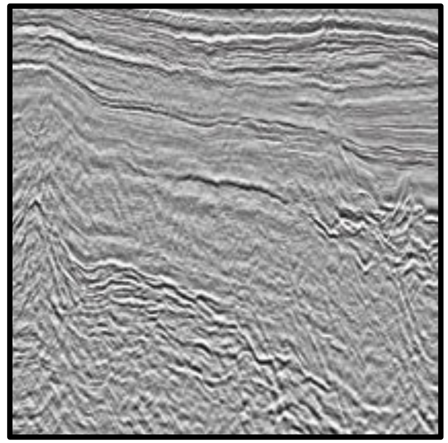
Single-molecule fluorescence (3-D)

Sparse Estimation

Recover pointwise sources from noisy low-resolution observations.

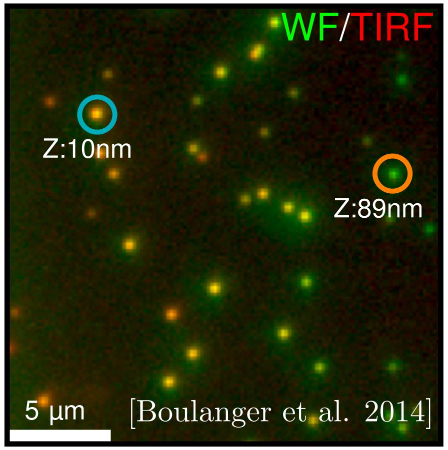


Neural spikes (1D)

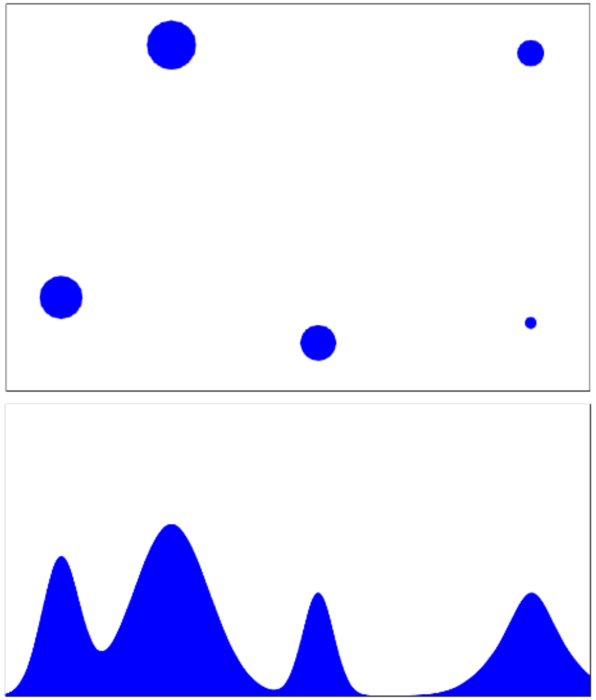


Seismic imaging (1.5D)

Astrophysics (2D)



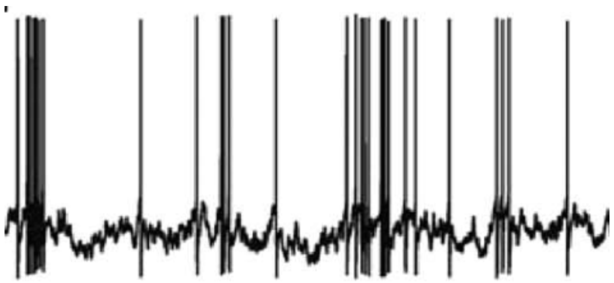
Single-molecule fluorescence (3-D)



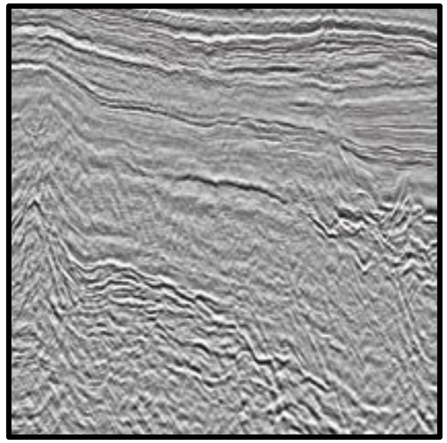
Mixture estimation

Sparse Estimation

Recover pointwise sources from noisy low-resolution observations.

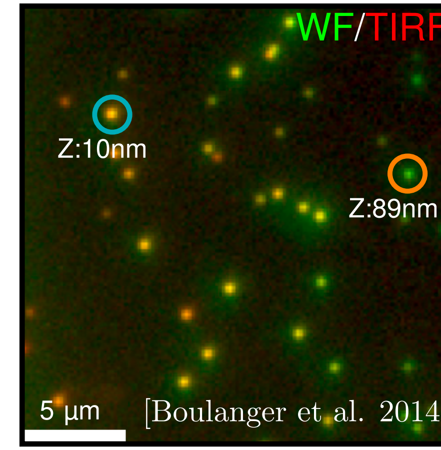


Neural spikes (1D)

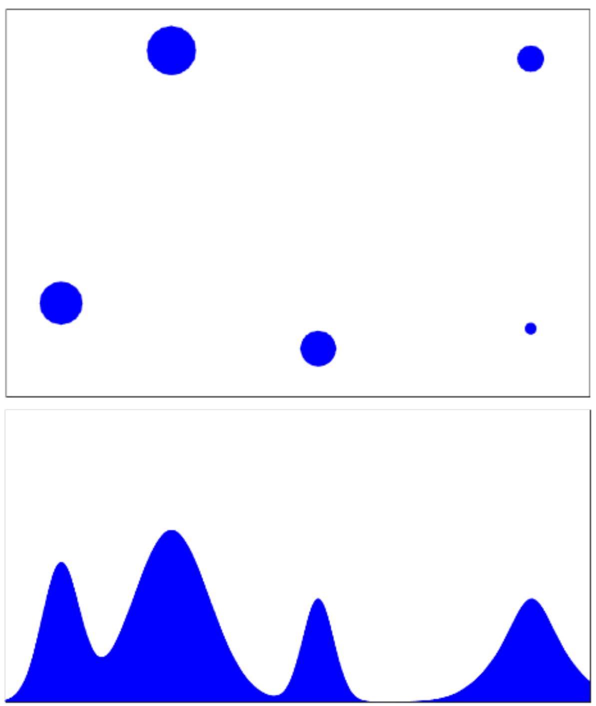


Seismic imaging (1.5D)

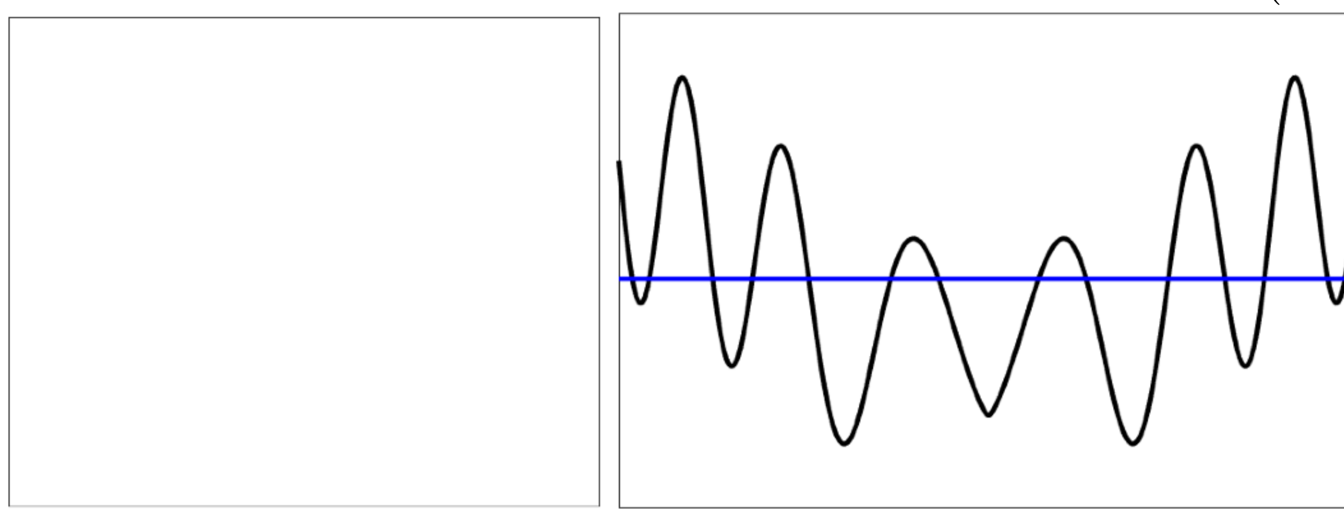
Astrophysics (2D)



Single-molecule fluorescence (3-D)



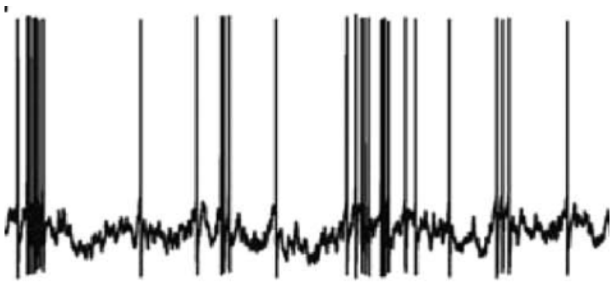
Mixture estimation

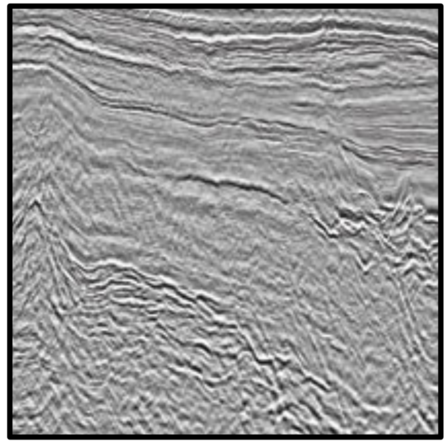


Perceptron with 1 hidden layer

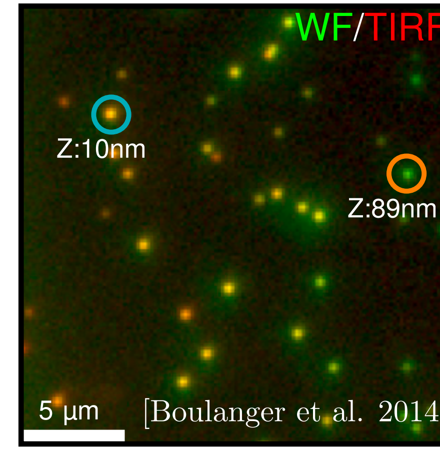
Sparse Estimation

Recover pointwise sources from noisy low-resolution observations.

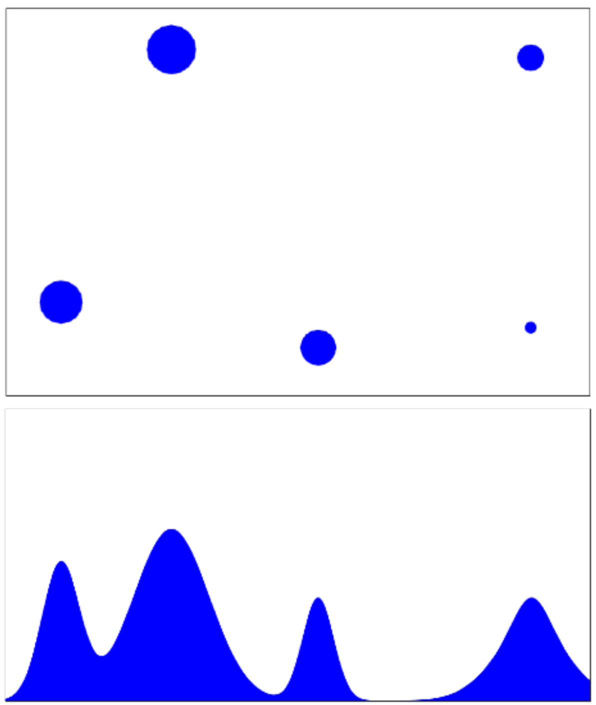

Neural spikes (1D)

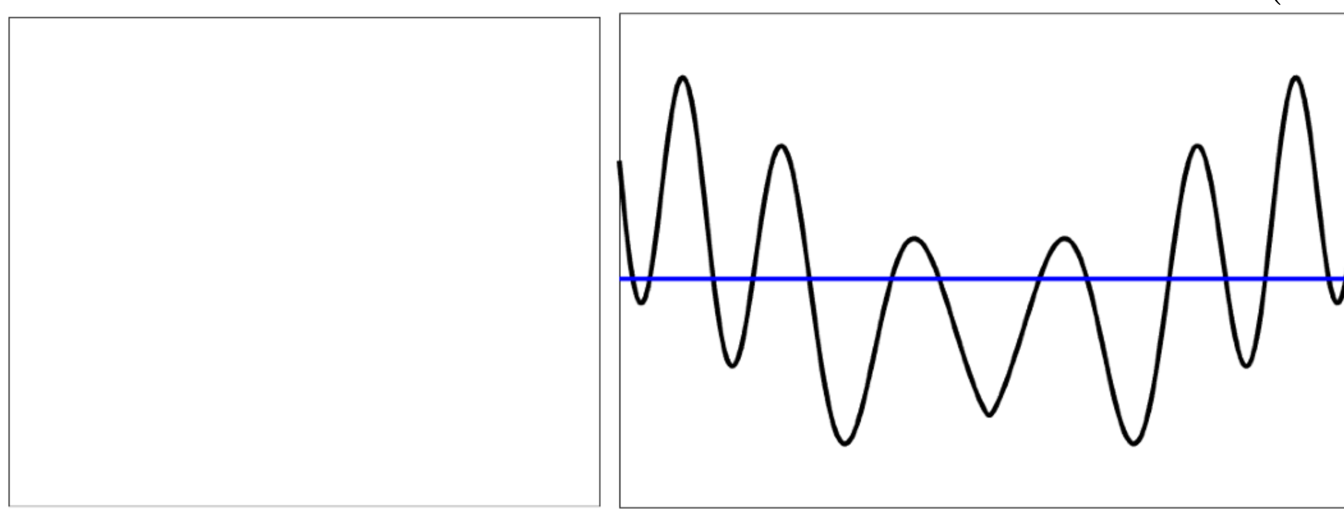

Seismic imaging (1.5D)


Astrophysics (2D)



Single-molecule fluorescence (3-D)


Mixture estimation


Perceptron with 1 hidden layer

Theory: Rayleigh limit?

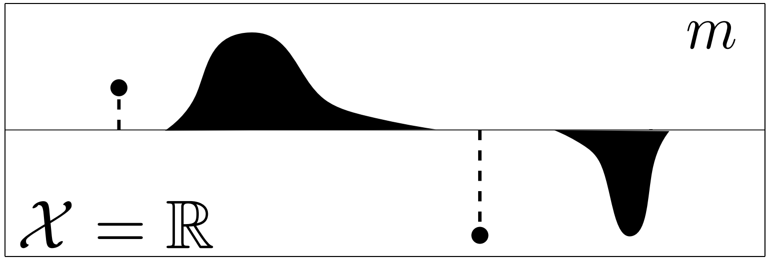
Practice: Scalable algorithms?

Overview

- **Sparse Linear Models**
- Sparse Estimation with Blasso
- Non-super-resolution Regime
- Super-resolution Regime

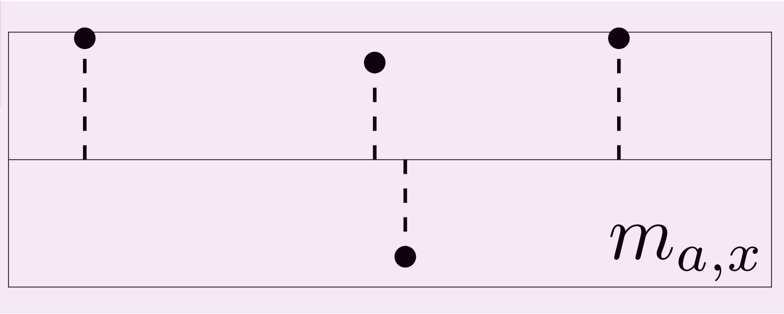
Sparse Linear Models

Measure m on $\mathcal{X}$ (e.g. $\subset \mathbb{R}^d$ or $\mathbb{R}/\mathbb{Z})^d$).



Discrete measure:

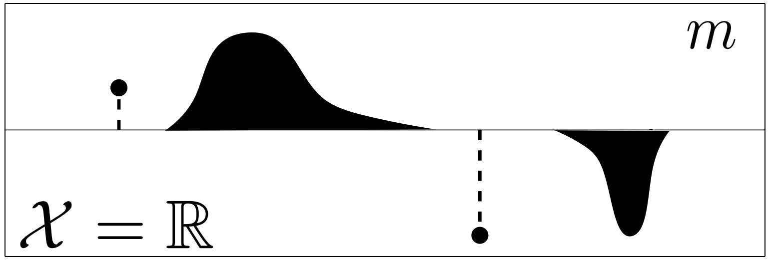
$$m_{a,x} = \sum_{i=1}^n a_i \delta_{x_i}, \quad a_i \in \mathbb{R}, x_i \in \mathcal{X}.$$



$\tilde{\varphi}$

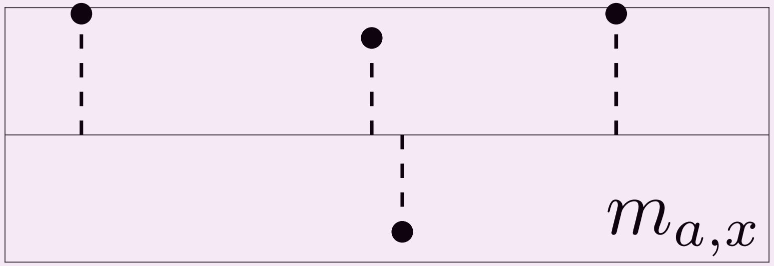
Sparse Linear Models

Measure m on $\mathcal{X}$ (e.g. $\subset \mathbb{R}^d$ or $\mathbb{R}/\mathbb{Z})^d$).



Discrete measure:

$$m_{a,x} = \sum_{i=1}^n a_i \delta_{x_i}, \quad a_i \in \mathbb{R}, x_i \in \mathcal{X}.$$



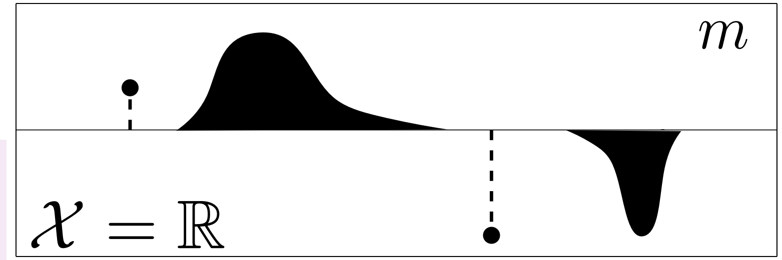
Linear model: $\varphi(x) \in \mathcal{H}$ φ continuous.

$$y = \Phi m \overset{\text{def.}}{=} \int_{\mathcal{X}} \varphi(x) \mathrm{d}m(x)$$

$\tilde{\varphi}$

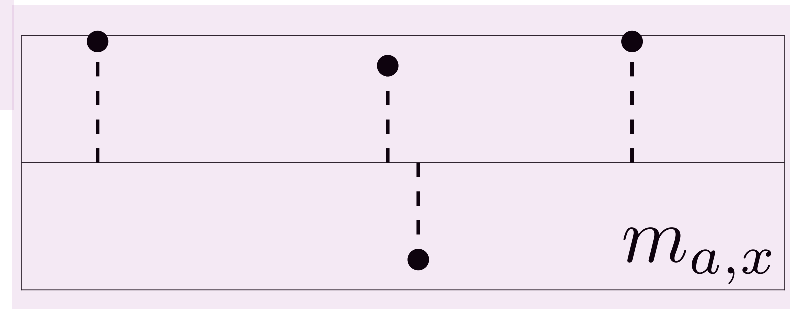
Sparse Linear Models

Measure m on $\mathcal{X}$ (e.g. $\subset \mathbb{R}^d$ or $\mathbb{R}/\mathbb{Z})^d$).



Discrete measure:

$$m_{a,x} = \sum_{i=1}^{n} a_i \delta_{x_i}, \quad a_i \in \mathbb{R}, x_i \in \mathcal{X}.$$

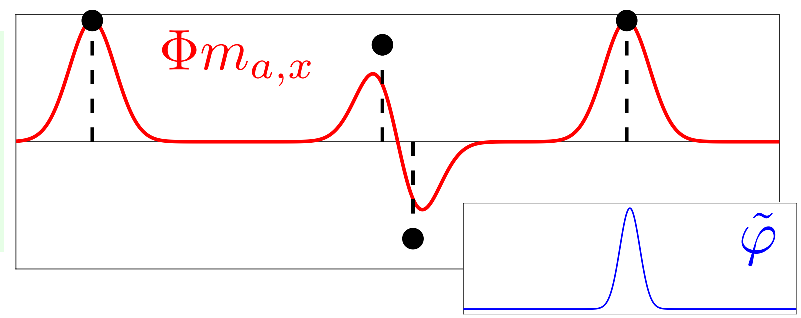


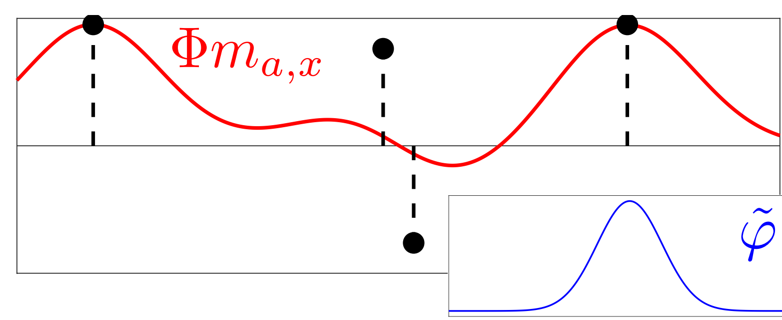
Linear model: $\varphi(x) \in \mathcal{H}$ φ continuous.

$$y = \Phi m \overset{\text{def.}}{=} \int_{\mathcal{X}} \varphi(x) \mathrm{d}m(x)$$

Deconvolution: $\varphi(x) = \tilde{\varphi}(\cdot - x)$

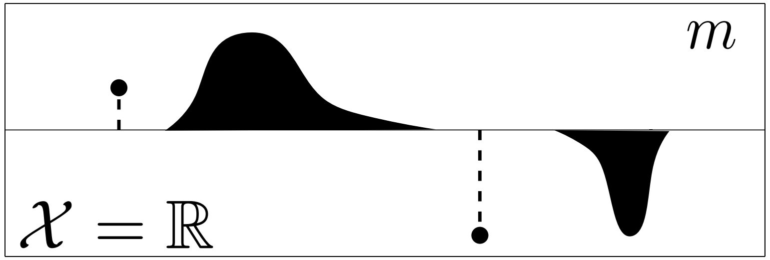
$\rightarrow$ Signal-dependent recovery criteria.





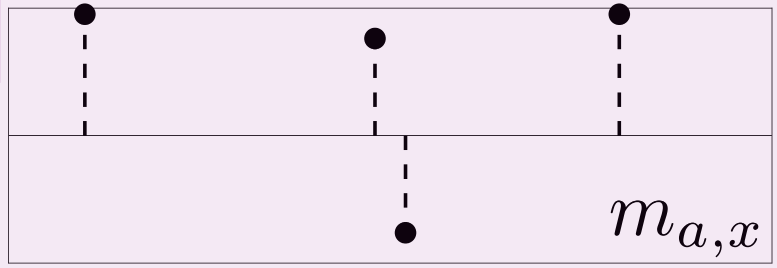
Sparse Linear Models

Measure m on $\mathcal{X}$ (e.g. $\subset \mathbb{R}^d$ or $\mathbb{R}/\mathbb{Z})^d$).



Discrete measure:

$$m_{a,x} = \sum_{i=1}^n a_i \delta_{x_i}, \quad a_i \in \mathbb{R}, x_i \in \mathcal{X}.$$

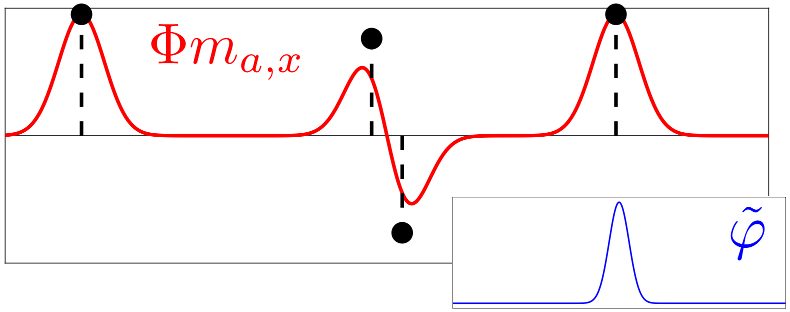


Linear model: $\varphi(x) \in \mathcal{H}$ φ continuous.

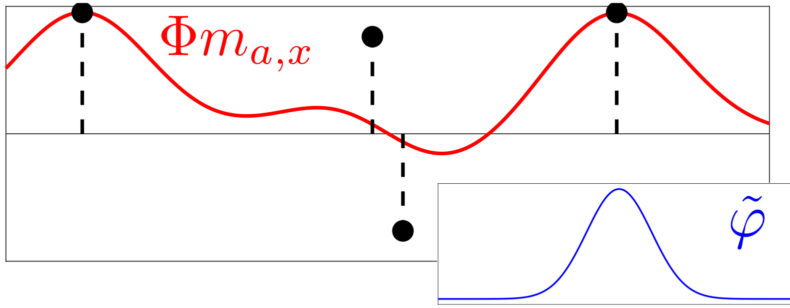
$$y = \Phi m \overset{\text{def.}}{=} \int_{\mathcal{X}} \varphi(x) \mathrm{d}m(x)$$

Deconvolution: $\varphi(x) = \tilde{\varphi}(\cdot - x)$

$\rightarrow$ Signal-dependent recovery criteria.



Fourier: $\varphi(x) = (e^{\mathrm{i}\ell x})_{\ell=-f_c}^{f_c} \in \mathbb{C}^{2f_c+1}$

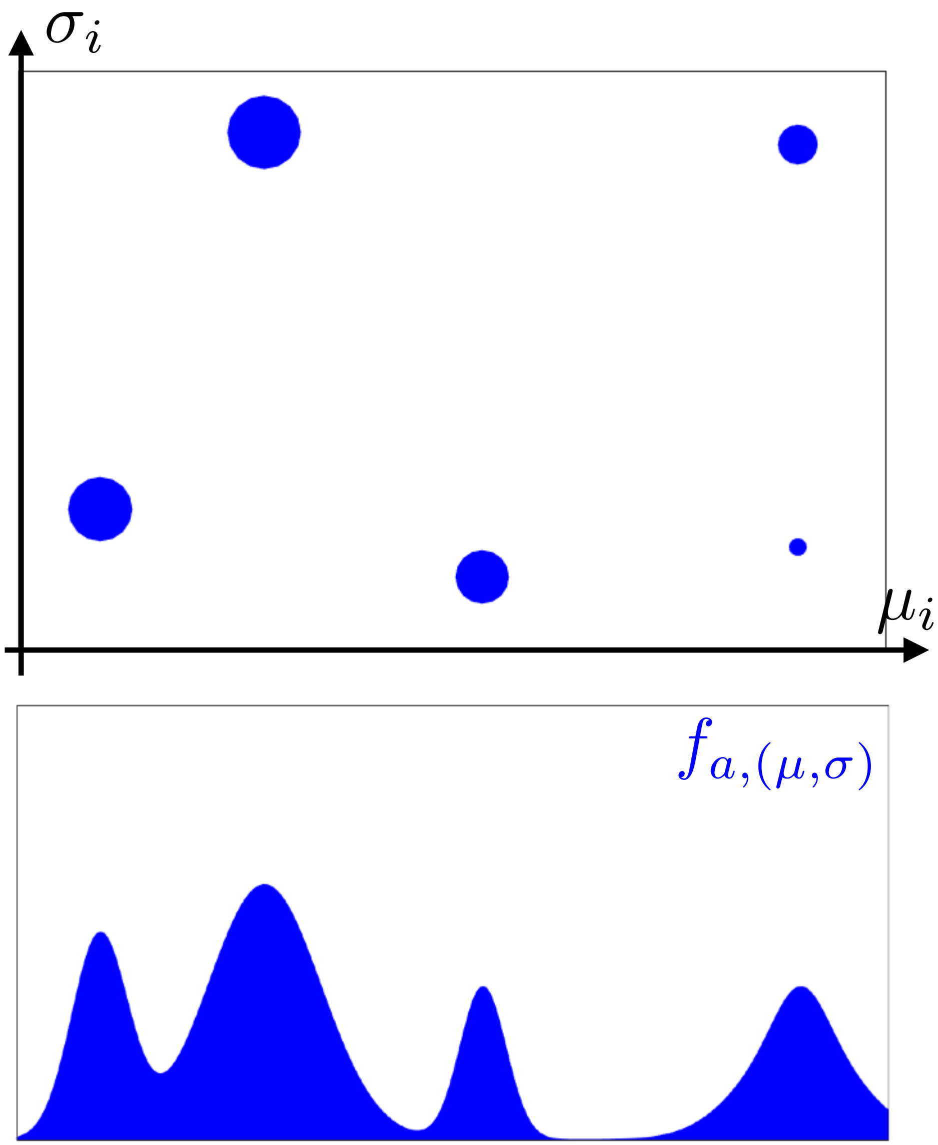


Laplace: $\varphi(x) = e^{-x\cdot} \in \mathcal{H} \overset{\text{def.}}{=} L^2(\mathbb{R}^+)$

Mixture Models

Position/scale mixture in $\mathbb{R}^d$: $(\mu, \sigma) = (\text{mean}, \text{std}) \in \mathcal{X} \overset{\text{def.}}{=} \mathbb{R}^d \times \mathbb{R}^+$:

$$f_{a,(\mu,\sigma)}(t) = \sum_{i=1}^{n} \frac{a_i}{\sigma_i^d} h\left(\frac{t-\mu_i}{\sigma_i}\right)$$



Mixture estimation:

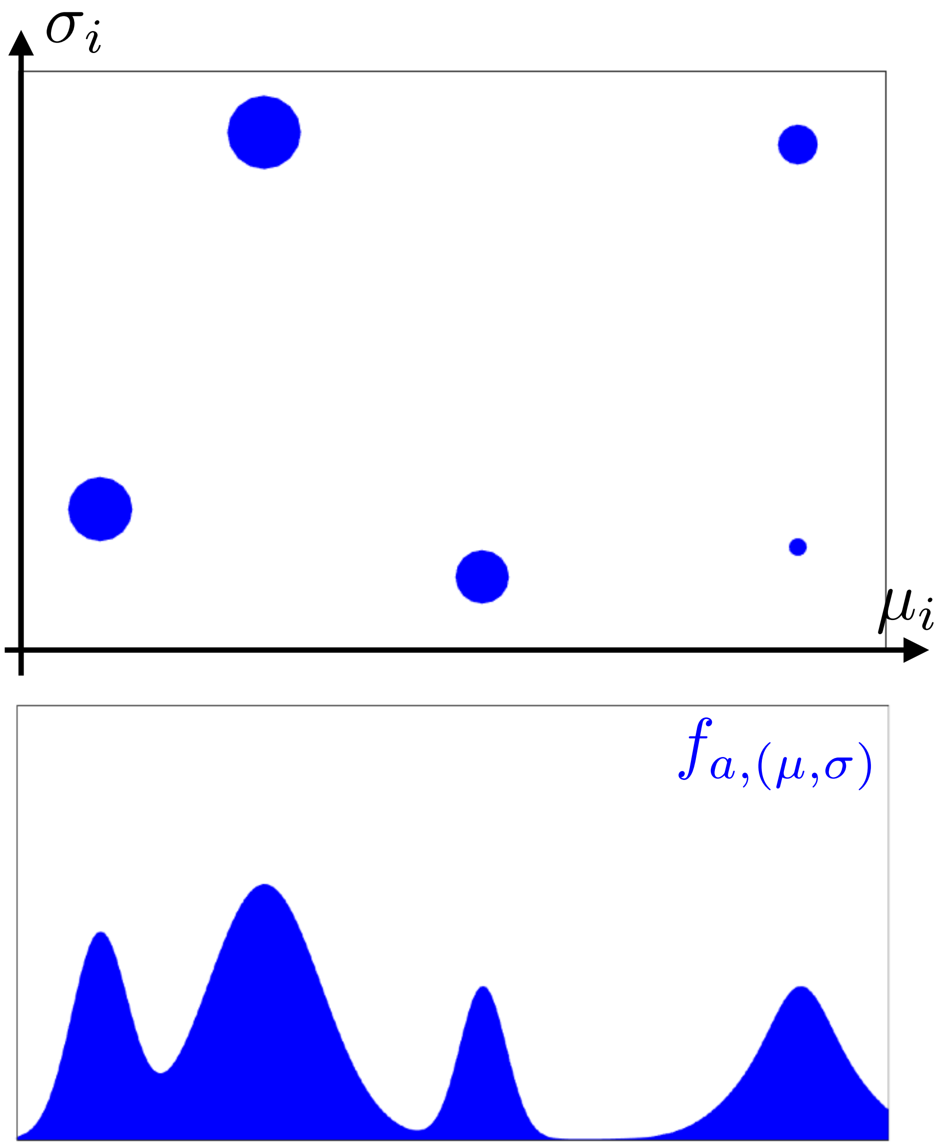
non-convex

$$\min_{a,(\mu,\sigma)} \int |f(t) - f_{a,(\mu,\sigma)}(t)|^2 \mathrm{d}t$$

Mixture Models

Position/scale mixture in $\mathbb{R}^d$: $\quad (\mu, \sigma) = (\text{mean}, \text{std}) \in \mathcal{X} \overset{\text{def.}}{=} \mathbb{R}^d \times \mathbb{R}^+$:

$$f_{a,(\mu,\sigma)}(t) = \sum_{i=1}^{n} \frac{a_i}{\sigma_i^d} h\left(\frac{t-\mu_i}{\sigma_i}\right)$$



Mixture estimation:

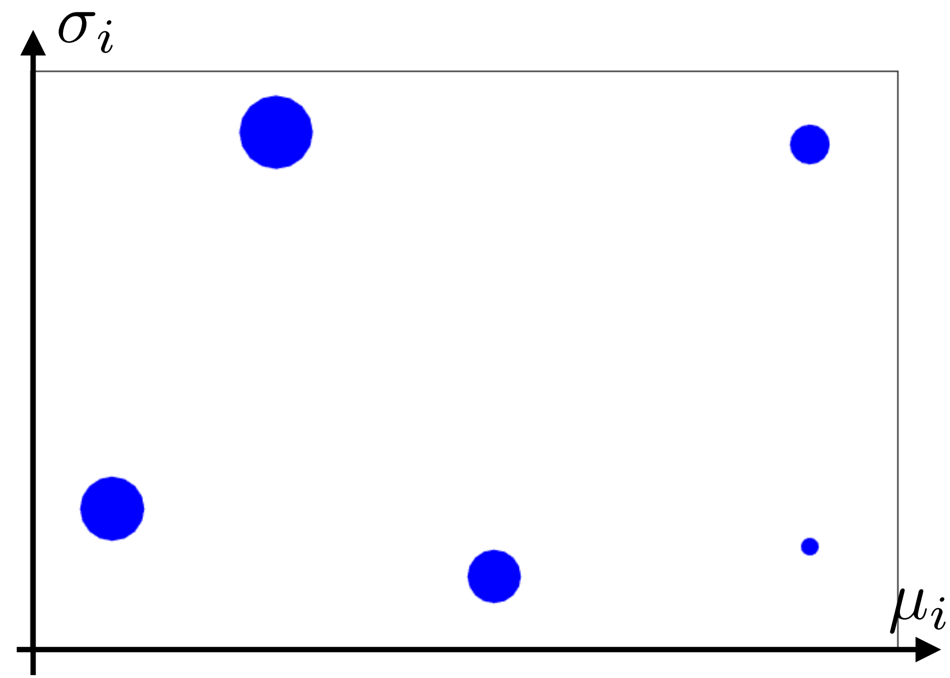
non-convex

$$\min_{a,(\mu,\sigma)} \int |f(t) - f_{a,(\mu,\sigma)}(t)|^2 \mathrm{d}t$$

Mixture Models

Position/scale mixture in $\mathbb{R}^d$: $(\mu, \sigma) = (\text{mean}, \text{std}) \in \mathcal{X} \overset{\text{def.}}{=} \mathbb{R}^d \times \mathbb{R}^+$:

$$f_{a,(\mu,\sigma)}(t) = \sum_{i=1}^{n} \frac{a_i}{\sigma_i^d} h\left(\frac{t-\mu_i}{\sigma_i}\right)$$

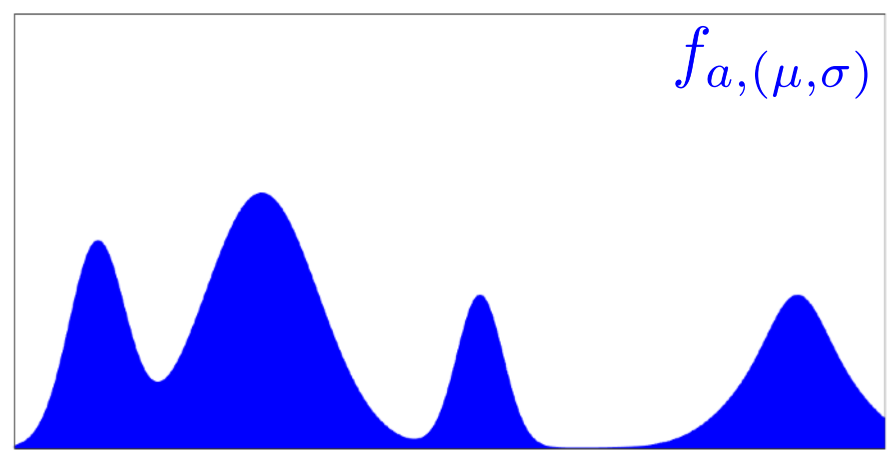


Mixture estimation:

non-convex $\min_{a,(\mu,\sigma)} \int |f(t) - f_{a,(\mu,\sigma)}(t)|^2 \mathrm{d}t$

$m_{a,(\mu,\sigma)} \rightarrow m$

convex $\min_{m \in \mathcal{M}_+^1(\mathcal{X})} \|f - \Phi m\|^2$

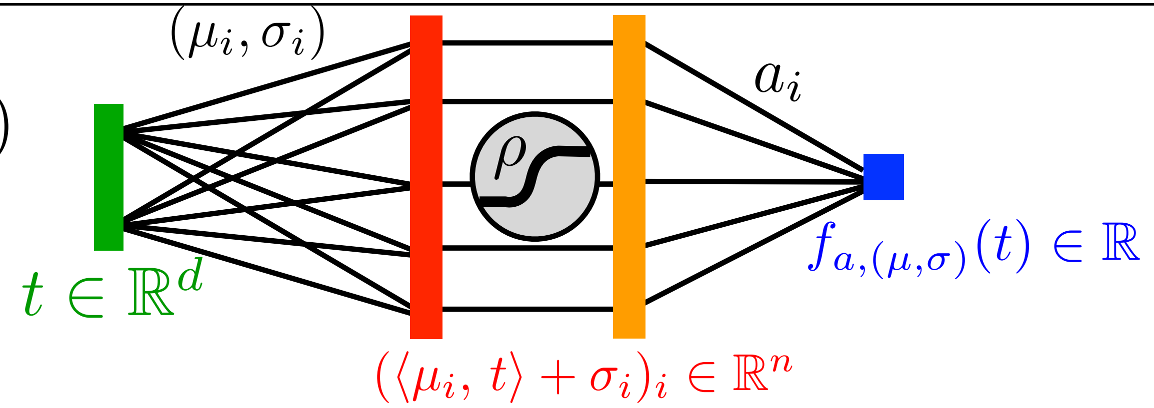


$$f_{a,(\mu,\sigma)} = \Phi m_{a,(\mu,\sigma)} \overset{\text{def.}}{=} \int_{\mathbb{R}^d} \varphi(x) \mathrm{d}m_{a,(\mu,\sigma)}(x)$$

where $\varphi(x) = \frac{1}{x_2^d} h\left(\frac{\cdot - x_1}{x_2}\right)$

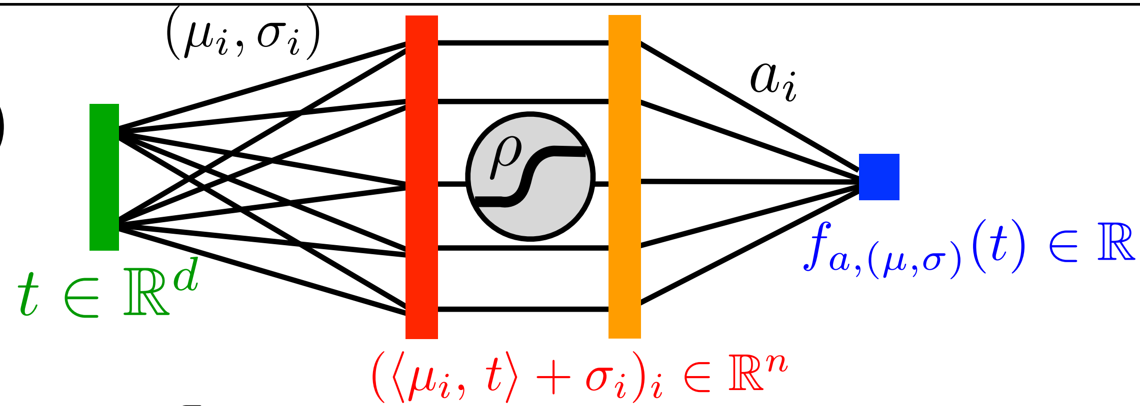
Multilayer Perceptron

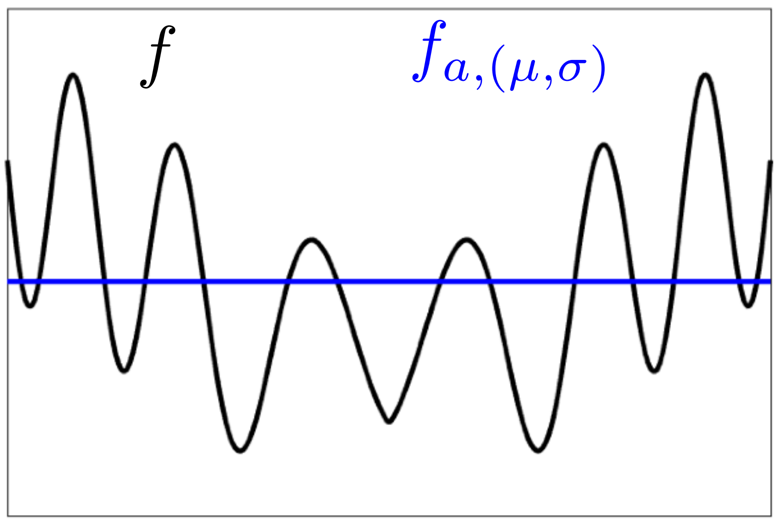
$$f_{a,(m,s)}(t) = \sum_{i=1}^{n} a_i \rho(\langle \mu_i,\, t\rangle + \sigma_i)$$

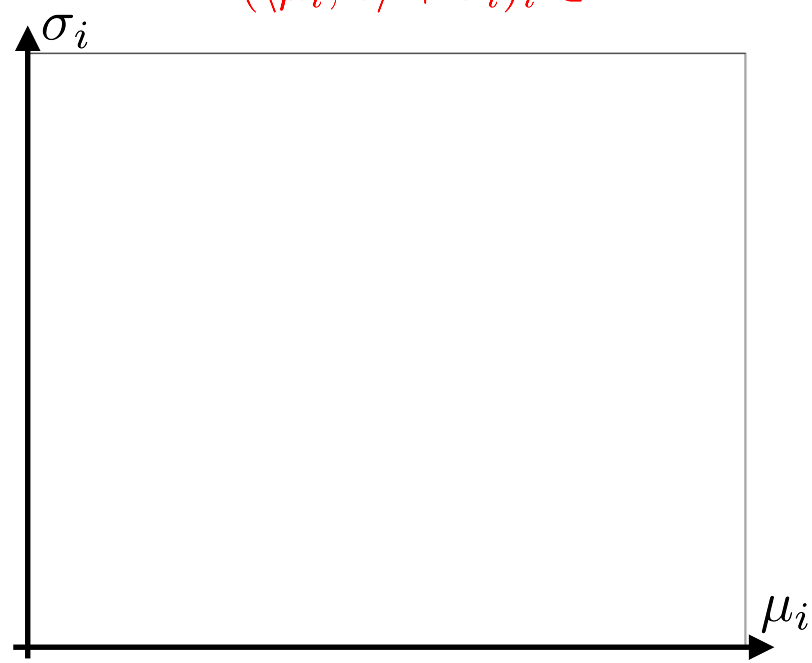


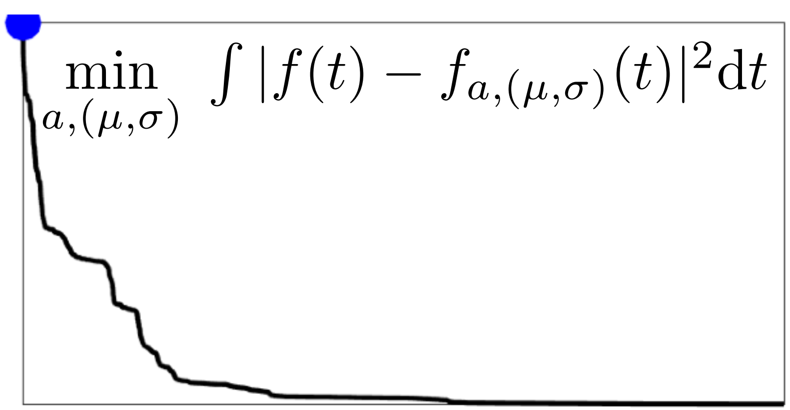
Multilayer Perceptron

$$f_{a,(m,s)}(t) = \sum_{i=1}^{n} a_i \rho(\langle \mu_i,\, t\rangle + \sigma_i)$$



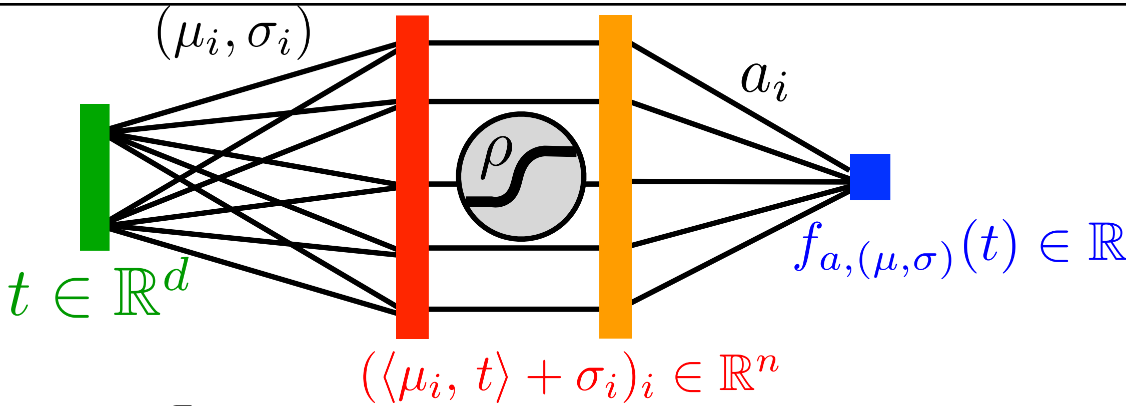


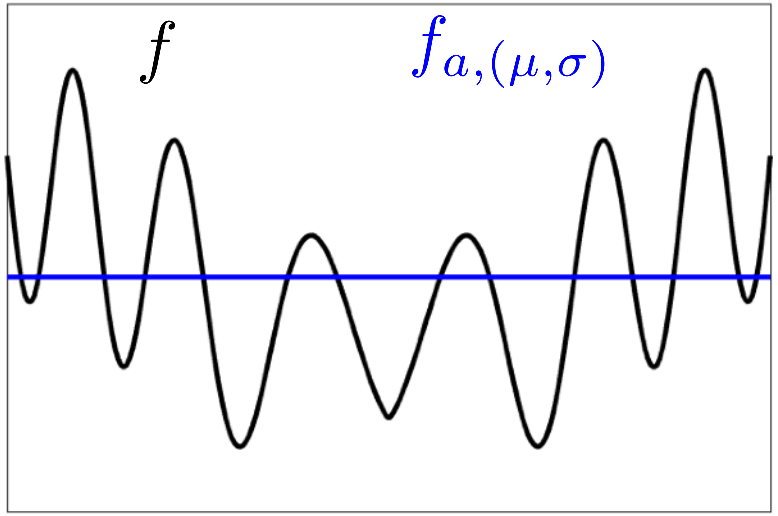


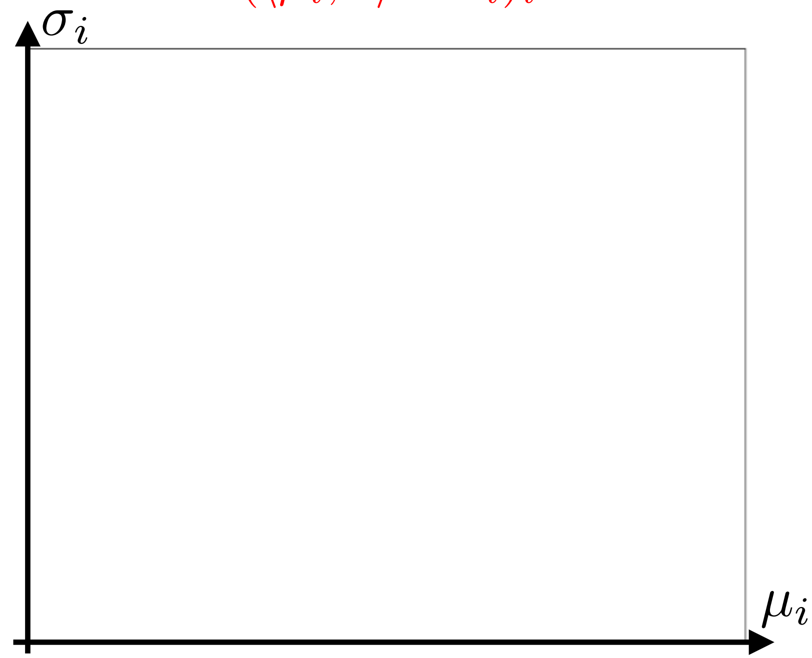


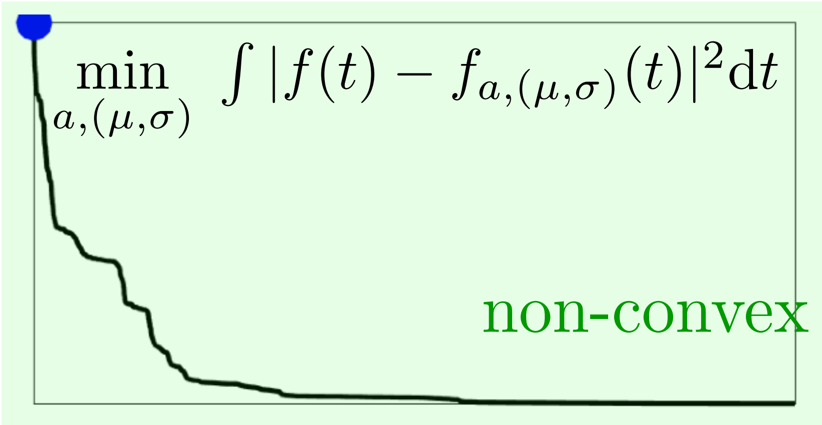
Multilayer Perceptron

$$f_{a,(m,s)}(t) = \sum_{i=1}^{n} a_i \rho(\langle \mu_i,\, t\rangle + \sigma_i)$$









$m_{a,w} \to m$

convex $\quad \min_{m \in \mathcal{M}(\mathcal{X})} \|f - \Phi m\|^2$

$$f_{a,(\mu,\sigma)} = \Phi m_{a,(\mu,\sigma)} \overset{\text{def.}}{=} \int_{\mathbb{R}^d} \varphi(x) \mathrm{d}m_{a,(\mu,\sigma)}(x) \quad \text{where} \quad \varphi(x) = \rho(\langle x_1, \cdot\rangle + x_2)$$

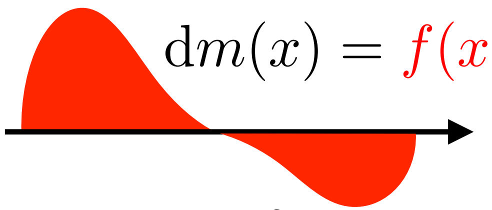
Overview

- Sparse Linear Models
- **Sparse Estimation with Blasso**
- Non-super-resolution Regime
- Super-resolution Regime

Off-the-grid Sparse Regularization

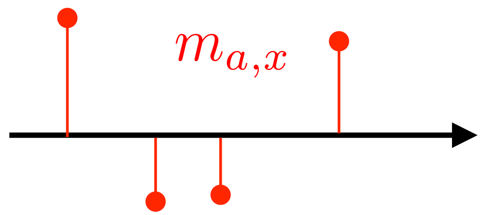
Grid-free regularization: total variation of measures:

$$|m|(\mathcal{X}) \overset{\text{def.}}{=} \sup \left\{ \int_{\mathcal{X}} \eta \mathrm{d}m \; ; \; f \in \mathcal{C}(\mathcal{X}), \|f\|_\infty \leqslant 1 \right\}$$



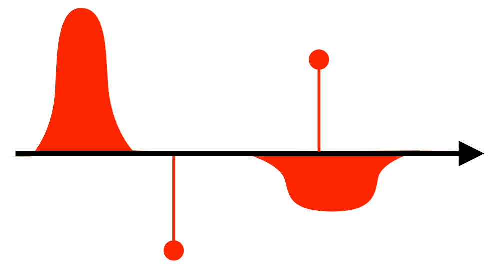
$$\mathrm{d}m(x) = f(x)\mathrm{d}x$$

$$|m|(\mathcal{X}) = \int |f| = \|f\|_{L^1}$$



$$m_{a,x}$$

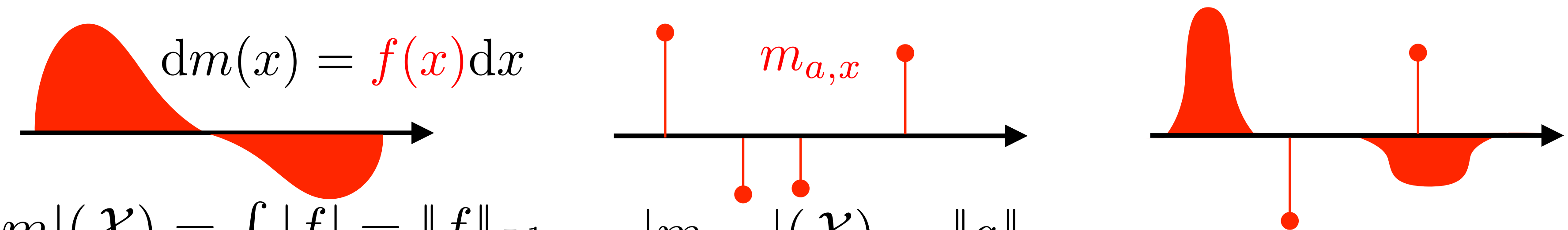
$$|m_{a,x}|(\mathcal{X}) = \|a\|_{\ell^1}$$



Off-the-grid Sparse Regularization

Grid-free regularization: total variation of measures:

$$|m|(\mathcal{X}) \overset{\text{def.}}{=} \sup \left\{ \int_{\mathcal{X}} \eta \mathrm{d}m \; ; \; f \in \mathcal{C}(\mathcal{X}), \|f\|_\infty \leqslant 1 \right\}$$



$|m|(\mathcal{X}) = \int |f| = \|f\|_{L^1}$ $\quad$ $|m_{a,x}|(\mathcal{X}) = \|a\|_{\ell^1}$

Beurling Lasso:
[De Castro, Gamboa, 2012]

$$\min_m \frac{1}{2}\|\Phi m - y\|^2 + \lambda |m|(\mathcal{X}) \qquad (\mathcal{P}_\lambda(y))$$

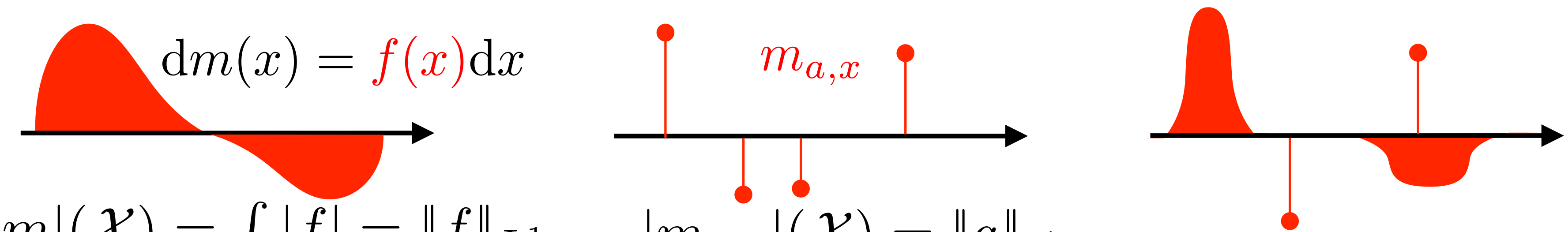
$$\min_m \; \{ |m|(\mathcal{X}) \; ; \; \Phi m = y \} \qquad (\mathcal{P}_0(y))$$

$\lambda \to 0^+$

Off-the-grid Sparse Regularization

Grid-free regularization: total variation of measures:

$$|m|(\mathcal{X}) \overset{\text{def.}}{=} \sup \left\{ \int_{\mathcal{X}} \eta \mathrm{d}m \;;\; f \in \mathcal{C}(\mathcal{X}), \|f\|_\infty \leqslant 1 \right\}$$



$|m|(\mathcal{X}) = \int |f| = \|f\|_{L^1}$ $\quad$ $|m_{a,x}|(\mathcal{X}) = \|a\|_{\ell^1}$

Beurling Lasso:
[De Castro, Gamboa, 2012]

$$\min_m \frac{1}{2}\|\Phi m - y\|^2 + \lambda |m|(\mathcal{X}) \qquad (\mathcal{P}_\lambda(y))$$

$\lambda \to 0^+$

$$\min_m \; \{|m|(\mathcal{X}) \;;\; \Phi m = y\} \qquad (\mathcal{P}_0(y))$$

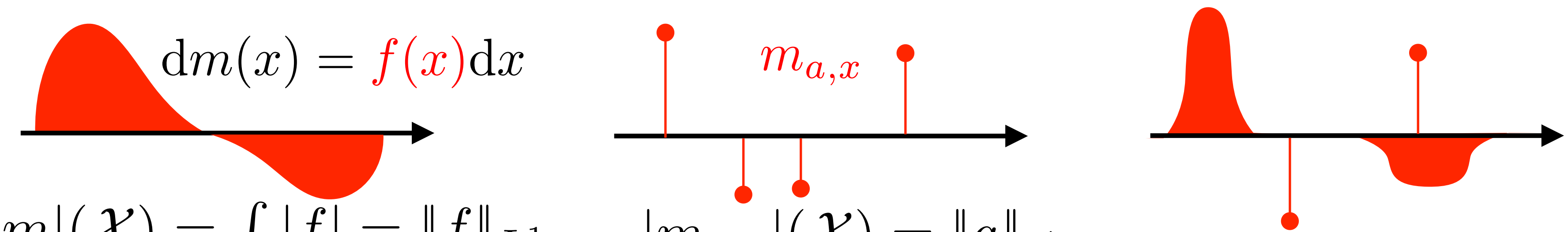
Proposition:
[Fischer Jerome, 1974]

If $\dim(\mathrm{Im}(\Phi)) < +\infty$, $\exists (a, x) \in \mathbb{R}^N \times \mathbb{T}^N$
with $N \leqslant \dim(\mathrm{Im}(\Phi))$
such that $m_{a,x}$ is a solution to $\mathcal{P}_\lambda(y)$.

Off-the-grid Sparse Regularization

Grid-free regularization: total variation of measures:

$$|m|(\mathcal{X}) \overset{\text{def.}}{=} \sup \left\{ \int_{\mathcal{X}} \eta \mathrm{d}m \; ; \; f \in \mathcal{C}(\mathcal{X}), \|f\|_\infty \leqslant 1 \right\}$$



$|m|(\mathcal{X}) = \int |f| = \|f\|_{L^1}$ $\quad$ $|m_{a,x}|(\mathcal{X}) = \|a\|_{\ell^1}$

Beurling Lasso:
[De Castro, Gamboa, 2012]

$$\min_m \frac{1}{2}\|\Phi m - y\|^2 + \lambda |m|(\mathcal{X}) \qquad (\mathcal{P}_\lambda(y))$$

$$\min_m \{|m|(\mathcal{X}) \; ; \; \Phi m = y\} \qquad (\mathcal{P}_0(y))$$

$\lambda \to 0^+$

Proposition:
[Fischer Jerome, 1974]

If $\dim(\mathrm{Im}(\Phi)) < +\infty$, $\exists (a, x) \in \mathbb{R}^N \times \mathbb{T}^N$
with $N \leqslant \dim(\mathrm{Im}(\Phi))$
such that $m_{a,x}$ is a solution to $\mathcal{P}_\lambda(y)$.

Questions: is m_0 solution of $\mathcal{P}_0(\Phi m_0)$? Robustness to noise?

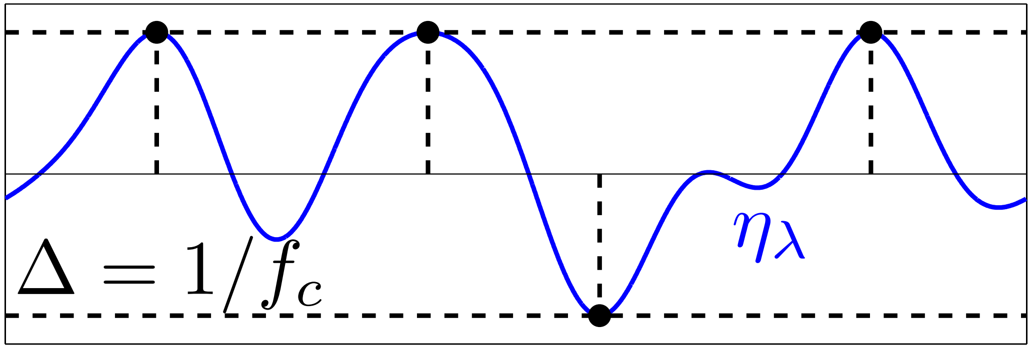
Limit Certificate

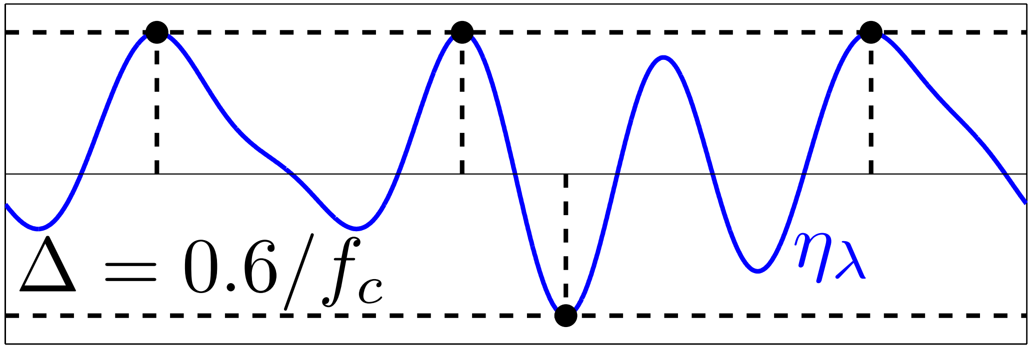
$\mathcal{P}_\lambda(y)$ $$\min_m |m|(\mathcal{X}) + \frac{1}{2\lambda}\|\Phi m - y\|^2$$

$$\eta_\lambda \overset{\text{def.}}{=} \frac{1}{\lambda}\Phi^*(y - \Phi m_\lambda)$$

Proposition: m_λ solves $(\mathcal{P}_\lambda(y)) \Leftrightarrow \eta_\lambda \in \partial|m_\lambda|(\mathcal{X})$

(if $m_\lambda = m_{a,x}$) $\Leftrightarrow |\eta_\lambda| \leqslant 1$ and $\eta_\lambda(x_i) = \text{sign}(a_i)$





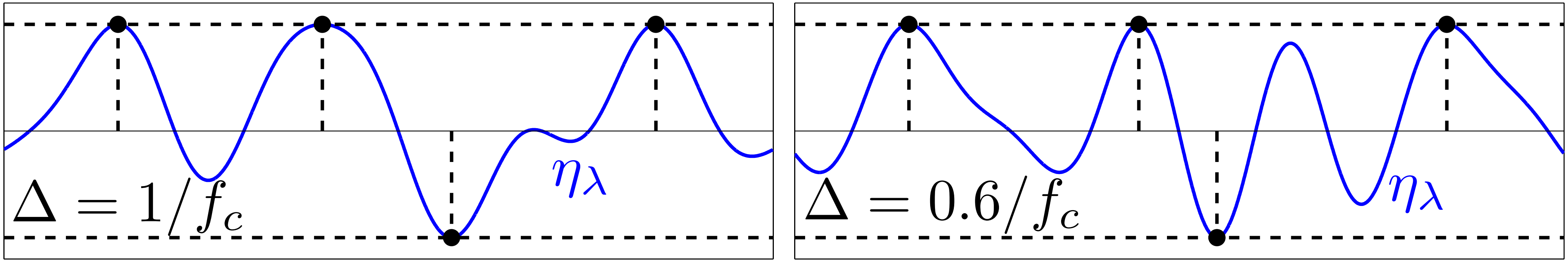
Limit Certificate

$\mathcal{P}_\lambda(y)$: $$\min_m |m|(\mathcal{X}) + \frac{1}{2\lambda}\|\Phi m - y\|^2$$

$$\eta_\lambda \overset{\text{def.}}{=} \frac{1}{\lambda}\Phi^*(y - \Phi m_\lambda)$$

Proposition: m_λ solves $(\mathcal{P}_\lambda(y)) \Leftrightarrow \eta_\lambda \in \partial|m_\lambda|(\mathcal{X})$

(if $m_\lambda = m_{a,x}$) $\Leftrightarrow |\eta_\lambda| \leqslant 1$ and $\eta_\lambda(x_i) = \text{sign}(a_i)$



Theorem: If $(\lambda, \|w\|/\lambda) \to 0$, then $m_\lambda \to m_0 = m_{a,x}$, $\eta_\lambda \to \eta_0$.

$$\eta_0 \overset{\text{def.}}{=} \underset{\eta=\Phi^* p}{\text{argmin}} \|p\| \text{ s.t. } \begin{cases} \forall\, i,\ \eta(x_i) = \text{sign}(a_i), \\ \|\eta\|_\infty \leqslant 1. \end{cases}$$

Vanishing Derivative Pre-certificate

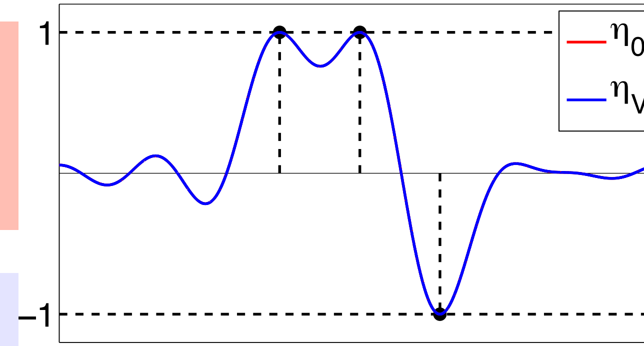
Input measure: $m_0 = m_{a,x}$.

$$\eta_0 \overset{\text{def.}}{=} \underset{\eta=\Phi^* p}{\text{argmin}} \|p\| \text{ s.t. } \begin{cases} \forall\, i,\ \eta(x_i) = \text{sign}(a_i), \\ \|\eta\|_\infty \leqslant 1. \end{cases}$$

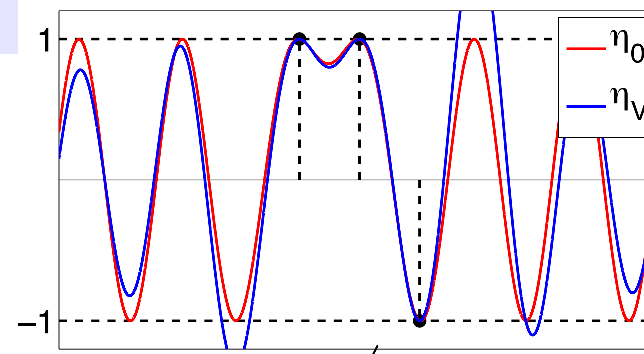
$$\eta_V \overset{\text{def.}}{=} \underset{\eta=\Phi^* p}{\text{argmin}} \|p\| \text{ s.t. } \begin{cases} \forall\, i,\ \eta(x_i) = \text{sign}(a_i), \\ \forall\, i,\ \eta'(x_i) = 0. \end{cases}$$

Proposition: $\eta_V = \Phi^* A_x^+(\text{sign}(a); 0)$

where $A_x(b) = \sum_i b_i^1 \varphi(x_i, \cdot) + b_i^2 \varphi'(x_i, \cdot)$



$\eta_0 = \eta_V$



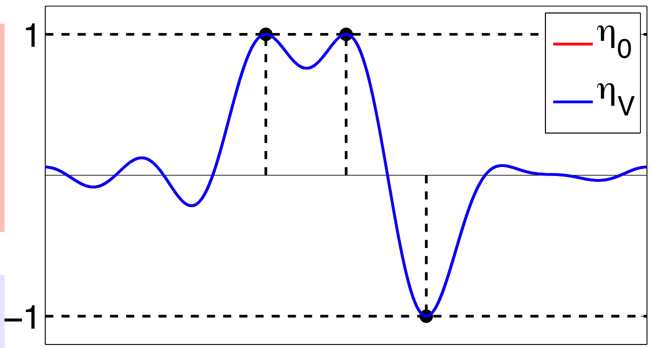
$\eta_0 \neq \eta_V$

Vanishing Derivative Pre-certificate

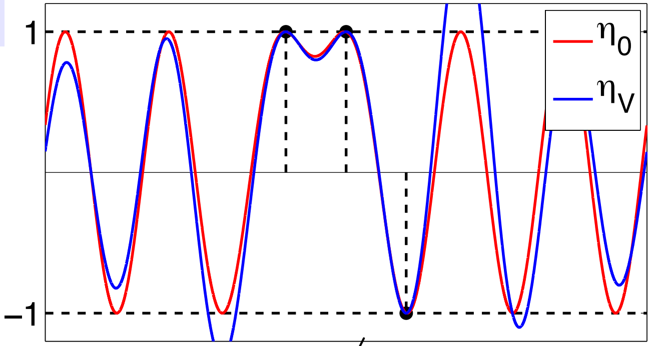
Input measure: $m_0 = m_{a,x}$.

$$\eta_0 \overset{\text{def.}}{=} \underset{\eta=\Phi^* p}{\text{argmin}} \|p\| \text{ s.t. } \begin{cases} \forall\, i,\ \eta(x_i) = \text{sign}(a_i), \\ \|\eta\|_\infty \leqslant 1. \end{cases}$$

$$\eta_V \overset{\text{def.}}{=} \underset{\eta=\Phi^* p}{\text{argmin}} \|p\| \text{ s.t. } \begin{cases} \forall\, i,\ \eta(x_i) = \text{sign}(a_i), \\ \forall\, i,\ \eta'(x_i) = 0. \end{cases}$$



$\eta_0 = \eta_V$



$\eta_0 \neq \eta_V$

Proposition: $\eta_V = \Phi^* A_x^+(\text{sign}(a); 0)$

where $A_x(b) = \sum_i b_i^1 \varphi(x_i, \cdot) + b_i^2 \varphi'(x_i, \cdot)$

Non-degenerate certificate: $\eta \in \text{ND}(m_{a,x})$:

$\iff \forall\, t \notin \{x_1, \ldots, x_N\}, |\eta(t)| < 1$ and $\forall\, i, \eta''(x_i) \neq 0$

Theorem: $\eta_V \in \text{ND}(m_0) \implies \eta_V = \eta_0$

Support Stability Theorem

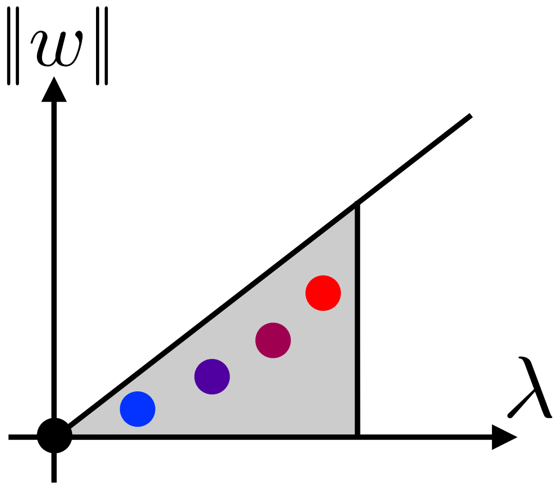
[Duval, Peyré 2014]

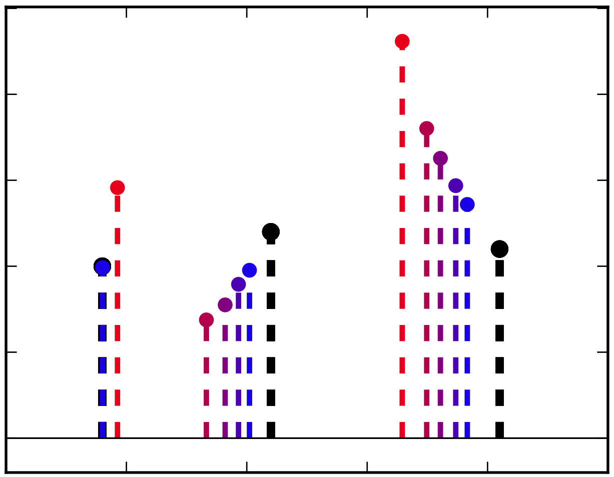
Theorem: If $\eta_V \in \mathrm{ND}(m_0)$ for $m_0 = m_{a,x}$, then

$$\text{for } (\|w\|/\lambda, \lambda) = O(1),$$

the solution of $\mathcal{P}_\lambda(y)$ for $y = \Phi(m_0) + w$ is

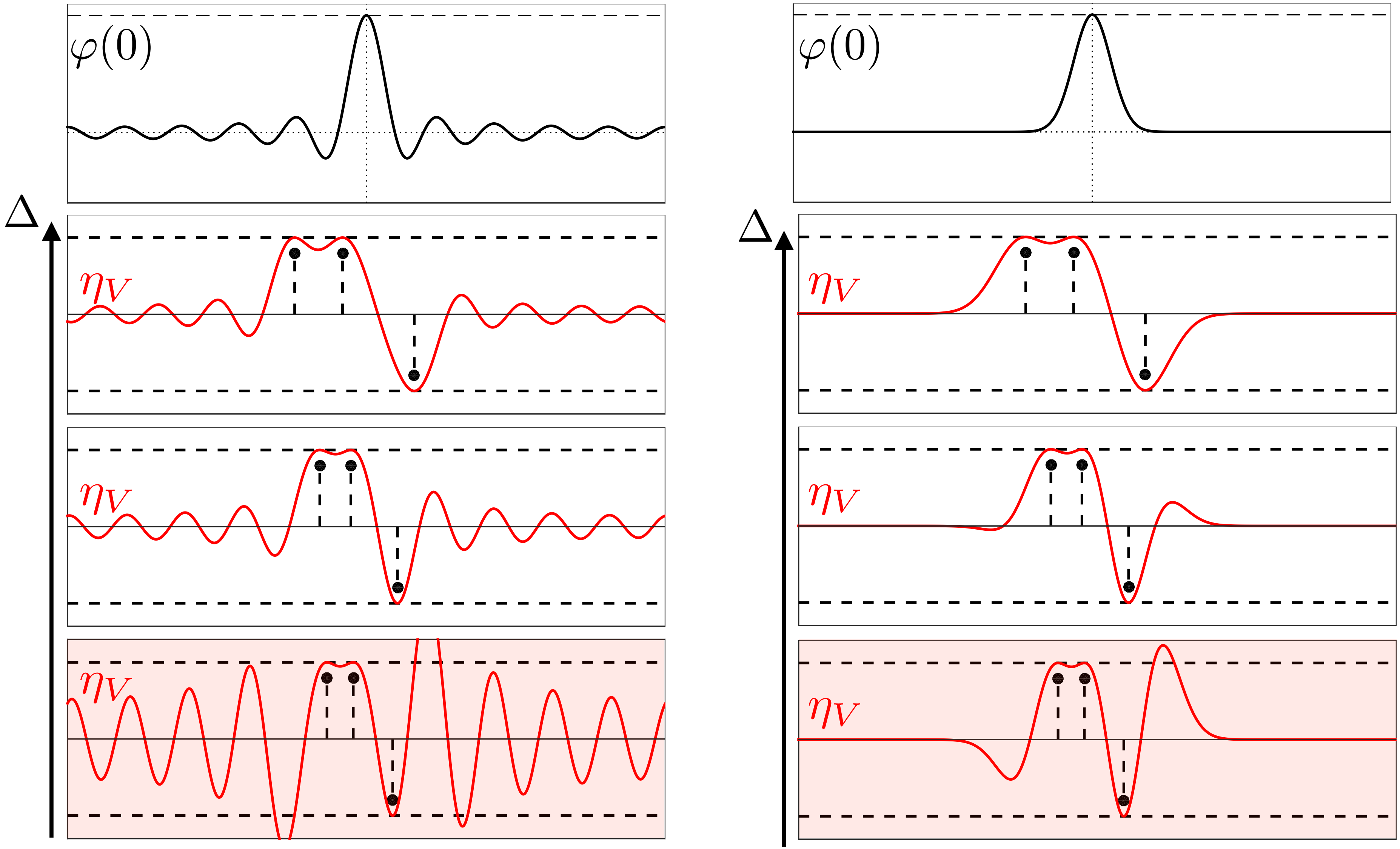
$$m_\lambda = \sum_{i=1}^{N} a_i^\star \delta_{x_i^\star} \quad \text{where} \quad \|(x,a) - (x^\star, a^\star)\| = O(\|w\|)$$





When is η_V Non-degenerate ?

Input measure: $m_0 = m_{a,\Delta x}$, $\Delta \to 0$



Intuition: for **signed** measures, spikes should be "separated".

Overview

- Sparse Linear Models
- Sparse Estimation with Blasso
- **Non-super-resolution Regime**
- Super-resolution Regime

Fisher-Rao Distance

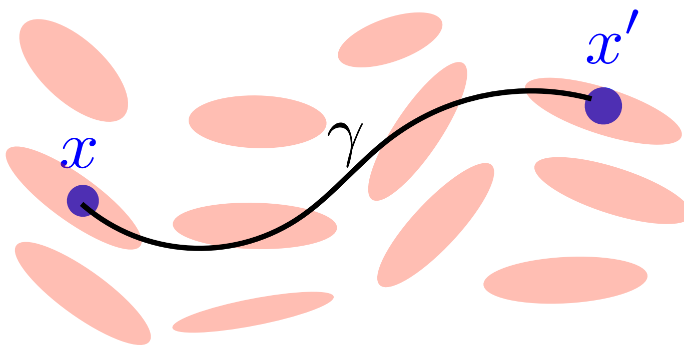
Fisher metric: $H_x \stackrel{\text{def.}}{=} [\partial\varphi(x)][\partial\varphi(x)]^\top \in \mathbb{R}^{d\times d}$ $\quad (\mathcal{X} \subset \mathbb{R}^d)$

Fisher-Rao Distance

Fisher metric: $H_x \overset{\text{def.}}{=} [\partial\varphi(x)][\partial\varphi(x)]^\top \in \mathbb{R}^{d\times d}$ $\quad (\mathcal{X} \subset \mathbb{R}^d)$

Fisher-Rao geodesic distance:

$$d_{\text{FR}}(x,x')^2 \overset{\text{def.}}{=} \inf_{\gamma : x \to x'} \int_0^1 \langle H_{\gamma(t)}\gamma'(t),\, \gamma'(t)\rangle \mathrm{d}t$$



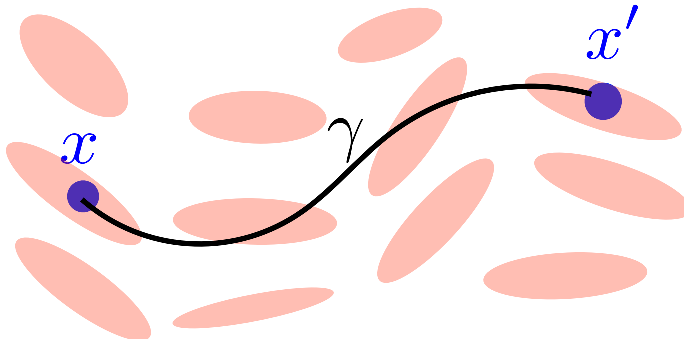
Statistical interpretation: $(|\varphi(x)_i|^2/\|\varphi(x)\|^2)_i \sim$ probability distribution.

Fisher-Rao Distance

Fisher metric: $H_x \overset{\text{def.}}{=} [\partial\varphi(x)][\partial\varphi(x)]^\top \in \mathbb{R}^{d\times d} \qquad (\mathcal{X} \subset \mathbb{R}^d)$

Fisher-Rao geodesic distance:

$$d_{\text{FR}}(x,x')^2 \overset{\text{def.}}{=} \inf_{\gamma:x\to x'} \int_0^1 \langle H_{\gamma(t)}\gamma'(t),\, \gamma'(t)\rangle \mathrm{d}t$$



Statistical interpretation: $(|\varphi(x)_i|^2/\|\varphi(x)\|^2)_i \sim$ probability distribution.

Deconvolution: $\mathcal{X} = \mathbb{R}^d$

$\varphi(x) = h((\cdot - x)/s)$

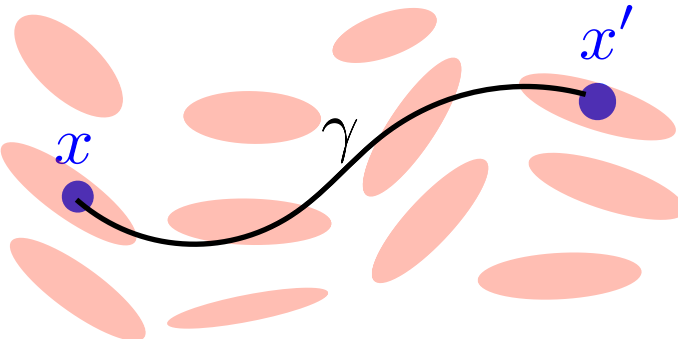
$$d_{\text{FR}}(x,x') = \sigma^{-1}\|x - x'\|$$

Fisher-Rao Distance

Fisher metric: $H_x \overset{\text{def.}}{=} [\partial\varphi(x)][\partial\varphi(x)]^\top \in \mathbb{R}^{d\times d}$ $\quad (\mathcal{X} \subset \mathbb{R}^d)$

Fisher-Rao geodesic distance:

$$d_{\mathrm{FR}}(x,x')^2 \overset{\text{def.}}{=} \inf_{\gamma: x\to x'} \int_0^1 \langle H_{\gamma(t)}\gamma'(t), \gamma'(t)\rangle \mathrm{d}t$$



Statistical interpretation: $(|\varphi(x)_i|^2/\|\varphi(x)\|^2)_i \sim$ probability distribution.

Deconvolution: $\mathcal{X} = \mathbb{R}^d$

$\varphi(x) = h((\cdot - x)/s)$

$$d_{\mathrm{FR}}(x,x') = \sigma^{-1}\|x - x'\|$$

Fourier coefficients: $\mathcal{X} = (\mathbb{R}/\mathbb{Z})^d$

$\varphi(x) = (e^{2\mathrm{i}\pi kx})_{k=-f_c}^{f_c}$

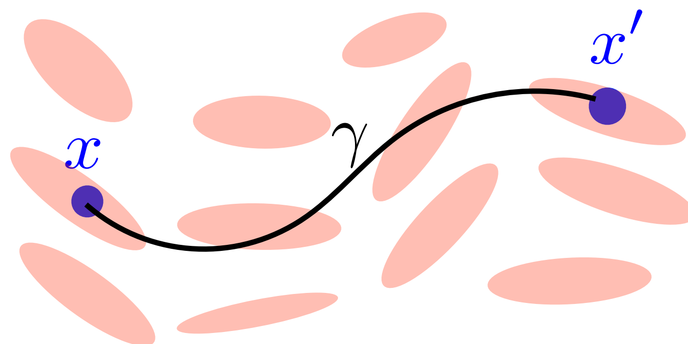
$$d_{\mathrm{FR}}(x,x') = f_c\|x - x'\|$$

Fisher-Rao Distance

Fisher metric: $H_x \stackrel{\text{def.}}{=} [\partial\varphi(x)][\partial\varphi(x)]^\top \in \mathbb{R}^{d\times d} \qquad (\mathcal{X} \subset \mathbb{R}^d)$

Fisher-Rao geodesic distance:

$$d_{\text{FR}}(x,x')^2 \stackrel{\text{def.}}{=} \inf_{\gamma:x\to x'} \int_0^1 \langle H_{\gamma(t)}\gamma'(t),\, \gamma'(t)\rangle \mathrm{d}t$$



Statistical interpretation: $(|\varphi(x)_i|^2/\|\varphi(x)\|^2)_i \sim$ probability distribution.

Deconvolution: $\mathcal{X} = \mathbb{R}^d$
$\varphi(x) = h((\cdot - x)/s)$
$d_{\text{FR}}(x,x') = \sigma^{-1}\|x - x'\|$

Fourier coefficients: $\mathcal{X} = (\mathbb{R}/\mathbb{Z})^d$
$\varphi(x) = (e^{2i\pi kx})_{k=-f_c}^{f_c}$
$d_{\text{FR}}(x,x') = f_c\|x - x'\|$

Laplace transform: $\mathcal{X} = \mathbb{R}_+^*$
$\varphi(x) = e^{-x\cdot}$
$d_{\text{FR}}(x,x') = |\log(x/x')|$

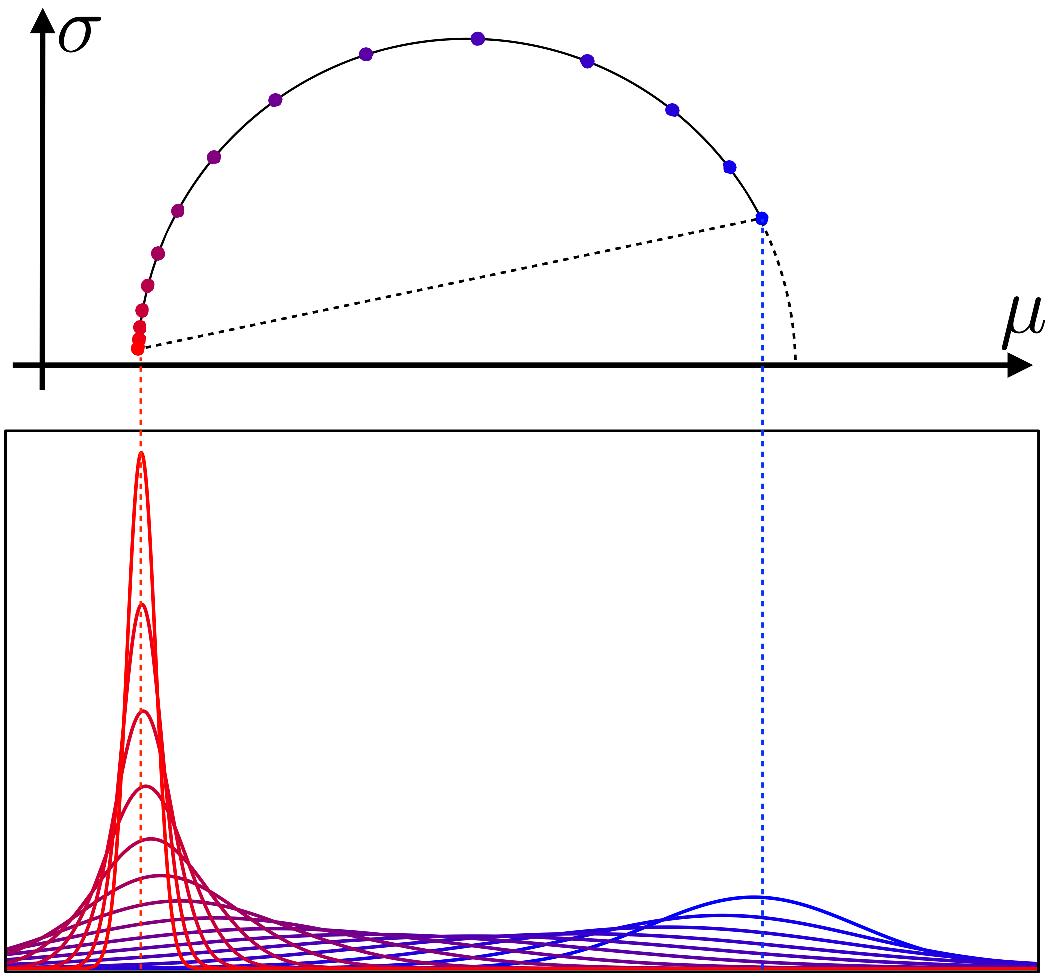
Gaussian Mixtures

$$x = (\mu, \sigma) \in \mathcal{X} = \mathbb{R} \times \mathbb{R}^+$$

$$\varphi(x) = \frac{1}{\sigma} e^{-\frac{(\cdot - \mu)^2}{2\sigma^2}}$$

$$H_x = \frac{1}{\sigma^2} \mathrm{Id}_{2\times 2}$$

$$d_{\mathrm{FR}}(x, x') = \operatorname{arcsinh} \frac{\|x - x'\|}{\sqrt{\sigma \sigma'}}$$



Exact Support Recovery

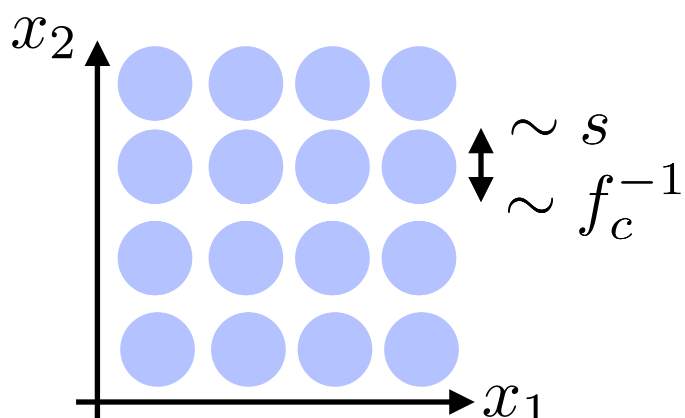
Input measure: $m_0 = \sum_i a_i \delta_{x_i}$.

Theorem: there exists Δ such that [Poon, Keriven, P, 2018]

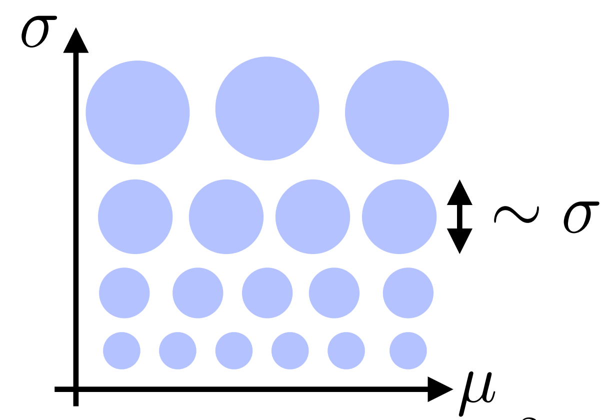
$$\max_{i \neq i'} d_{\mathrm{FR}}(x_i, x_{i'}) \geqslant \Delta \quad \Longrightarrow \quad \eta_V \text{ is non-degenerate.}$$

Remark: additional smoothness/decay assumption required.

Exclusion balls: $= \{x' \; ; \; d_{\mathrm{FR}}(x, x') \leqslant \Delta\}$



$\varphi(x) = (e^{2i\pi k x})_{k=-f_c}^{f_c}$
$\varphi(x) = h((\cdot - x)/s)$



$\varphi(x) = \frac{1}{\sigma} e^{-\frac{(\cdot - \mu)^2}{2\sigma^2}}$

Exact Support Recovery

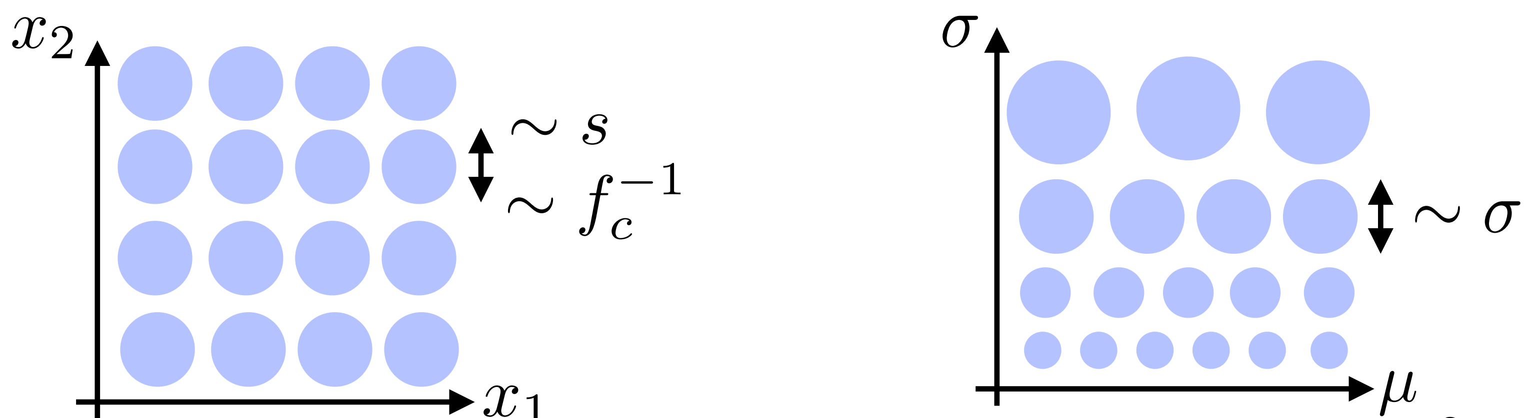
Input measure: $m_0 = \sum_i a_i \delta_{x_i}$.

Theorem: there exists Δ such that [Poon, Keriven, P, 2018]

$$\max_{i \neq i'} d_{\mathrm{FR}}(x_i, x_{i'}) \geqslant \Delta \implies \eta_V \text{ is non-degenerate.}$$

Remark: additional smoothness/decay assumption required.

Exclusion balls:  $= \{x' \; ; \; d_{\mathrm{FR}}(x, x') \leqslant \Delta\}$



$\varphi(x) = (e^{2\mathrm{i}\pi k x})_{k=-f_c}^{f_c}$

$\varphi(x) = h((\cdot - x)/s)$

$\varphi(x) = \frac{1}{\sigma} e^{-\frac{(\cdot - \mu)^2}{2\sigma^2}}$

Remark: for Fourier (or convolution), Δ independent of f_c (or s).

Remark: extends to randomized (compressed sensing) measurements.

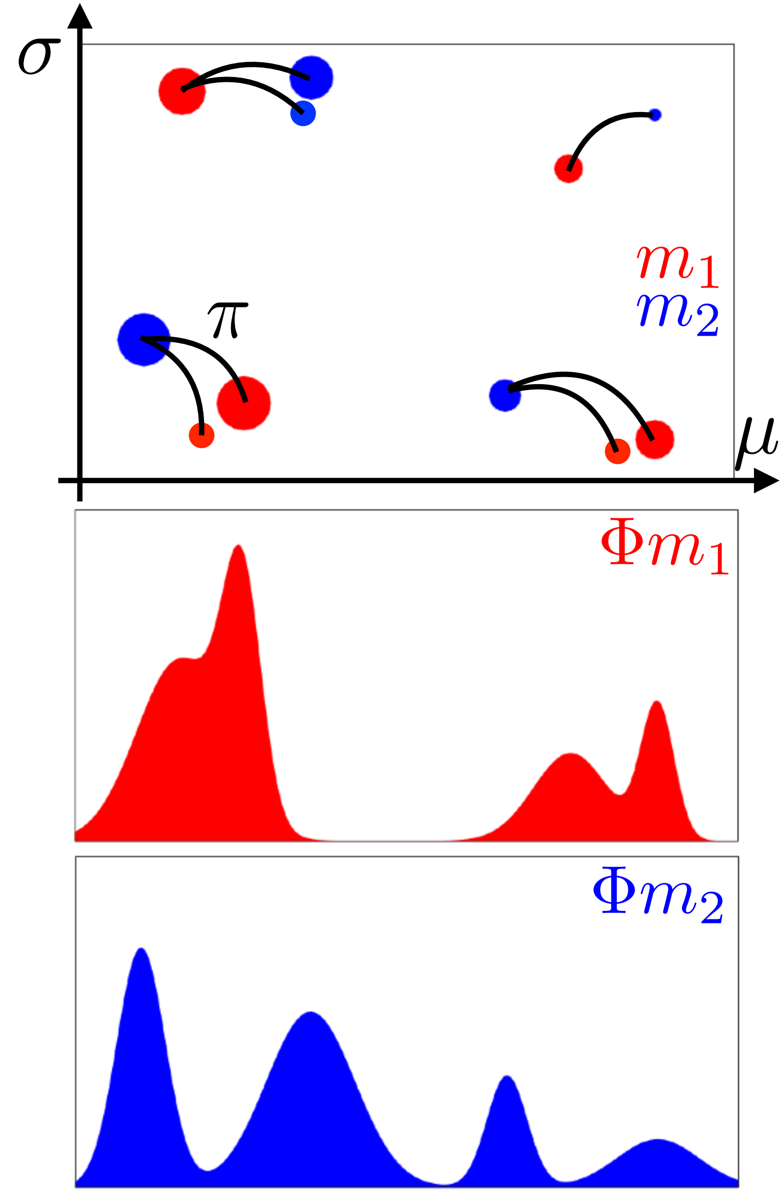
Optimal Transport Approximate Recovery

"Unbalanced" Fisher-Rao Optimal Transport:

$$W_2^2(m_1, m_2) \overset{\text{def.}}{=} \inf_{\pi \in \mathcal{M}_+(\mathcal{X}^2)} \int_{\mathcal{X}^2} d_{\text{FR}}^2(x_1, x_2) \mathrm{d}\pi(x_1, x_2) + |\pi_1 - m_1|(\mathcal{X}) + |\pi_2 - m_2|(\mathcal{X})$$

Example: Gaussian mixtures.

$$\varphi(x) = \frac{1}{\sigma} e^{-\frac{(\cdot - \mu)^2}{2\sigma^2}}$$



Optimal Transport Approximate Recovery

"Unbalanced" Fisher-Rao Optimal Transport:

$$W_2^2(m_1, m_2) \overset{\text{def.}}{=} \inf_{\pi \in \mathcal{M}_+(\mathcal{X}^2)} \int_{\mathcal{X}^2} d_{\text{FR}}^2(x_1, x_2) \mathrm{d}\pi(x_1, x_2) + |\pi_1 - m_1|(\mathcal{X}) + |\pi_2 - m_2|(\mathcal{X})$$

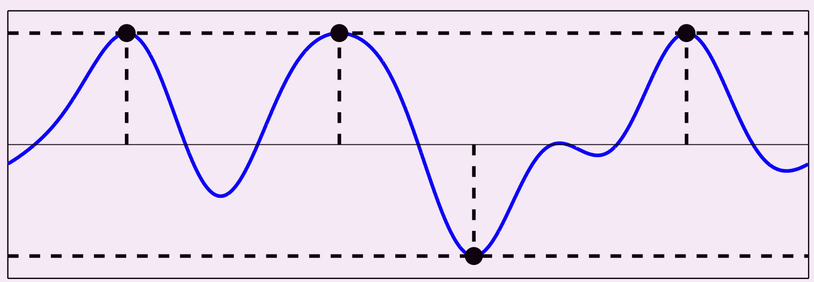
Example: Gaussian mixtures.

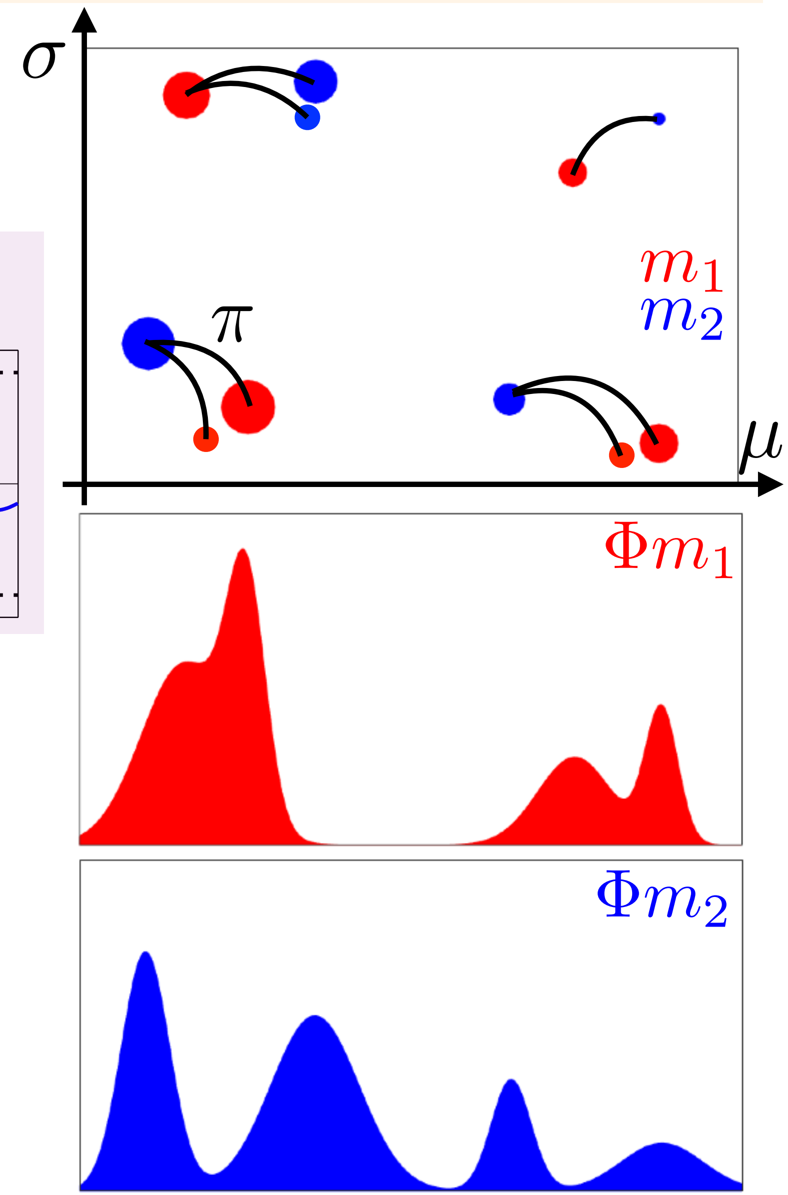
$$\varphi(x) = \frac{1}{\sigma} e^{-\frac{(\cdot - \mu)^2}{2\sigma^2}}$$

Dual certificate: $\eta \in \mathrm{Im}(\Phi^*)$,

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

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





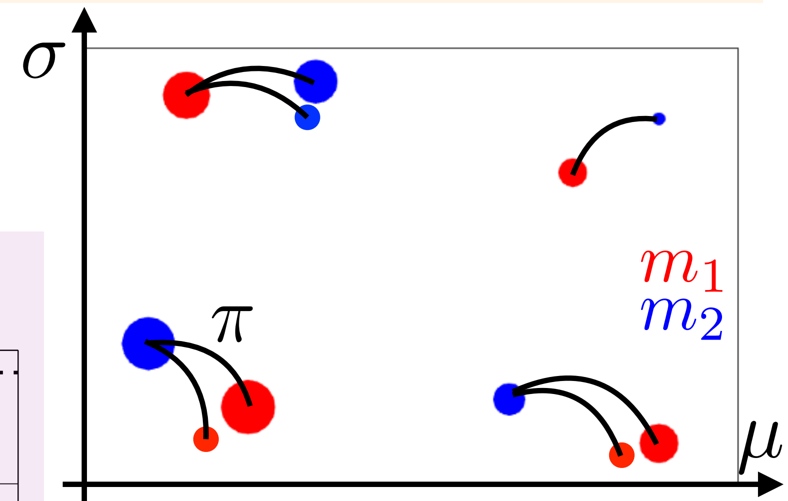
Optimal Transport Approximate Recovery

"Unbalanced" Fisher-Rao Optimal Transport:

$$W_2^2(m_1, m_2) \overset{\text{def.}}{=} \inf_{\pi \in \mathcal{M}_+(\mathcal{X}^2)} \int_{\mathcal{X}^2} d_{\mathrm{FR}}^2(x_1, x_2) \mathrm{d}\pi(x_1, x_2) + |\pi_1 - m_1|(\mathcal{X}) + |\pi_2 - m_2|(\mathcal{X})$$

Example: Gaussian mixtures.

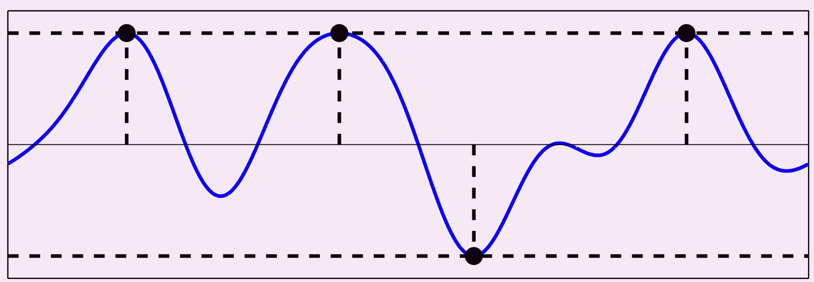
$$\varphi(x) = \frac{1}{\sigma} e^{-\frac{(\cdot - \mu)^2}{2\sigma^2}}$$

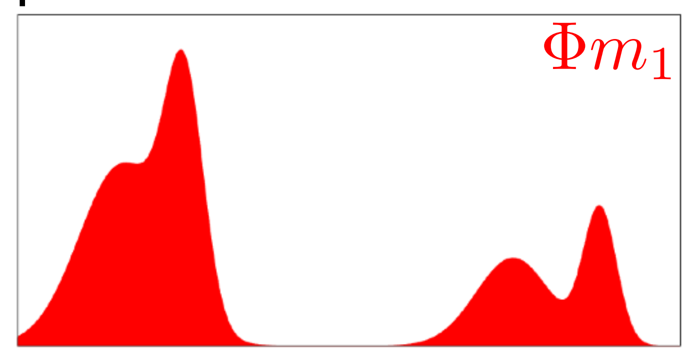


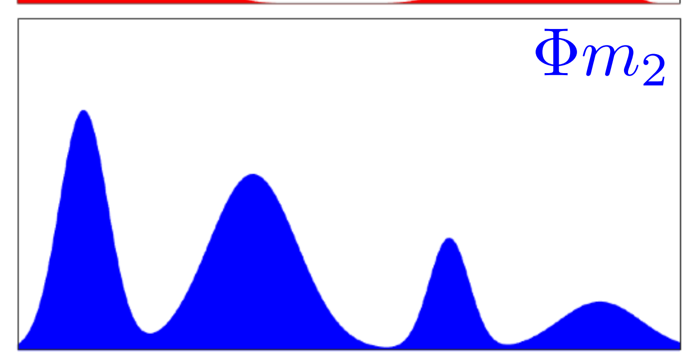
Dual certificate: $\eta \in \mathrm{Im}(\Phi^*)$,

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

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







Theorem: [Poon, Keriven, Peyré 2018]

If $\exists \eta$ non-degenerate certificate

then for any solution m_λ of $(\mathcal{P}_\lambda(\Phi m_0 + w))$

$$W_2^2(m_\lambda, m_0) = O(\|w\|) \quad \text{when} \quad \lambda \sim \|w\|$$

$\rightarrow$ golfing scheme construction of η.

Overview

- Sparse Linear Models
- Sparse Estimation with Blasso
- Non-super-resolution Regime
- **Super-resolution Regime**

Recovery of Positive Measures

Input measure: $m_0 = m_{a,x}$ where $a \in \mathbb{R}_+^N$.

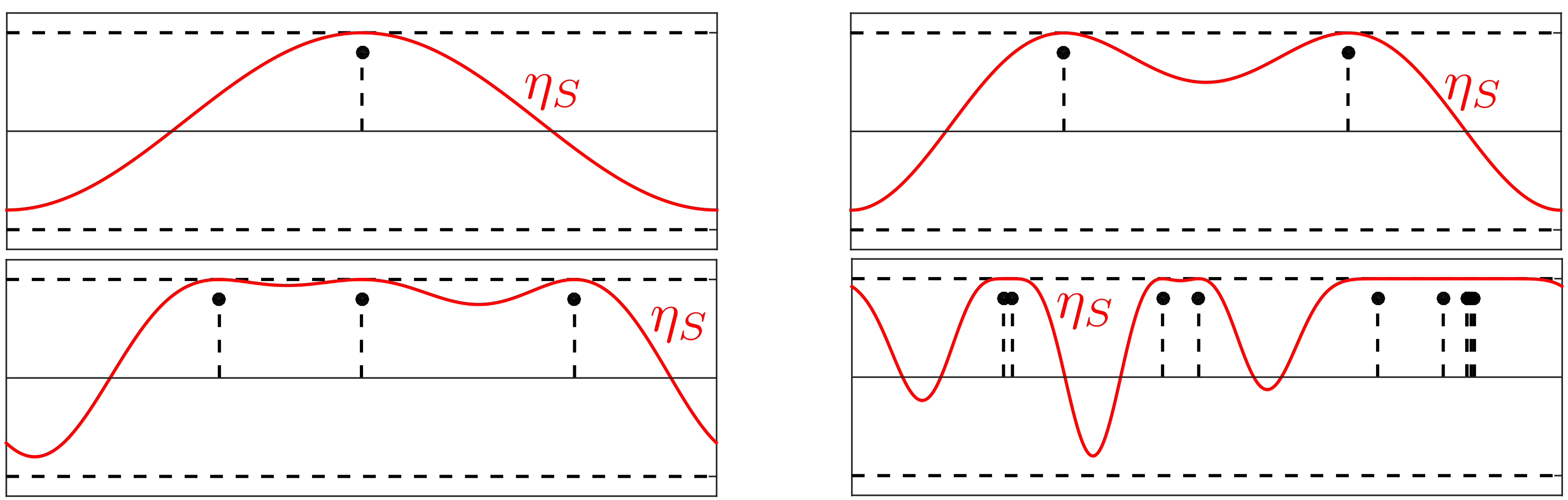
Theorem: let $\varphi(x) = (e^{\mathrm{i}\ell x})_{\ell=-f_c}^{f_c}$ and

$$\eta_S(t) = 1 - \rho \prod_{i=1}^N \sin(\pi(t - x_i))^2$$

for $N \leqslant f_c$ and ρ small enough, $\eta_S \in \mathrm{ND}(m_0)$.

$\longrightarrow$ m_0 is recovered when there is no noise.

[de Castro et al. 2011]



Recovery of Positive Measures

Input measure: $m_0 = m_{a,x}$ where $a \in \mathbb{R}_+^N$.

Theorem: let $\varphi(x) = (e^{i\ell x})_{\ell=-f_c}^{f_c}$ and

$$\eta_S(t) = 1 - \rho \prod_{i=1}^N \sin(\pi(t - x_i))^2$$

for $N \leqslant f_c$ and ρ small enough, $\eta_S \in \mathrm{ND}(m_0)$.

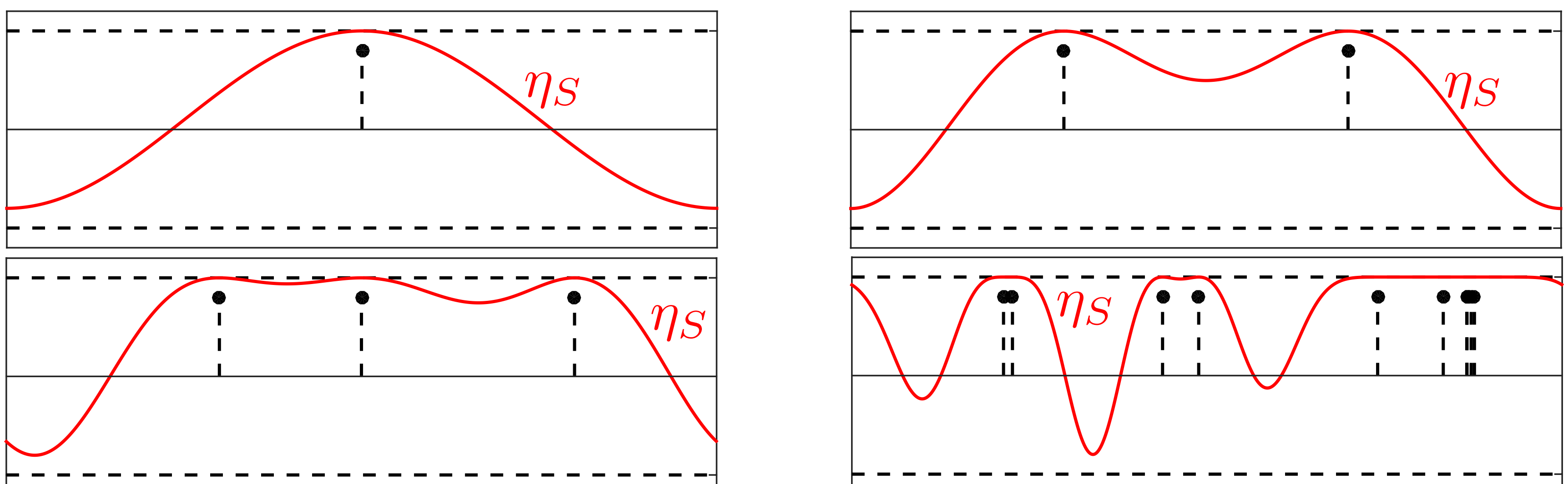
[de Castro et al. 2011]

$\longrightarrow$ m_0 is recovered when there is no noise.

$\longrightarrow$ behavior as $\forall\, i, x_i \to 0$?

[Morgenshtern, Candès, 2015] discrete ℓ^1 robustness.

[Demanet, Nguyen, 2015] discrete ℓ^0 robustness.



Recovery of Positive Measures

Input measure: $m_0 = m_{a,x}$ where $a \in \mathbb{R}_+^N$.

Theorem: let $\varphi(x) = (e^{i\ell x})_{\ell=-f_c}^{f_c}$ and

$$\eta_S(t) = 1 - \rho \prod_{i=1}^N \sin(\pi(t - x_i))^2$$

for $N \leqslant f_c$ and ρ small enough, $\eta_S \in \mathrm{ND}(m_0)$.

[de Castro et al. 2011]

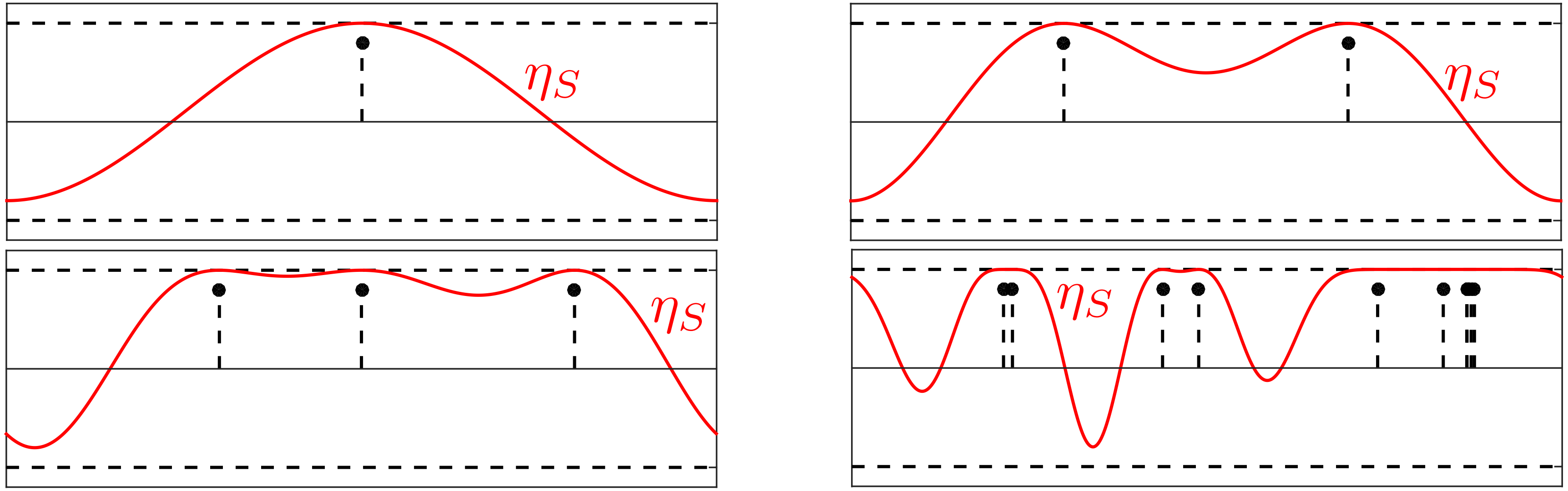
$\longrightarrow$ m_0 is recovered when there is no noise.

$\longrightarrow$ behavior as $\forall\, i, x_i \to 0$?

[Morgenshtern, Candès, 2015] discrete ℓ^1 robustness.

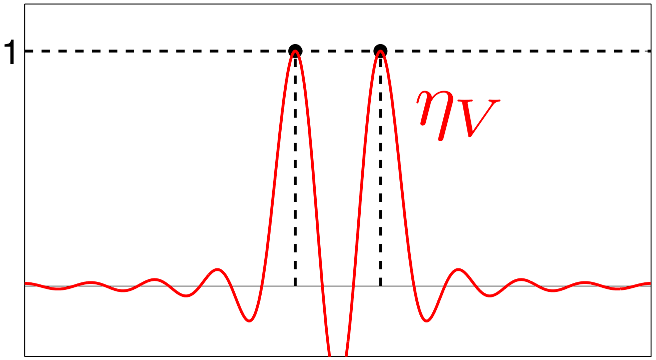
[Demanet, Nguyen, 2015] discrete ℓ^0 robustness.

$\longrightarrow$ noise robustness of support recovery ?



Asymptotic of Vanishing Certificate

Valid only in 1-D, i.e. $\mathbb{T} = \mathbb{R}$ or $\mathbb{T} = \mathbb{R}/\mathbb{Z}$.



$$m_0 = m_{a,\Delta x} \quad \text{where} \quad \Delta \to 0$$

Vanishing Derivative pre-certificate:

$$\eta_V \overset{\text{def.}}{=} \underset{\eta = \Phi^* p}{\operatorname{argmin}} \|p\| \quad \text{s.t.} \quad \forall\, i, \begin{cases} \eta(\Delta x_i) = 1, \\ \eta'(\Delta x_i) = 0. \end{cases}$$

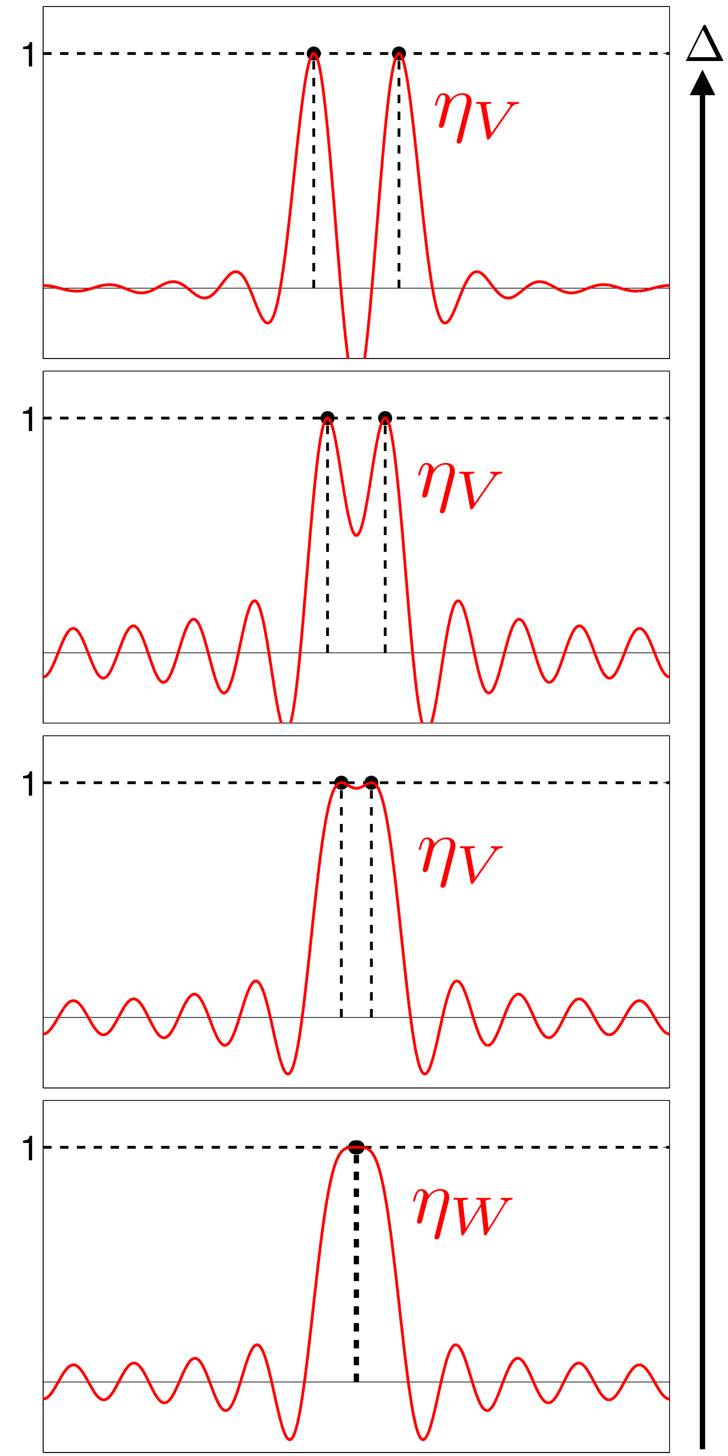
Asymptotic of Vanishing Certificate

Valid only in 1-D, i.e. $\mathbb{T} = \mathbb{R}$ or $\mathbb{T} = \mathbb{R}/\mathbb{Z}$.

$$m_0 = m_{a,\Delta x} \quad \text{where} \quad \Delta \to 0$$

Vanishing Derivative pre-certificate:

$$\eta_V \overset{\text{def.}}{=} \underset{\eta=\Phi^* p}{\text{argmin}} \|p\| \quad \text{s.t.} \quad \forall\, i, \begin{cases} \eta(\Delta x_i) = 1, \\ \eta'(\Delta x_i) = 0. \end{cases}$$



Asymptotic of Vanishing Certificate

Valid only in 1-D, i.e. $\mathbb{T} = \mathbb{R}$ or $\mathbb{T} = \mathbb{R}/\mathbb{Z}$.

$$m_0 = m_{a,\Delta x} \quad \text{where} \quad \Delta \to 0$$

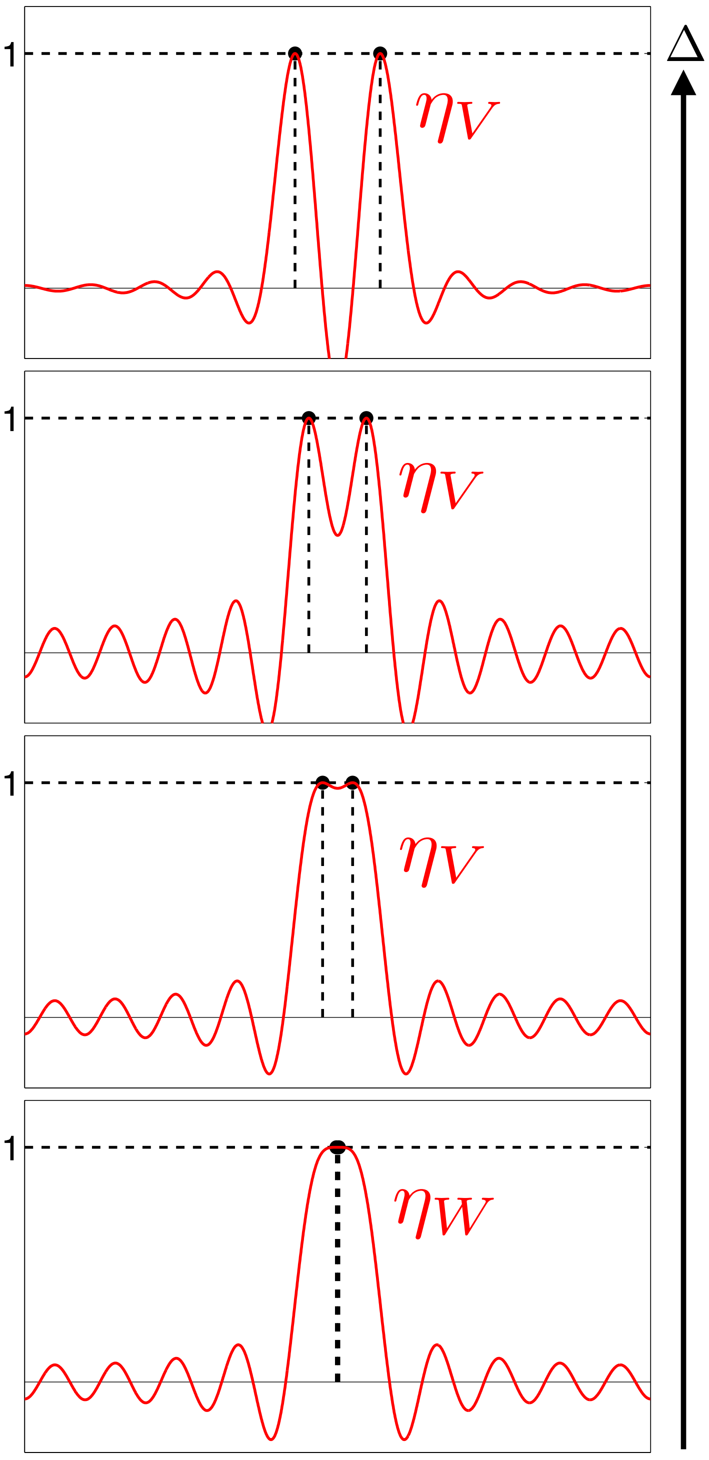
Vanishing Derivative pre-certificate:

$$\eta_V \overset{\text{def.}}{=} \underset{\eta=\Phi^* p}{\text{argmin}} \|p\| \quad \text{s.t.} \quad \forall\, i, \begin{cases} \eta(\Delta x_i) = 1, \\ \eta'(\Delta x_i) = 0. \end{cases}$$

$\Delta \to 0$

Asymptotic pre-certificate:

$$\eta_W \overset{\text{def.}}{=} \underset{\eta=\Phi^* p}{\text{argmin}} \|p\| \quad \text{s.t.} \quad \begin{cases} \eta(0) = 1, \\ \eta'(0) = \ldots = \eta^{(2N-1)}(0) = 0. \end{cases}$$



Asymptotic Robustness

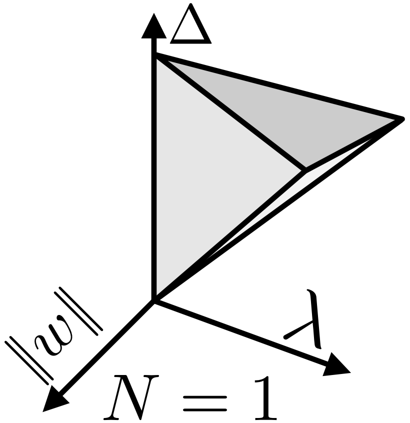
Theorem: If $\eta_W \in \mathrm{ND}_N$, letting $m_0 = m_{a,\Delta x}$, then

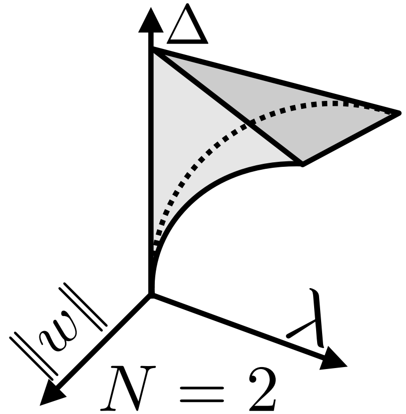
for $\left(\frac{w}{\lambda}, \frac{w}{\Delta^{2N-1}}, \frac{\lambda}{\Delta^{2N-1}}\right) = O(1)$

the solution of $\mathcal{P}_\lambda(y)$ for $y = \Phi(m_0) + w$ is

$\sum_{i=1}^N a_i^\star \delta_{\Delta x_i^\star}$ where $\|(x,a) - (x^\star, a^\star)\| = O\left(\frac{\|w\| + \lambda}{\Delta^{2N-1}}\right)$

[Denoyelle, D., P. 2015]





Asymptotic Robustness

Theorem: If $\eta_W \in \mathrm{ND}_N$, letting $m_0 = m_{a,\Delta x}$, then

for $\left(\frac{w}{\lambda}, \frac{w}{\Delta^{2N-1}}, \frac{\lambda}{\Delta^{2N-1}}\right) = O(1)$

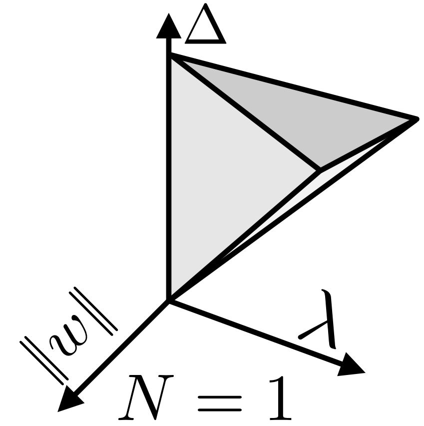
the solution of $\mathcal{P}_\lambda(y)$ for $y = \Phi(m_0) + w$ is

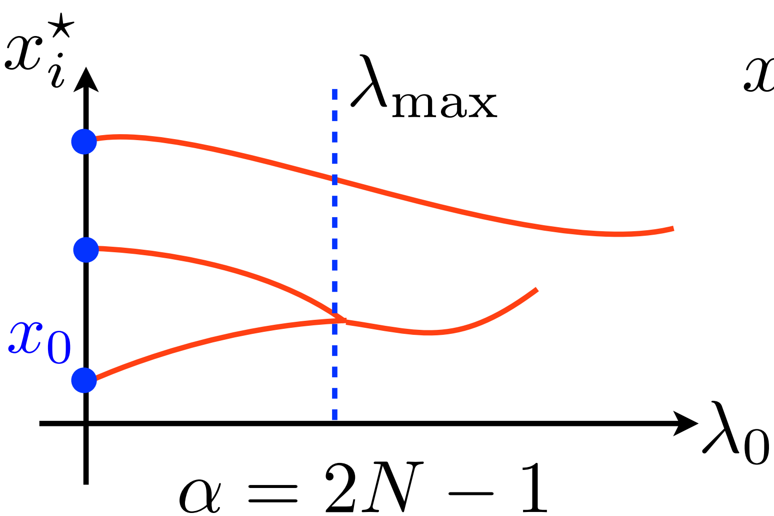
$\sum_{i=1}^N a_i^\star \delta_{\Delta x_i^\star}$ where $\|(x,a) - (x^\star, a^\star)\| = O\left(\frac{\|w\| + \lambda}{\Delta^{2N-1}}\right)$

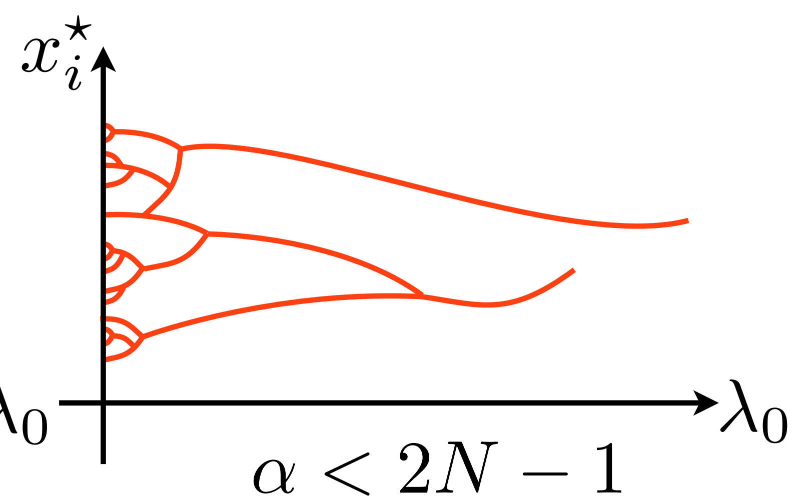
[Denoyelle, D., P. 2015]

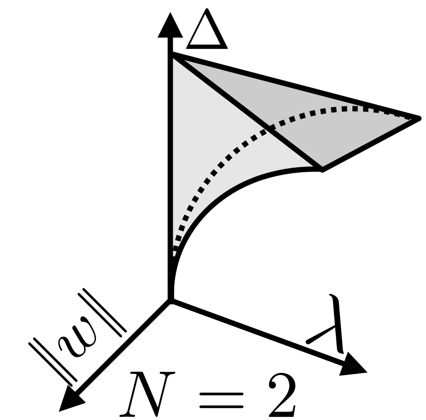
$y = \Phi m_{a,\Delta x} + w$

Noise: $w = \lambda w_0$.
Regularization: $\lambda = \lambda_0 \Delta^\alpha$









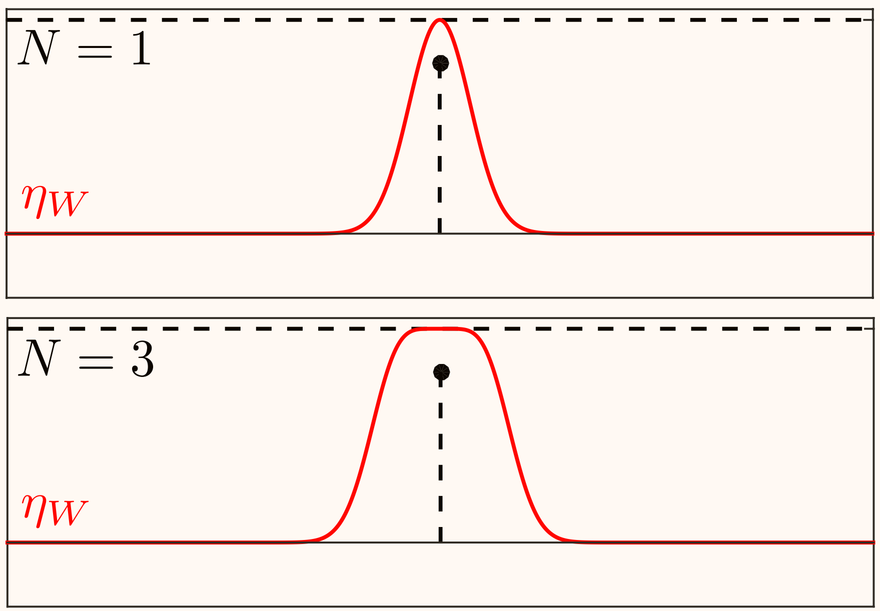
Deconvolution and Laplace

Gaussian convolution: $\varphi(x) = e^{-\frac{|x-\cdot|^2}{2\sigma^2}}$

Proposition: $\eta_W(x) = e^{-\frac{x^2}{4\sigma^2}} \sum_{k=0}^{N-1} \frac{(x/2\sigma)^{2k}}{k!}$

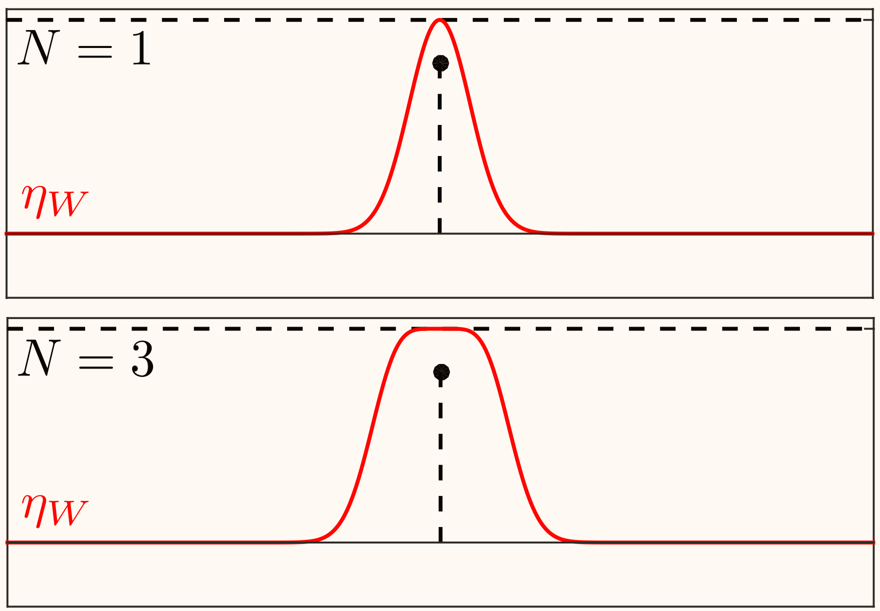
In particular, η_W is non-degenerate.

$\longrightarrow$ Gaussian deconvolution is support-stable.



Deconvolution and Laplace

Gaussian convolution: $\varphi(x) = e^{-\frac{|x-\cdot|^2}{2\sigma^2}}$



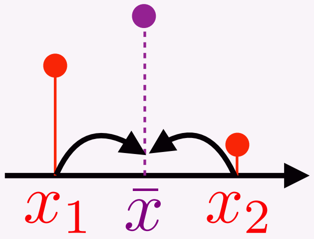
Proposition: $\eta_W(x) = e^{-\frac{x^2}{4\sigma^2}} \sum_{k=0}^{N-1} \frac{(x/2\sigma)^{2k}}{k!}$

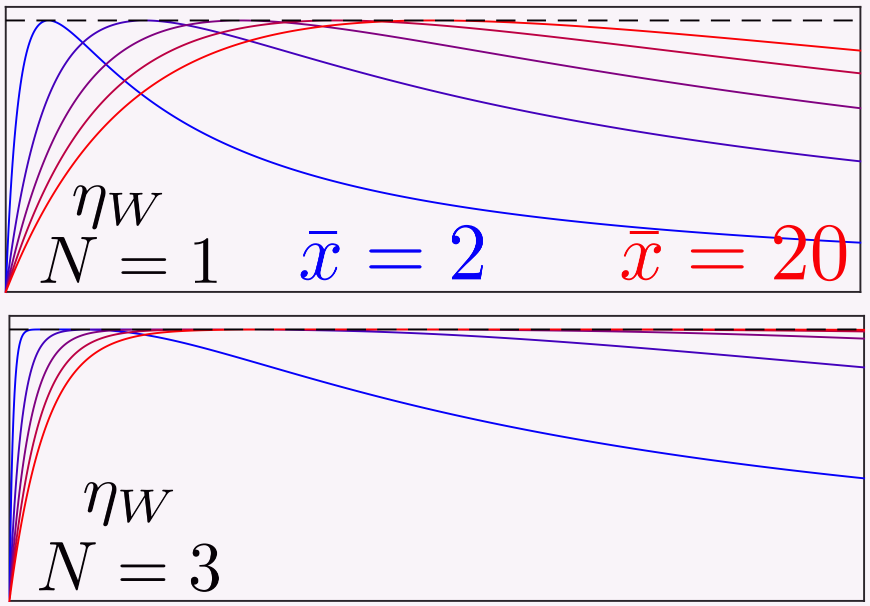
In particular, η_W is non-degenerate.

$\longrightarrow$ Gaussian deconvolution is support-stable.

Laplace transform: $\varphi(x) = e^{-x\cdot}$

Non-translation-invariant operator
$\rightarrow$ η_W depends on $\bar{x}$!



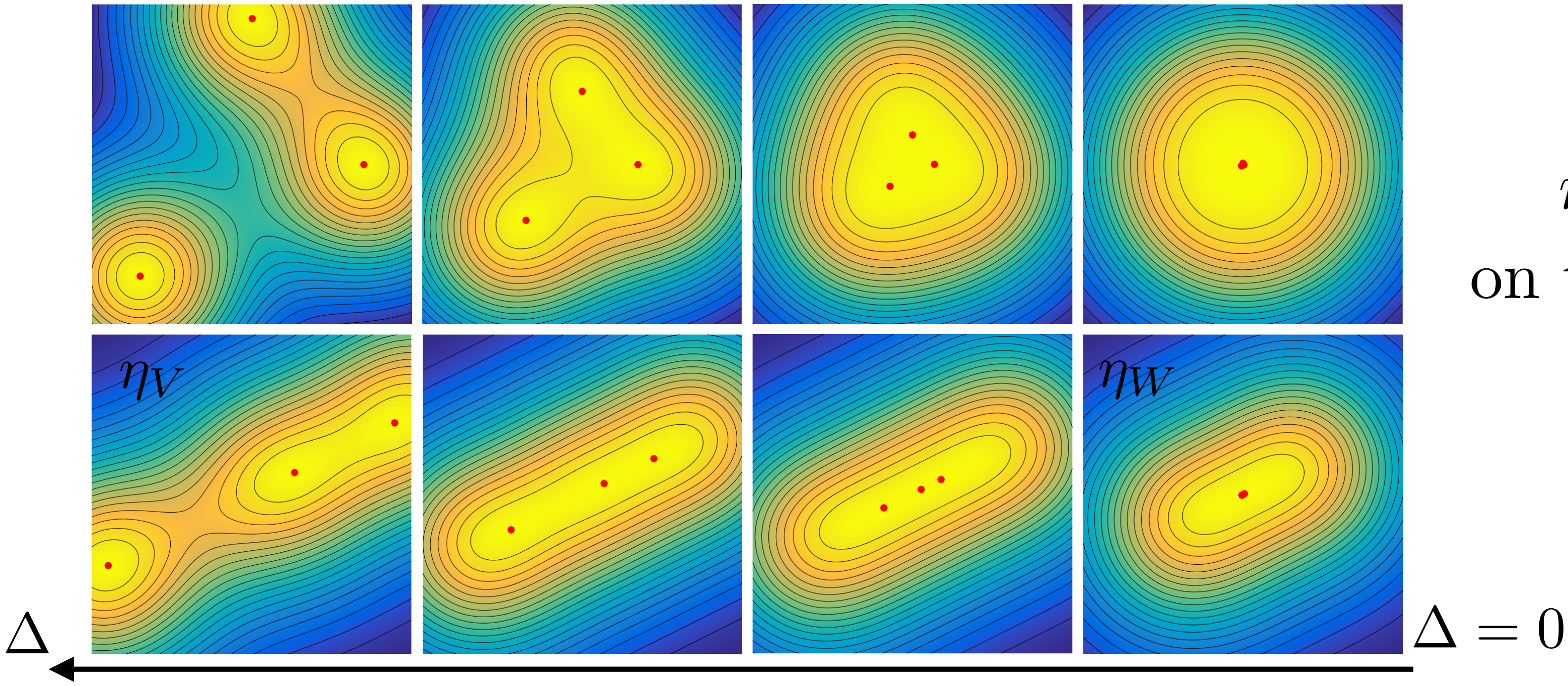


Proposition: $\eta_W(x) = 1 - \left(\frac{x - \bar{x}}{x + \bar{x}}\right)^{2N}$

In particular, η_W is non-degenerate.

Higher Dimension

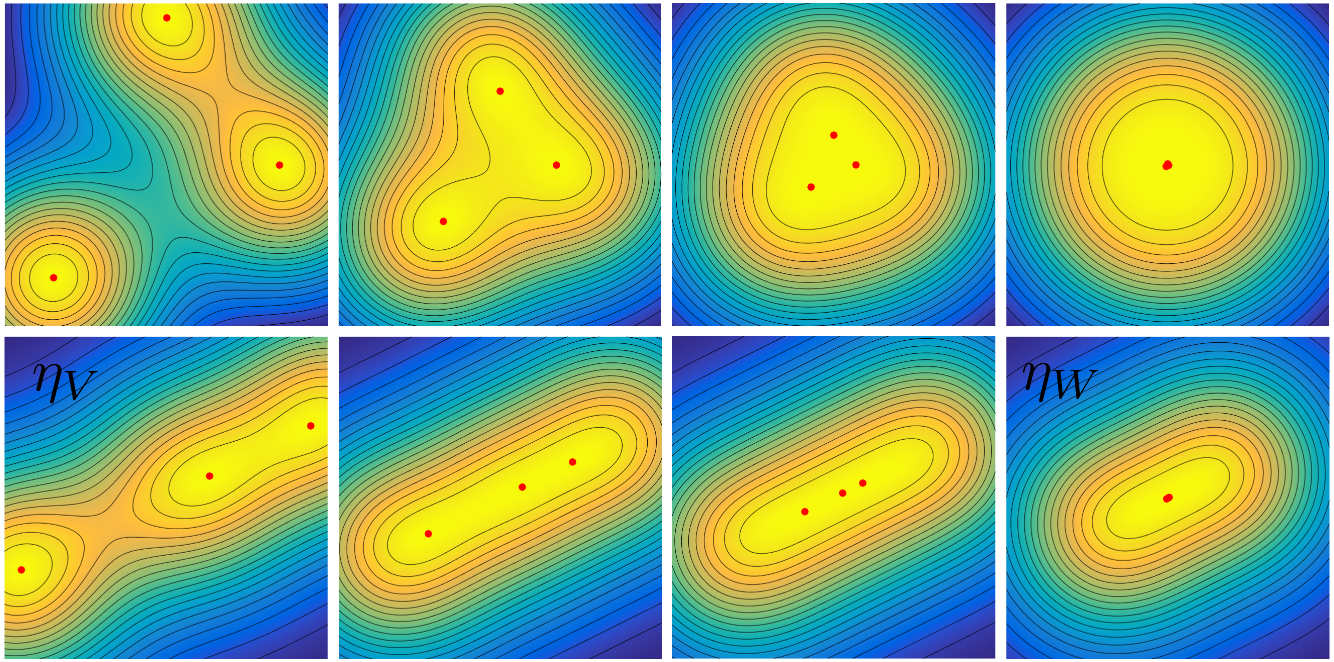
Gaussian convolution: $m_0 = m_{a,\Delta x}$ where $\Delta \to 0$



η_W depends on the directions x!

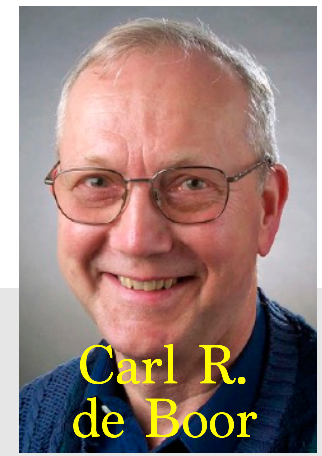
Higher Dimension

Gaussian convolution: $m_0 = m_{a,\Delta x}$ where $\Delta \to 0$



η_W depends on the directions x!

Δ ← $\Delta = 0$



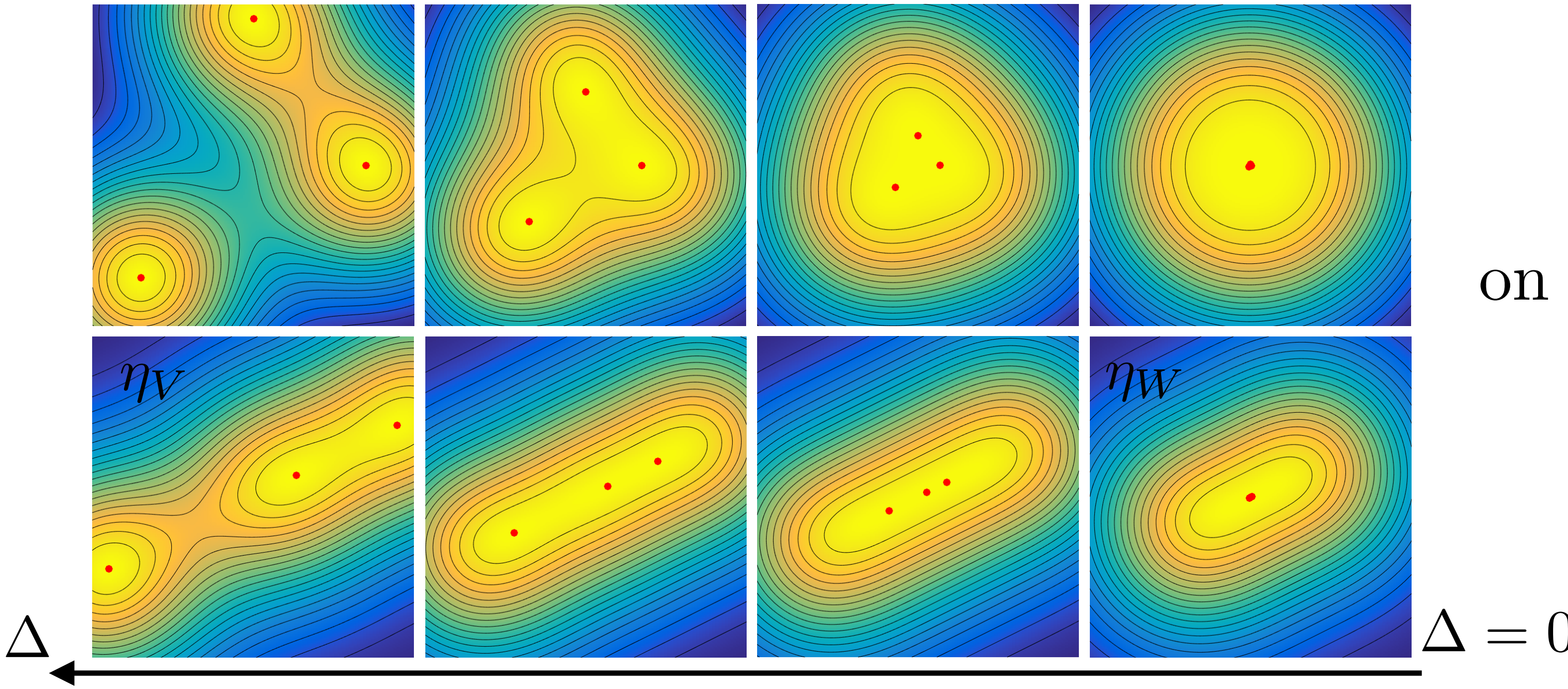
Proposition: $\eta_V \xrightarrow{\Delta \to 0} \eta_W$ where [Poon, P. 2018]

$$\eta_W \overset{\text{def.}}{=} \underset{\eta=\Phi^* p}{\operatorname{argmin}} \|p\| \quad \text{s.t.} \quad \begin{cases} \eta(0) = 1, \\ P_1(\partial)\eta(0) = \ldots = P_{(d+1)N-1}(\partial)\eta(0) = 0. \end{cases}$$

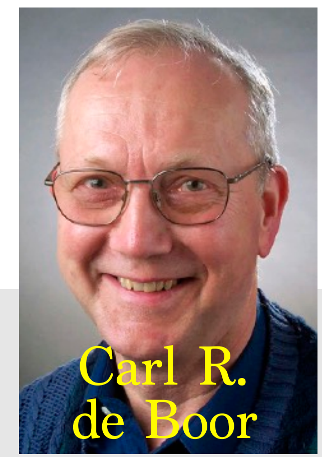
De Boor basis

Higher Dimension

Gaussian convolution: $m_0 = m_{a,\Delta x}$ where $\Delta \to 0$



η_W depends on the directions x!



Proposition: $\eta_V \xrightarrow{\Delta \to 0} \eta_W$ where [Poon, P. 2018]

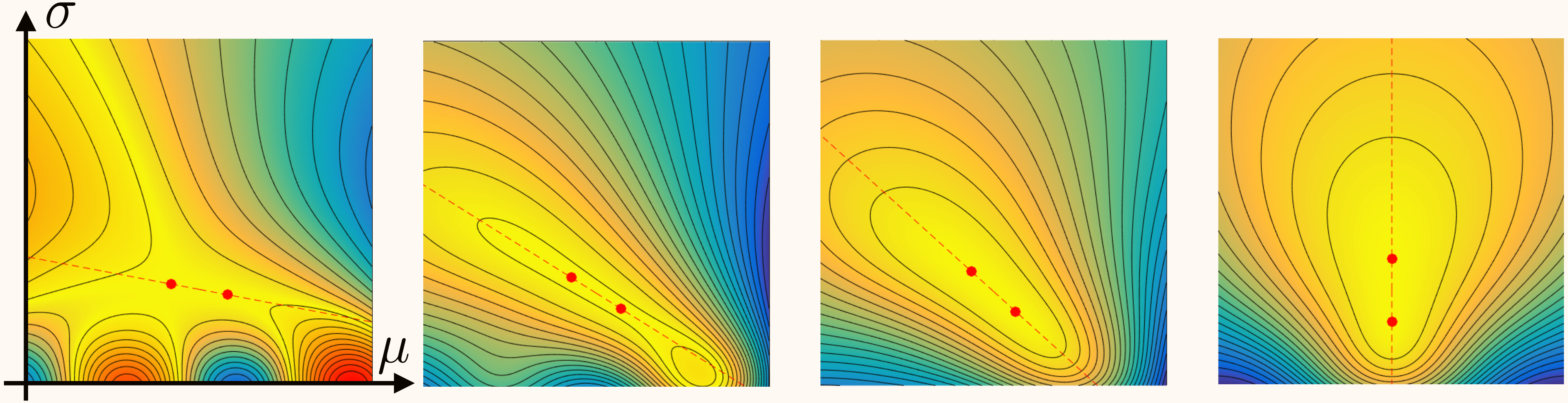
$$\eta_W \overset{\text{def.}}{=} \underset{\eta=\Phi^* p}{\operatorname{argmin}} \|p\| \quad \text{s.t.} \quad \begin{cases} \eta(0) = 1, \\ P_1(\partial)\eta(0) = \ldots = P_{(d+1)N-1}(\partial)\eta(0) = 0. \end{cases}$$

De Boor basis

Proposition: Non-degeneracy of η_W is necessary for stable recovery.
necessary and sufficient for $N = 2$ spikes.

Gaussian Mixtures

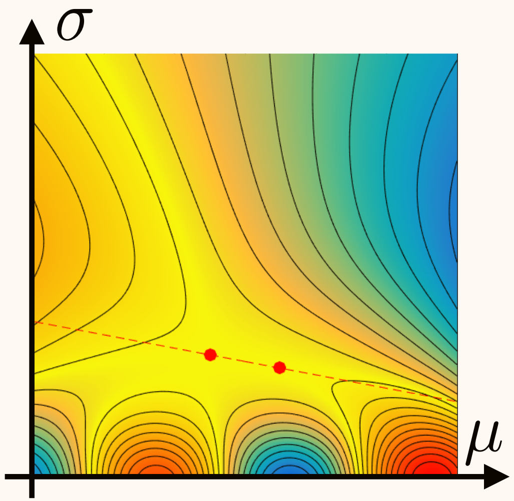
Gaussian mixtures: $\mathcal{X} = \mathbb{R} \times \mathbb{R}^+, \quad \varphi(x) = \frac{1}{\sigma} e^{-\frac{(\cdot - \mu)^2}{2\sigma^2}}$



η_W is non-degenerate if $\Delta_m \leqslant \Delta_\sigma$.

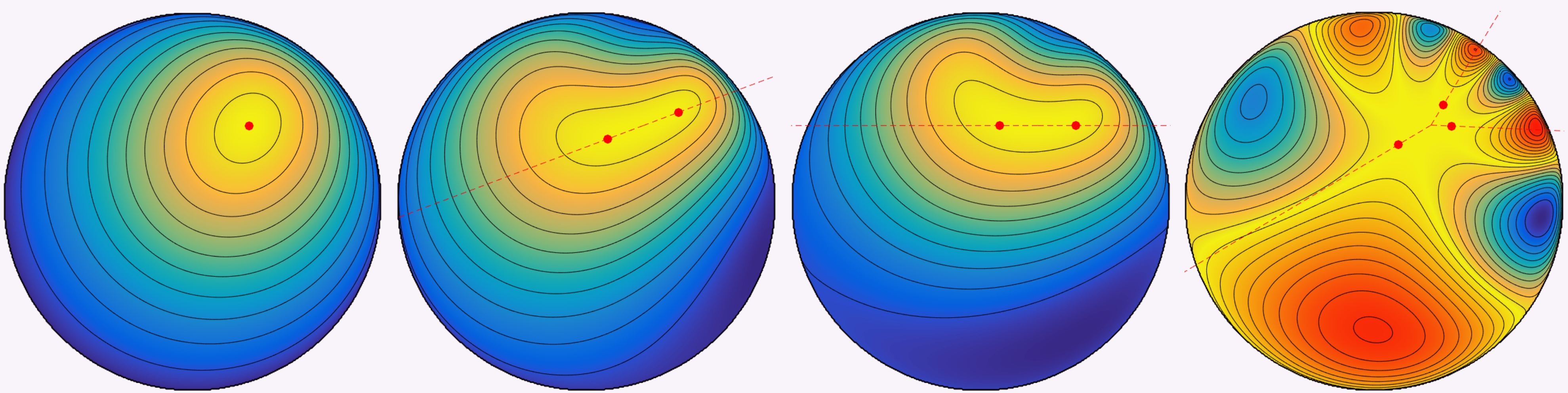
Gaussian Mixtures

Gaussian mixtures: $\mathcal{X} = \mathbb{R} \times \mathbb{R}^+, \quad \varphi(x) = \frac{1}{\sigma} e^{-\frac{(\cdot - \mu)^2}{2\sigma^2}}$



η_W is non-degenerate if $\Delta_m \leqslant \Delta_\sigma$.

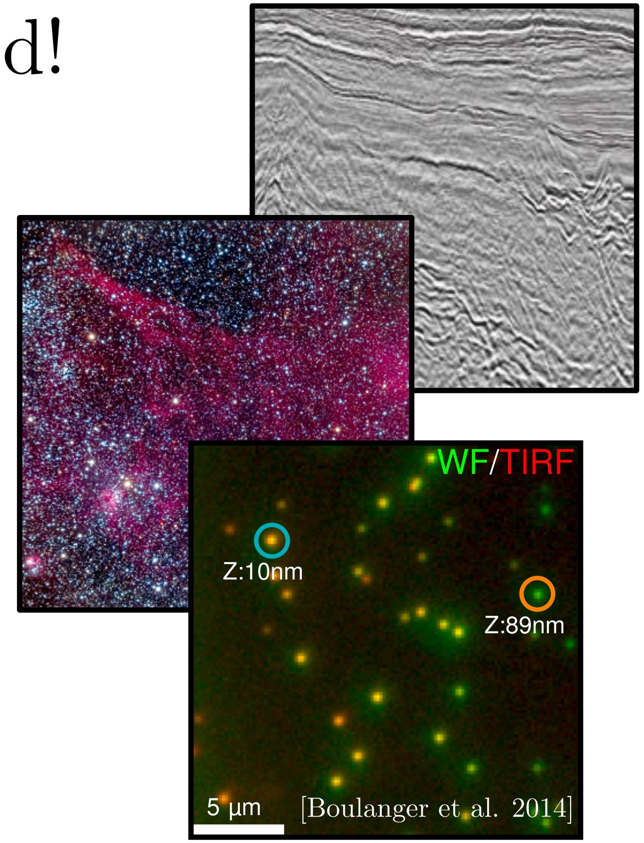
Neuro-imaging: $\mathcal{X} = \left\{ x \in \mathbb{R}^2 \; ; \; \|x\| \leqslant 1 \right\}, \quad \varphi(x) = (\|x - z\|^{-2})_{\|z\|=1}$



$N = 1$ $N = 2$ $N = 2$ $N = 3$

Conclusion

Super-resolution should be off-the-grid!



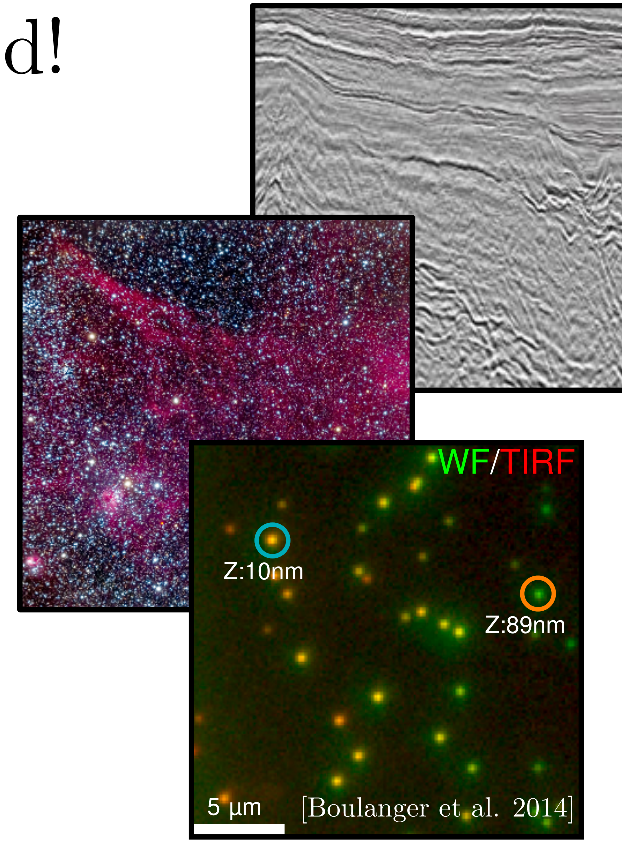
More efficient algorithms.

More accurate theory:

- "ℓ^2" error meaningless.
- Track the support.
- Fisher-Rao + Optimal Transport.

Conclusion

Super-resolution should be off-the-grid!



More efficient algorithms.

More accurate theory:

– "ℓ^2" error meaningless.

– Track the support.

– Fisher-Rao + Optimal Transport.

Open problem: other regularizations (e.g. piecewise constant) ?
see [Chambolle, Duval, Peyré, Poon 2016] for TV *denoising*.