Mini-course on Sparse estimation off-the-grid

Introduction

Clarice Poon

Outline

- Session 1: Introduction to sparse estimation as optimisation over the space of measures

- Session 2: Algorithms

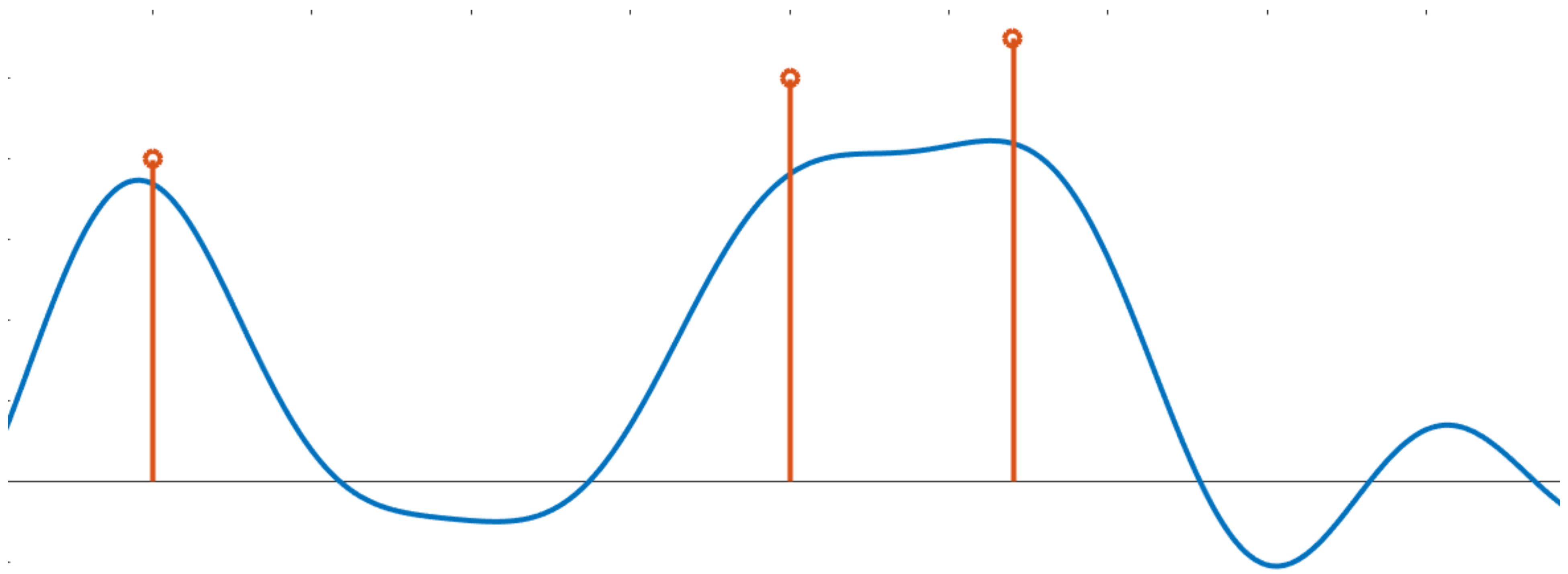
- Session 3: Super resolution and compressed sensing guarantees

Sparse Estimation

Recovering point wise sources from low resolution data

Let $\mathscr{X} \subseteq \mathbb{R}^n$ and let $\phi : \mathscr{X} \to \mathscr{H}$ where $\mathscr{H}$ is a Hilbert space.

Recover $a_j \in \mathbb{R}$ and $x_j \in \mathscr{X}$ given $\quad y = \sum_{j=1}^{s} a_j \phi(x_j)$



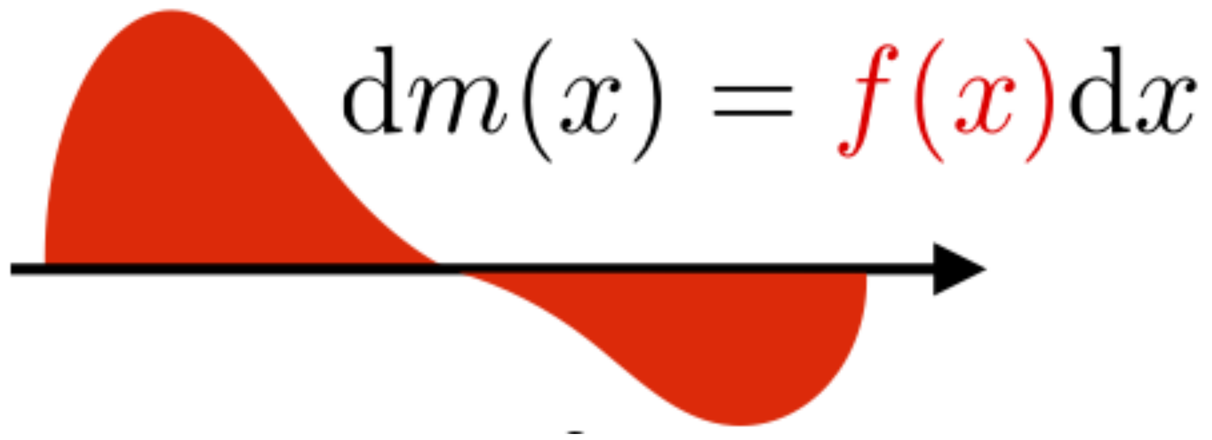
$$y = \sum_{j=1}^{s} a_j [\exp(2\pi\sqrt{-1}x_j k)]_k$$

Radon measures

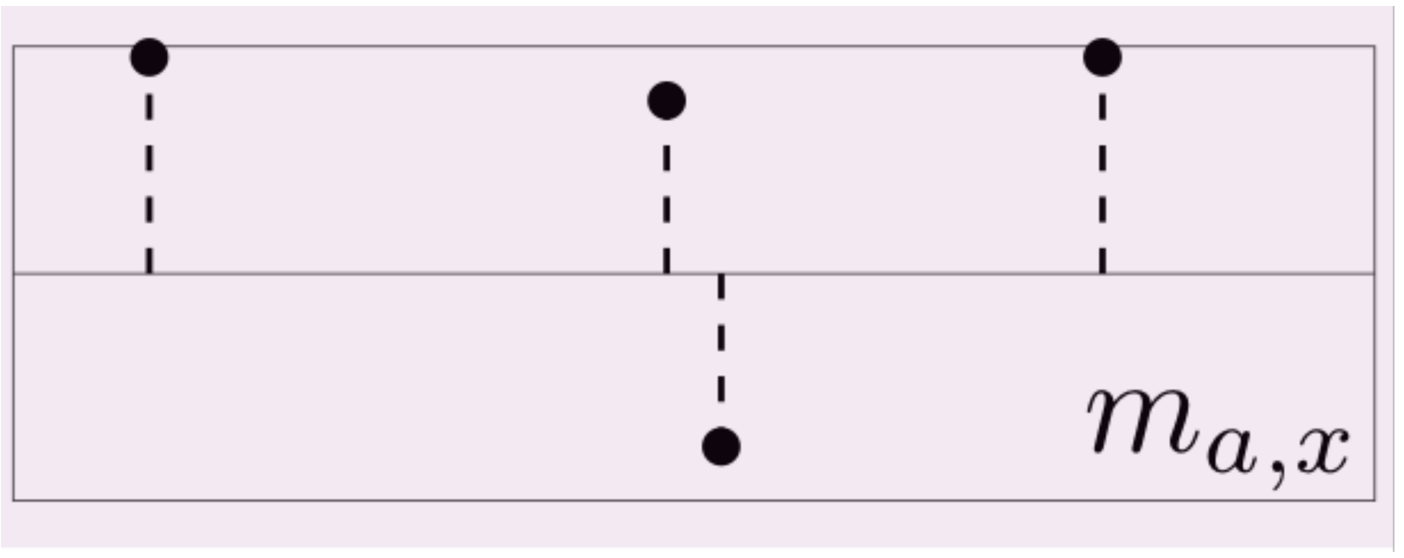
The space of Radon measures $\mathscr{M}(\mathscr{X})$ is the dual of

$$C_0(\mathscr{X}) = \overline{\left\{ f \in C(\mathscr{X}) : f \quad \text{has compact support in } \mathscr{X} \right\}}^{\|\cdot\|_\infty}$$

View $\mu \in \mathscr{M}(\mathscr{X})$ as linear functional on $C_0(\mathscr{X})$:

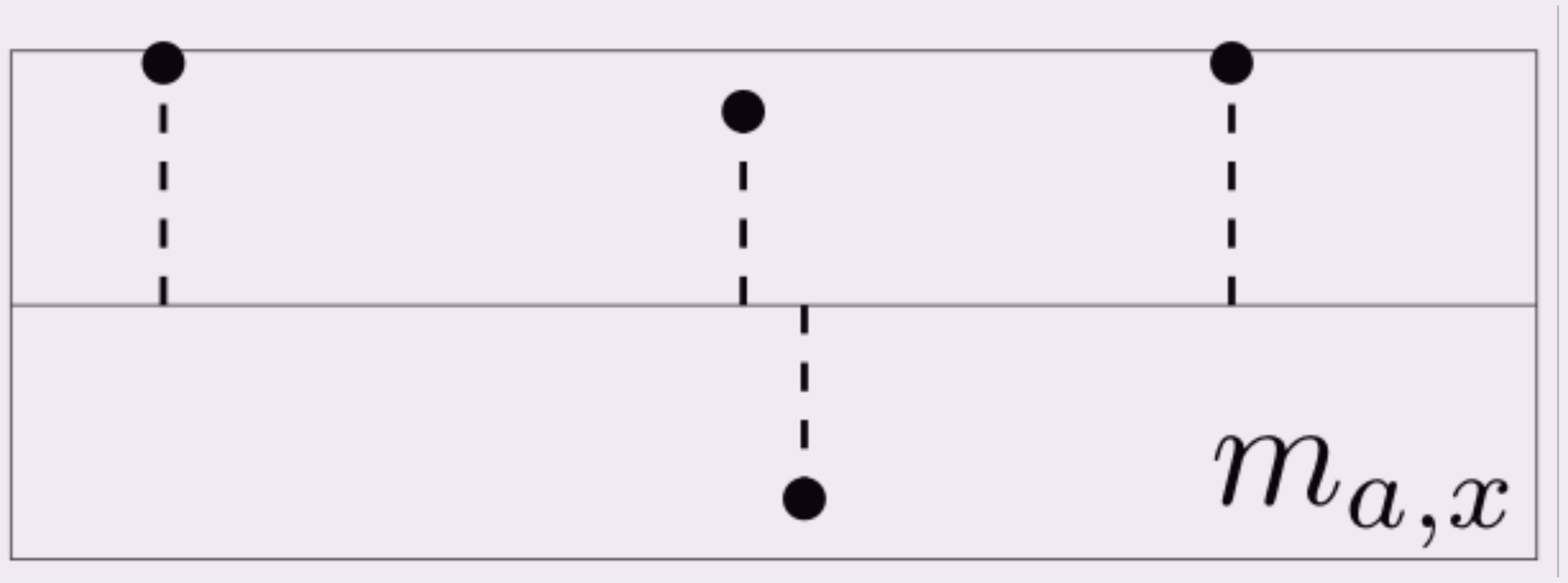


- For $f \in L^1(\mathscr{X})$, define μ by $\langle \phi, \mu \rangle = \int \phi(x) f(x) \mathrm{d}x$
- For $\mu = \sum_j a_j \delta_{x_j}$, $\quad \langle \phi, \mu \rangle = \sum_j \phi(x_j) a_j$

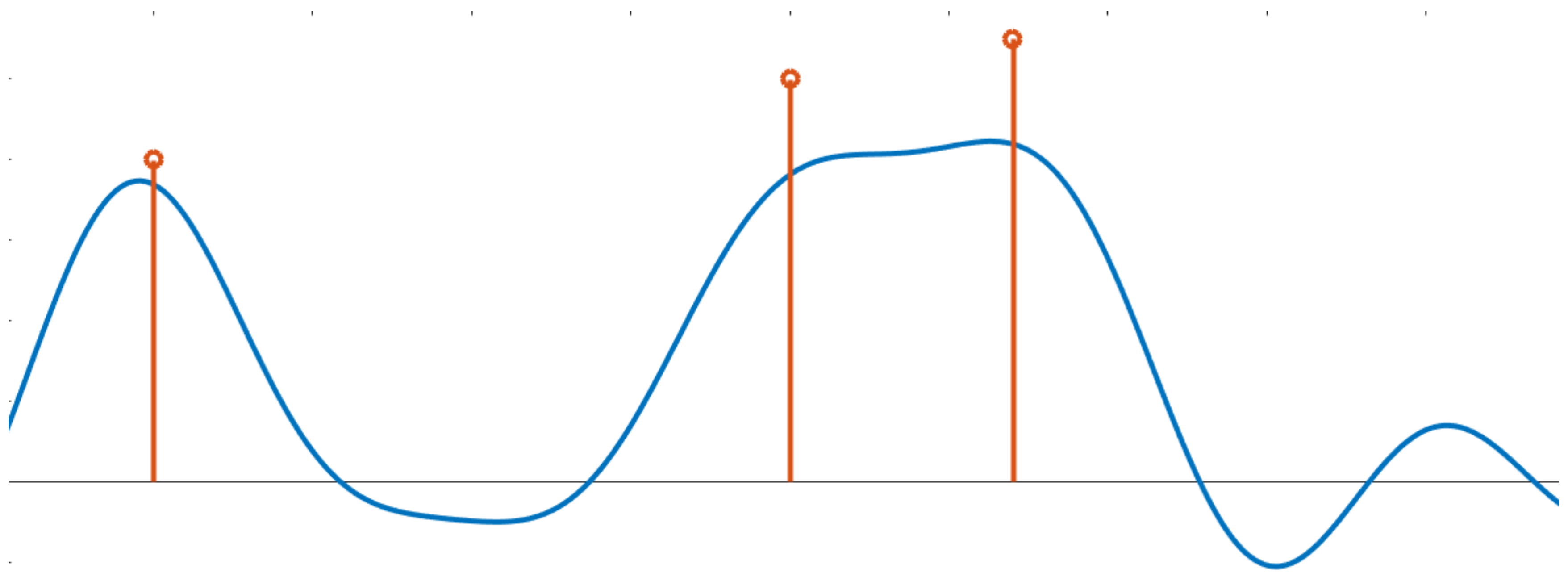


Linear inverse problem

Consider a measure μ on $\mathscr{X} \subseteq \mathbb{R}^n$



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



Observe linear measurements:

Define: $$\Phi\mu = \int_{\mathscr{X}} \phi(x) \mathrm{d}\mu(x)$$

$\phi(x) \in \mathscr{H}$ where $\phi : \mathscr{X} \to \mathscr{H}$

Observe: $$y = \Phi\mu + \text{noise}$$

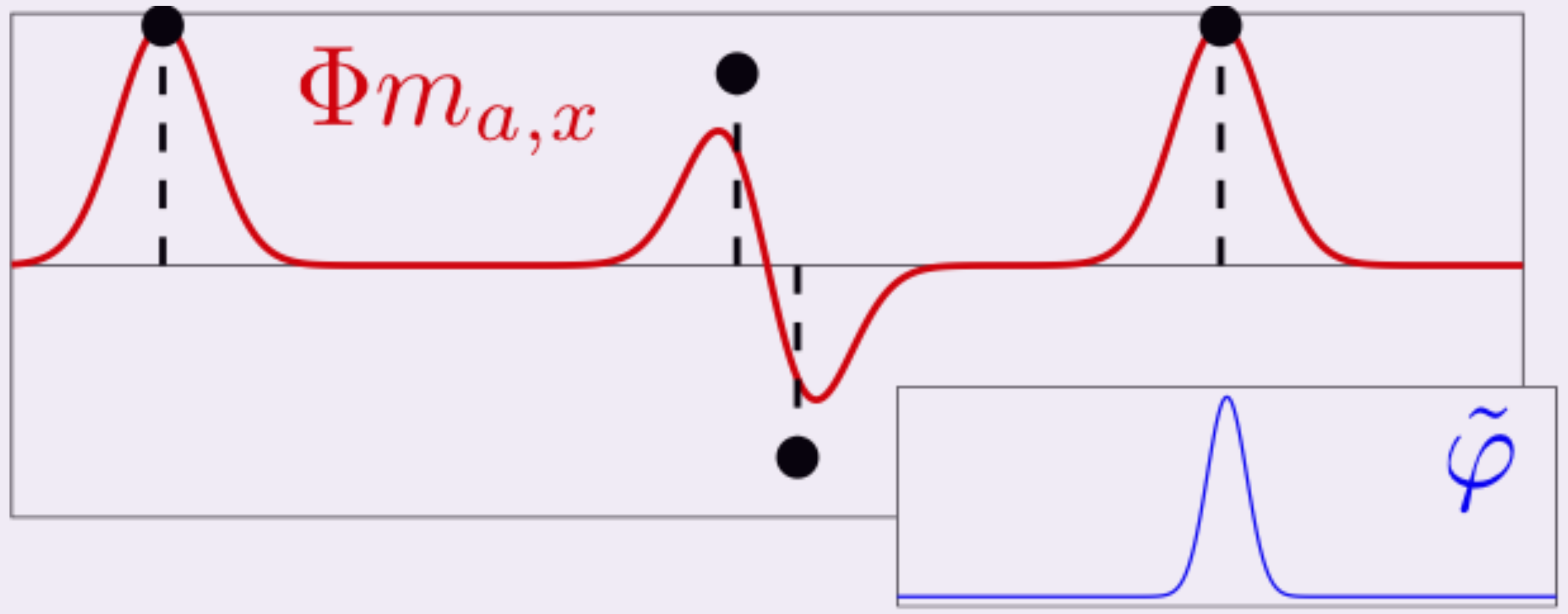
NB: $$\Phi\mu_{a,x} = \sum_{i=1}^{n} a_i \, \phi(x_i)$$

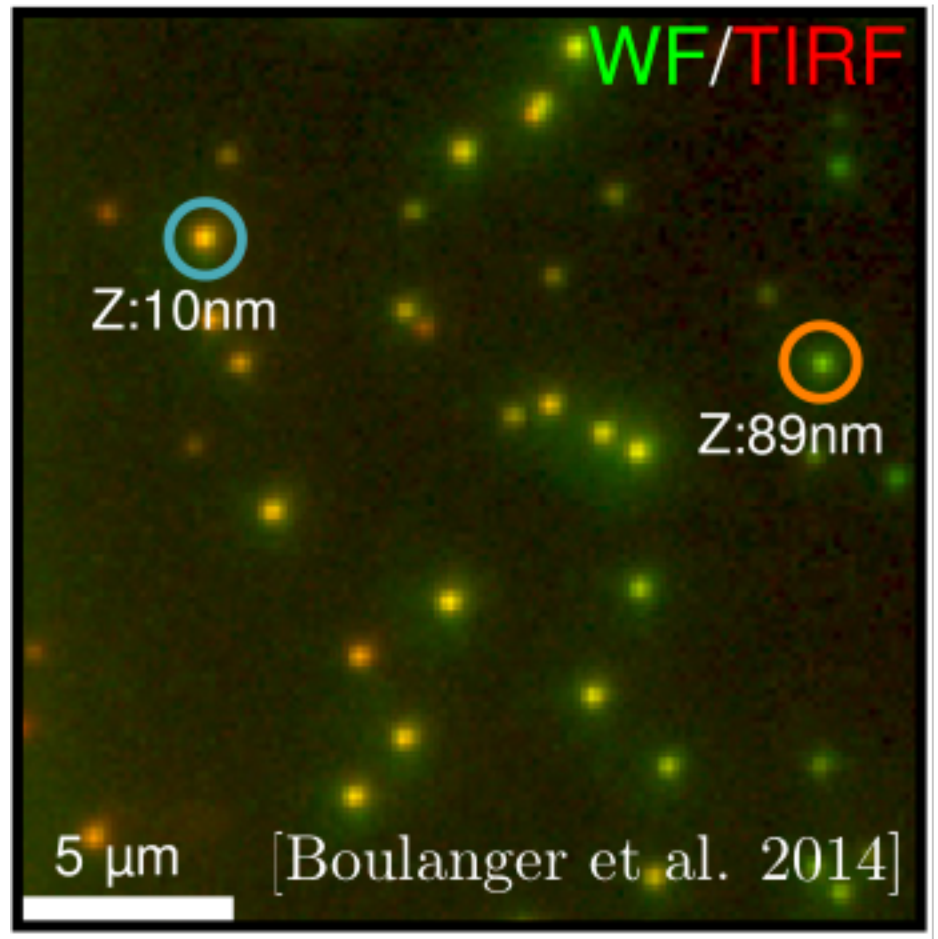
Signal/image processing

Deconvolution:

$$\phi(x) = \tilde{\phi}(\cdot - x) \in L^2(\mathbb{R}^d)$$

e.g. $\tilde{\phi}(x) = \exp(|x - \cdot|^2/\sigma)$





Laplace:

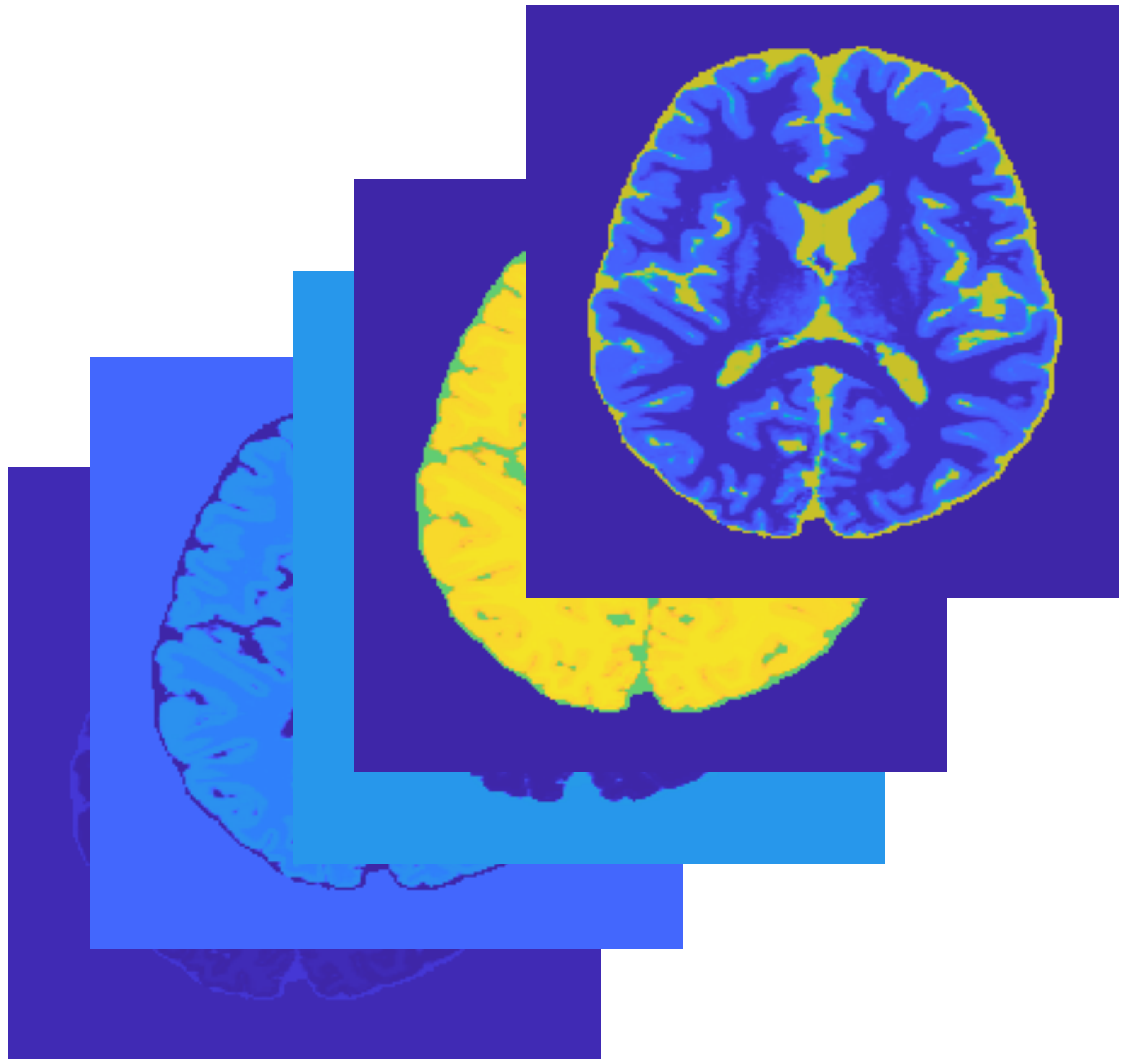
$$\phi(x) = \exp(-\langle x, \cdot \rangle) \in L^2(\mathbb{R}^d_+)$$

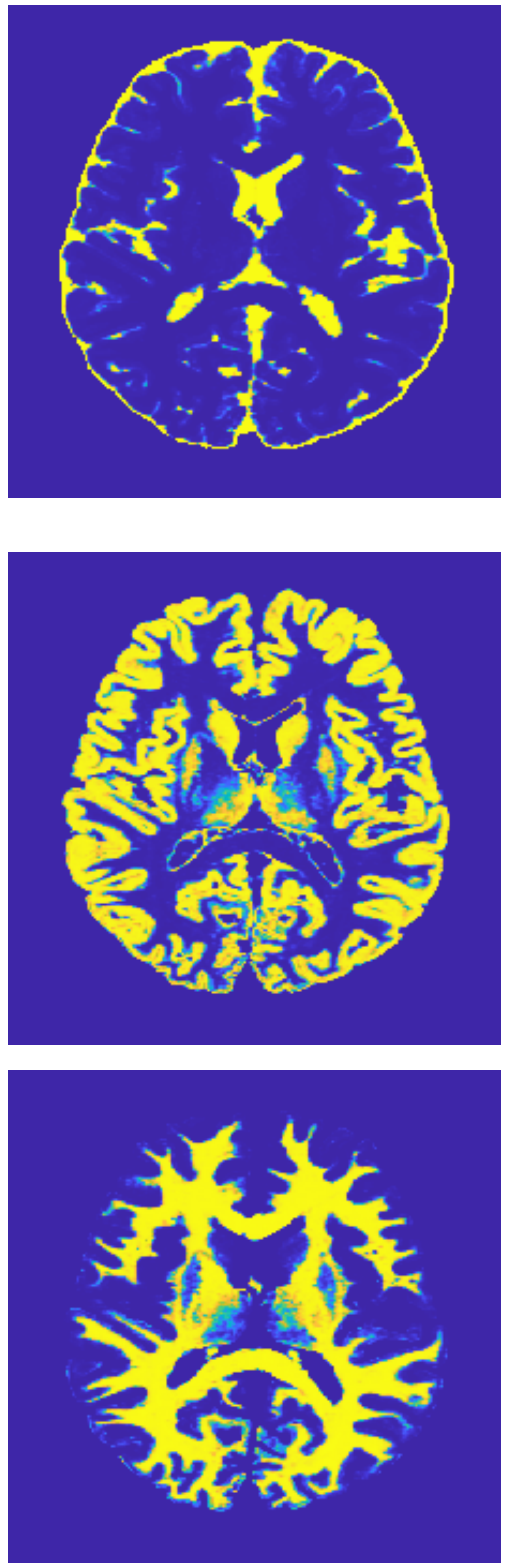
Fourier:

$$\phi(x) = (\exp(kx\sqrt{-1}))_{k=-f_c,\dots,f_c} \in \mathbb{C}^{2f_c+1}$$

Quantitative MRI

Time series data $Y = (y^v)$





θ_1

θ_2

θ_3

Time series measurements at voxel v:

$$y^v = [y_1, y_2, \ldots, y_T]$$

Recover the NMR properties

$$y = \sum_{i=1}^{s} a_i \, \phi(\theta_i) = \int \phi(\theta) \mathrm{d}\mu_{a,\theta}(\theta)$$

There can be more than 1 tissue type in each image voxel (so $n > 1$).

$\theta \in \mathcal{X}$ are parameters corresponding to different NMR properties.

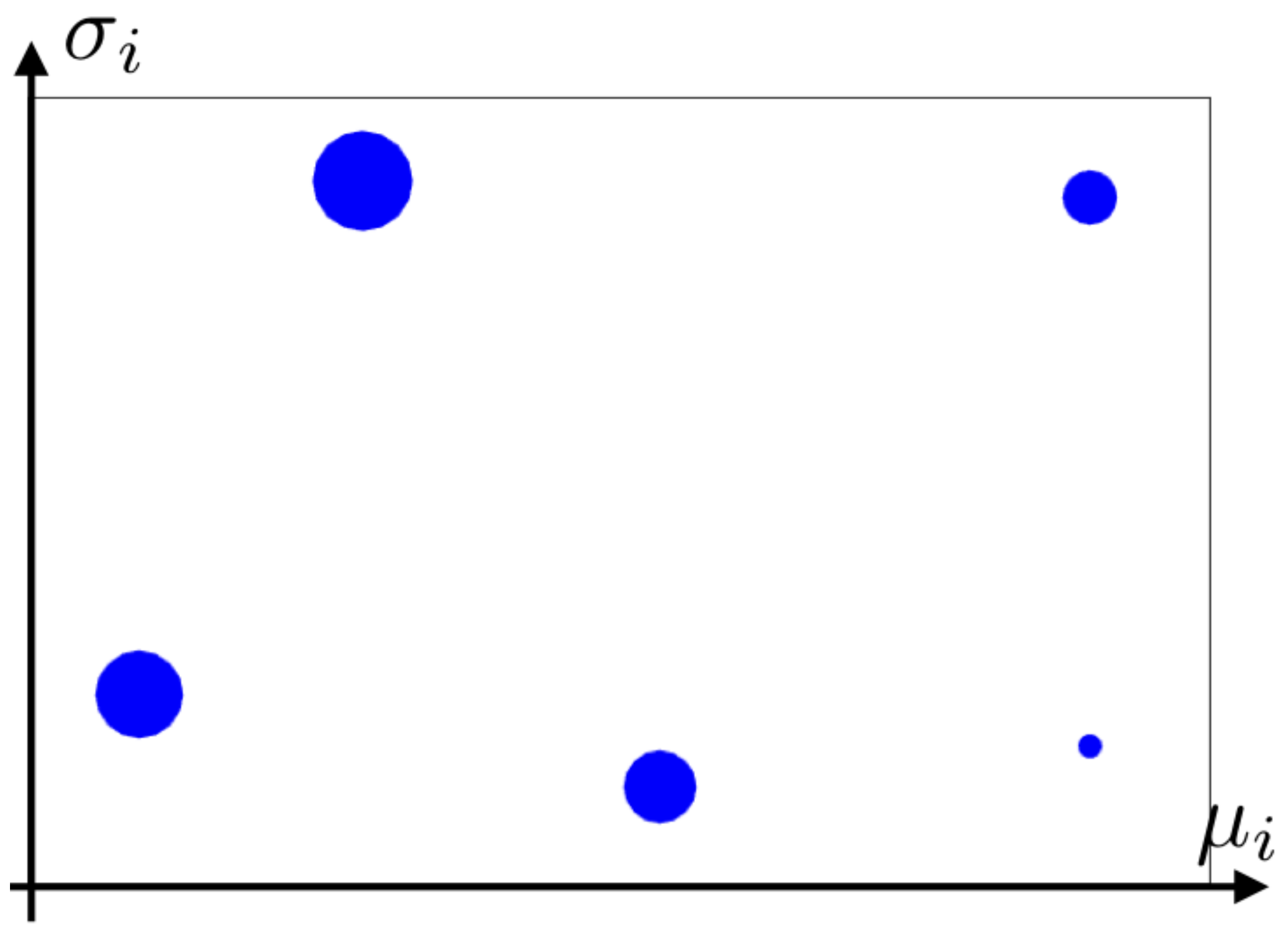
$\theta =$ T1/T2 representing tissue type

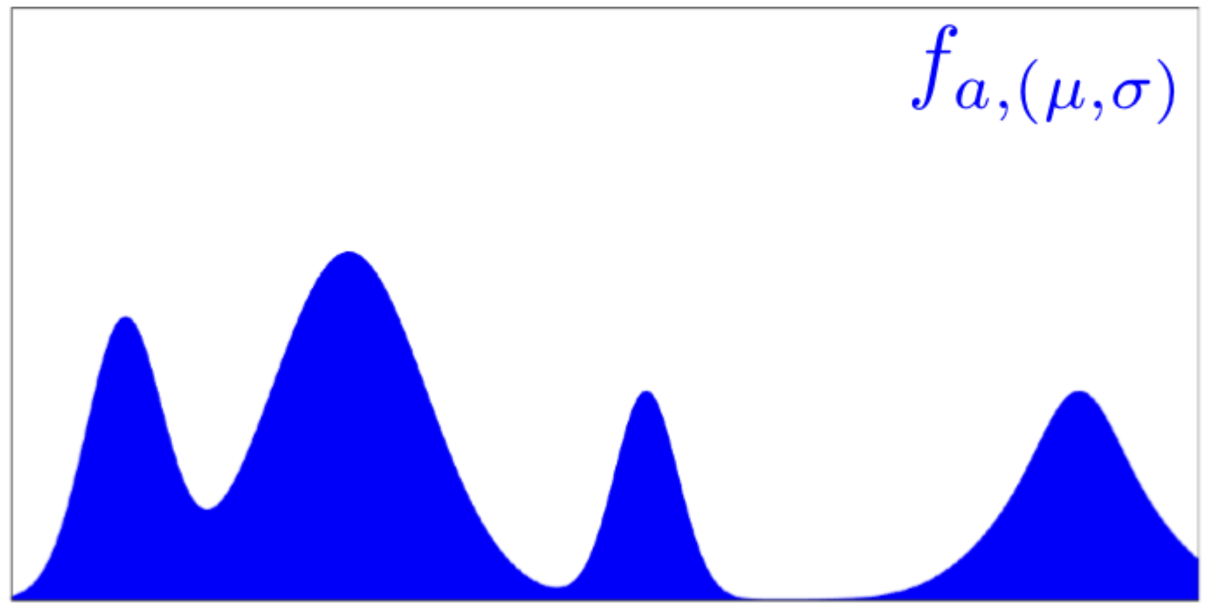
$\phi(\theta) =$ Block response of each tissue

Mixture models

Position/scale : $(z, \sigma) = (\text{mean}, \text{std}) \in \mathcal{X} = \mathbb{R}^d \times \mathbb{R}_+$

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





Convex

$$\min_{\mu \in \mathcal{M}(\mathcal{X})} \|f - \Phi\mu\|_{L^2}$$

Non-Convex

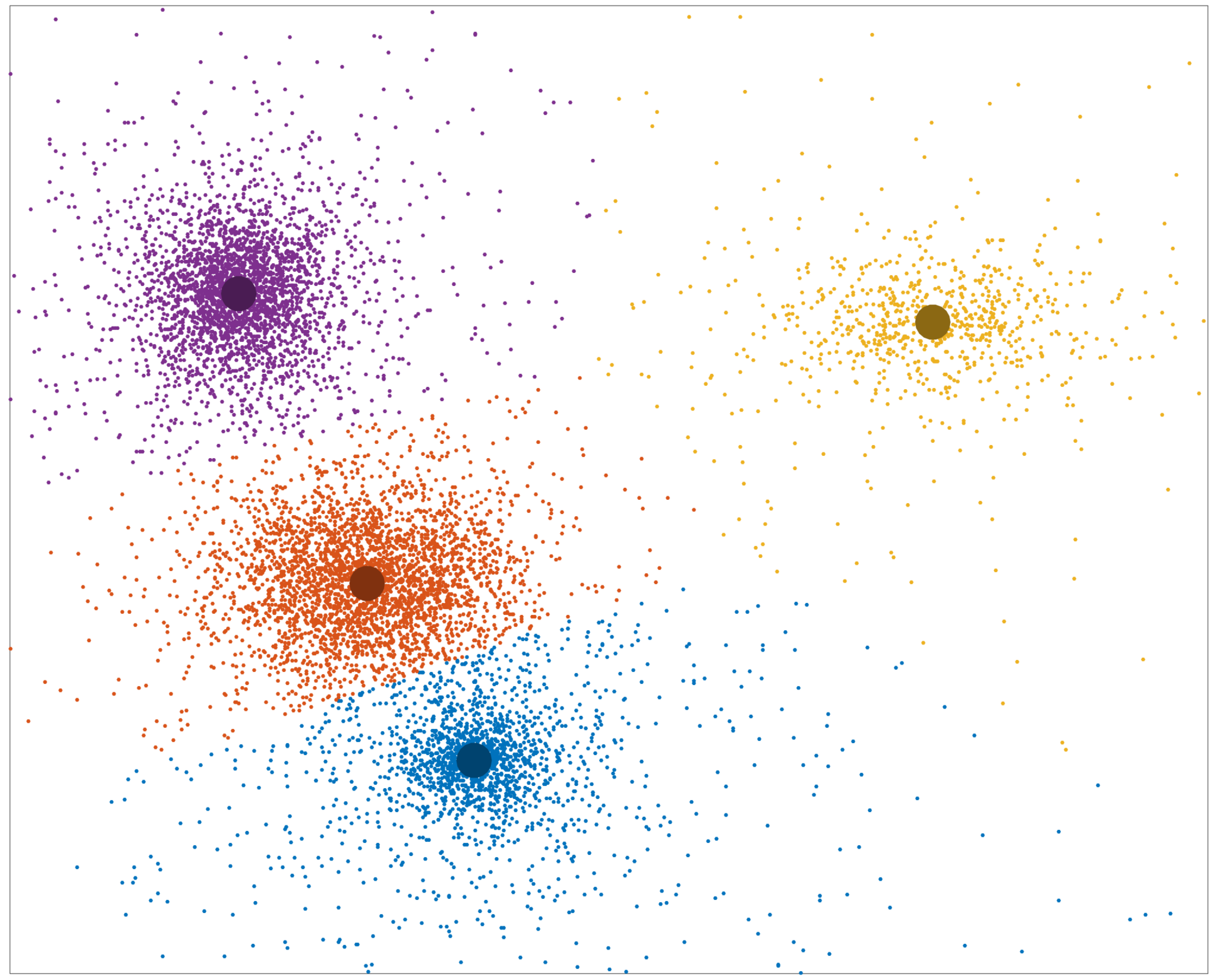
$$\min_{a,z,\sigma} \|f - f_{a,(z,\sigma)}\|_{L^2}$$

$$f_{a,(z,\sigma)} = \Phi\mu = \int_{\mathbb{R}^{d+1}} \phi(x) \mathrm{d}\mu(x)$$

Linear operator

$$\phi(z, \sigma) = h\left(\frac{\cdot - z}{\sigma}\right) \qquad \mu = \sum_{j=1}^{n} a_i \, \delta_{(z_i,\sigma_i)}$$

Density estimation with sketching



Given samples $t_1, t_2, \ldots, t_n$ iid from from density:

$$\bar{\xi}(t) = \sum_{j=1}^{s} a_j \xi(x_j, t) = \int \xi(x,t) \mathrm{d}\mu_{a,x}(x)$$

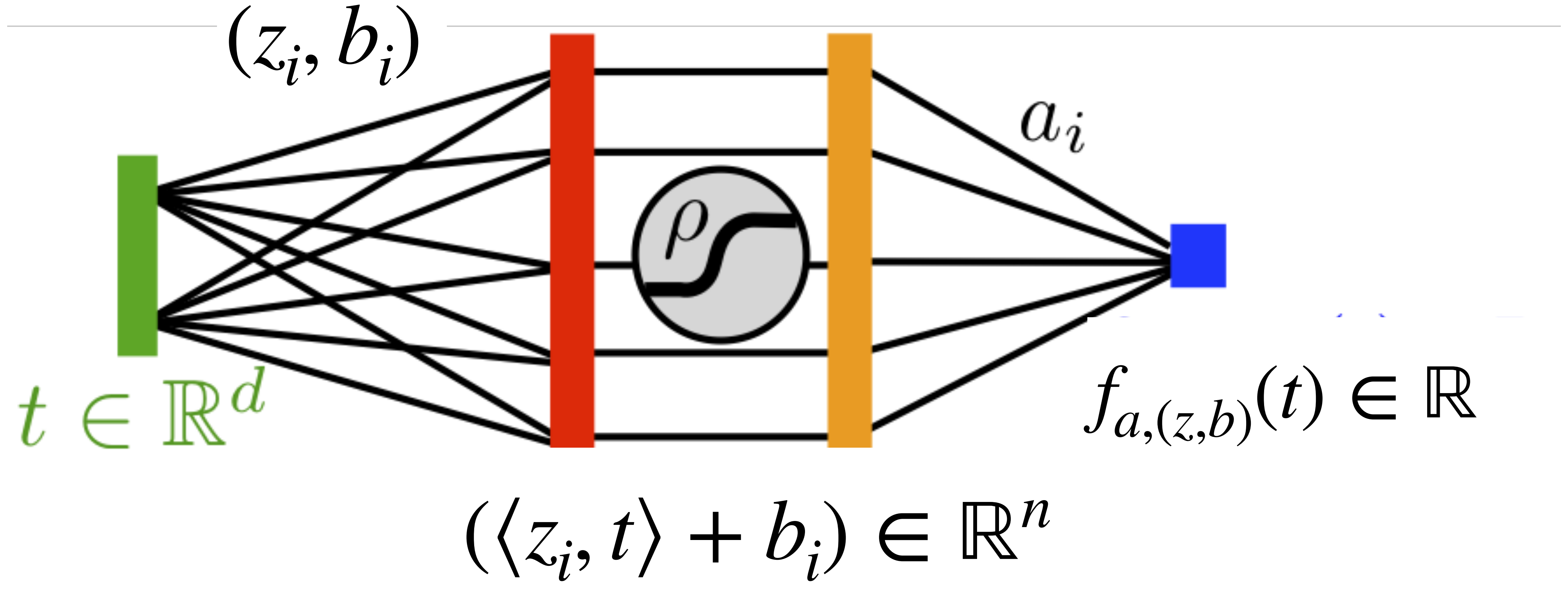
[Gribonval et al 2017]

Sketch using functions g_{ω_k}: $\quad y_k = \frac{1}{n} \sum_{j=1}^{n} g_{\omega_k}(t_j), \; k \in [m]$

Goal: recover a, x from $\quad y_k \approx \int g_{\omega_k}(t)\bar{\xi}(t)\mathrm{d}t = \int_{\mathcal{X}} \underbrace{\int g_{\omega_k}(t)\xi(x,t)\mathrm{d}t}_{\phi_{\omega_k}(x)} \, \mathrm{d}\mu_{a,x}(x)$$

Multi-layer perceptron

For training data $(t_i, y_i)_{i=1,\dots,N}$ fit $f_{a,z,b}(t_i) \approx y_i$



Non-convex

$$\min_{a,z,b} \sum_i |f_{a,z,b}(t_i) - y_i|^2$$

$$f_{a,z,b}(t) = \sum_{i=1}^{n} a_i \rho(\langle z_i, t\rangle + b_i)$$

Convex

$$\min_{\mu \in \mathcal{M}(\mathcal{X})} \|y - \Phi\mu\|^2$$

$$[f_{a,z,b}(t_i)]_i = \Phi\mu = \int_{\mathbb{R}^d} \phi(x)\mathrm{d}\mu(x)$$

$$\phi(x) = \Big[\rho\big(\langle z, t_i\rangle + b\big)\Big]_{i=1,\dots,N} \qquad \mu = \sum_{i=1}^{n} a_i\, \delta_{(z_i,b_i)}$$

Linear operator

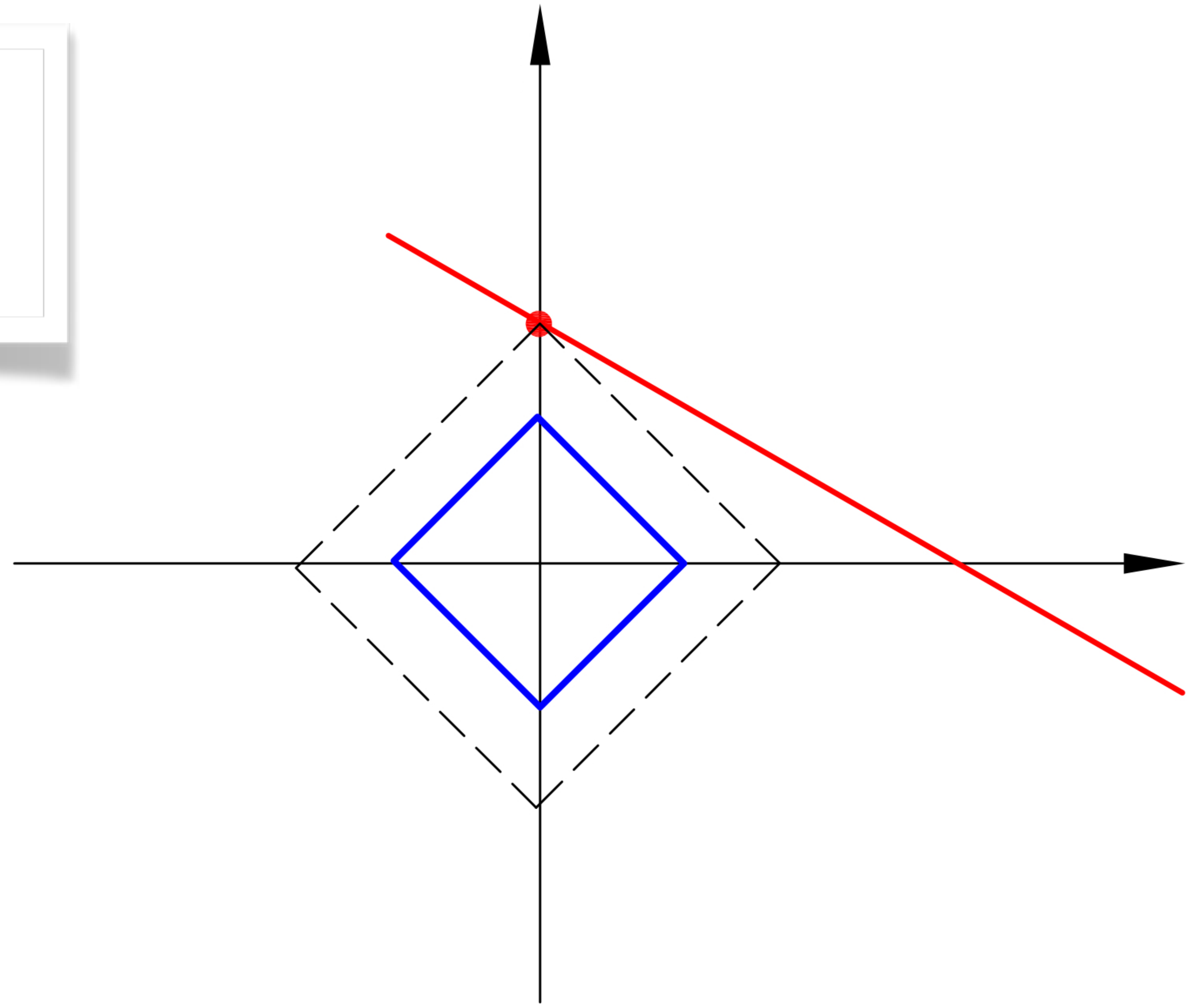
Total variation

$\mathscr{M}(\mathscr{X})$ is a Banach space with norm $||\mu||_{TV}$

$$||\mu||_{TV} = \sup\left\{ \int f(x)\mathrm{d}\mu(x) \; : \; f \in C_0(\mathscr{X}), \; ||f||_\infty \leq 1 \right\}$$

$$f \in L^1(\mathscr{X}), \mathrm{d}\mu(x) = f(x)\mathrm{d}x \longrightarrow ||\mu||_{TV} = \int |f(x)|\,\mathrm{d}x$$

$$\mu = \sum_j a_j \delta_{x_j} \longrightarrow ||\mu||_{TV} = \sum_j |a_j|$$



The extremal points of $\{\mu : ||\mu||_{TV} \leq 1\}$ are $\{\delta_x : x \in \mathscr{X}\}$

The Beurling-Lasso

$$P_\lambda(y) \qquad \inf_{\mu \in \mathcal{M}(\mathcal{X})} \lambda \|\mu\|_{TV} + \frac{1}{2}\|\Phi\mu - y\|^2$$

Relaxation for any K: $$\inf_{a,x} \lambda \sum_{j=1}^{K} |a_j| + \frac{1}{2}\|\sum_{j=1}^{K} \phi(x_j)a_j - y\|^2 \geq \inf P_\lambda(y)$$

Fisher-Jerome (1973):
If $\phi(x) \in \mathbb{R}^m$ with ϕ continuous, then there exists a solution to $P_\lambda(y)$ with at most m Diracs.

The relaxation is tight when $K \geq m$

$$P_0(y) \qquad \inf_{\mu \in \mathcal{M}(\mathcal{X})} \|\mu\|_{TV} \quad \text{s.t.} \quad \Phi\mu = y\,.$$

[Beurling (1973)]

[De Castro and Fabrice (2012)]

[Candès and Fernandez-Granda (2012)]

[Duval and Peyré (2015).]

The Beurling-Lasso

$$P_\lambda(y) \qquad \min_{\mu \in \mathcal{M}(\mathcal{X})} \lambda \|\mu\|_{TV} + \frac{1}{2}\|\Phi\mu - y\|^2$$

The Lasso: Given $y = Xa$, $y \in \mathbb{R}^m$, $X \in \mathbb{R}^{m \times n}$, to recover a sparse vector $a \in \mathbb{R}^n$

$$\min_{a \in \mathbb{R}^n} \frac{1}{2\lambda}\|Xa - y\|^2 + \|a\|_1$$

- Optimisation is over the space of measures (not just Diracs) with no a-priori choice on the number of spikes.
- This is a convex problem, with strong recovery guarantees.
- Some non-convex problems can be placed into this framework

Questions

- When is $\mu_0 = \sum_j a_j \delta_{x_j}$ an exact solution to $(P_0(y))$?
- Are solutions to $P_\lambda(y)$ stable to noise?
- Numerical algorithms in the infinite dimensional space?
- Under what conditions do we recover the exact number of spikes?
- Compressed sensing — if Φ is a random operator, how many measurements to recover?

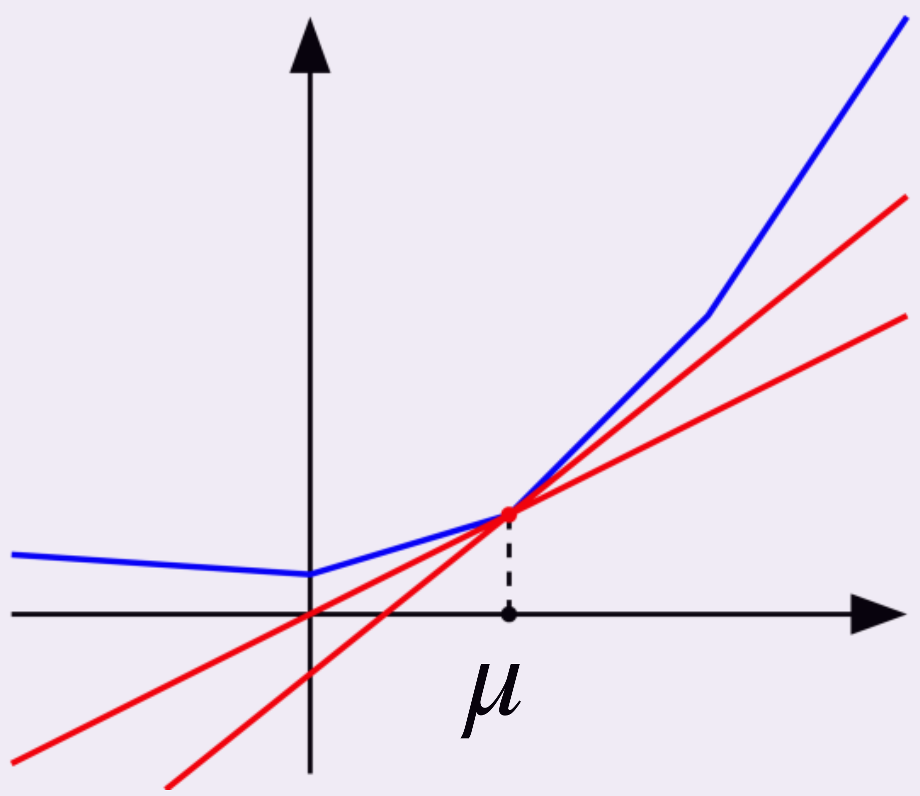
Optimality conditions

$\mu_* \in \text{argmin}_\mu F(\mu)$ $\Longleftrightarrow$ $\nabla F(\mu_*) = 0$

But $\|\mu\|_{TV}$ is not differentiable. Need to consider its **sub-differential**.

Let $\Psi : U \to \mathbb{R}$ be a convex function, its **sub-differential** is:

$$\partial\Psi(\mu) = \left\{ p \in U^* \ : \ \forall \hat{\mu},\ \ \Psi(\hat{\mu}) \geq \Psi(\mu) + \langle p, \hat{\mu} - \mu \rangle \right\}$$

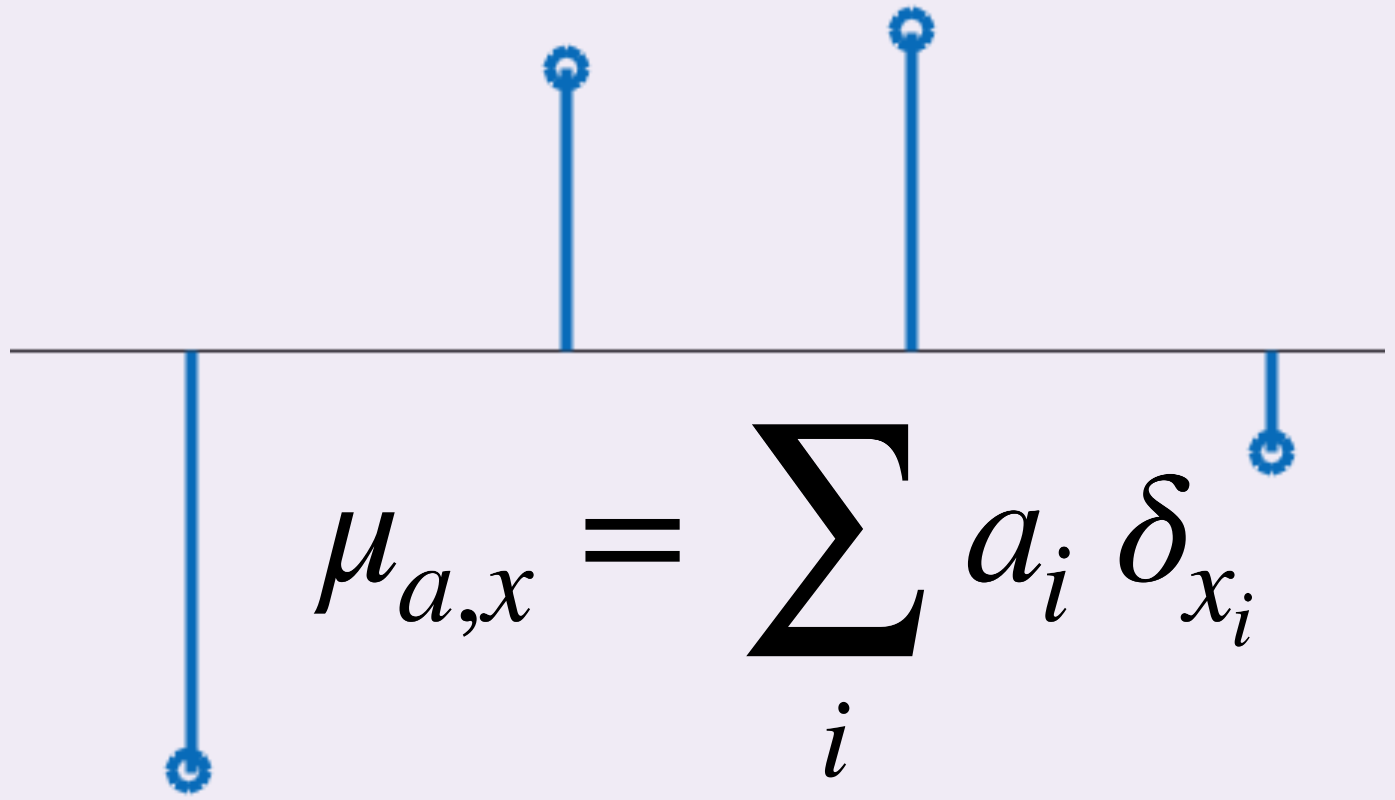


Optimality conditions

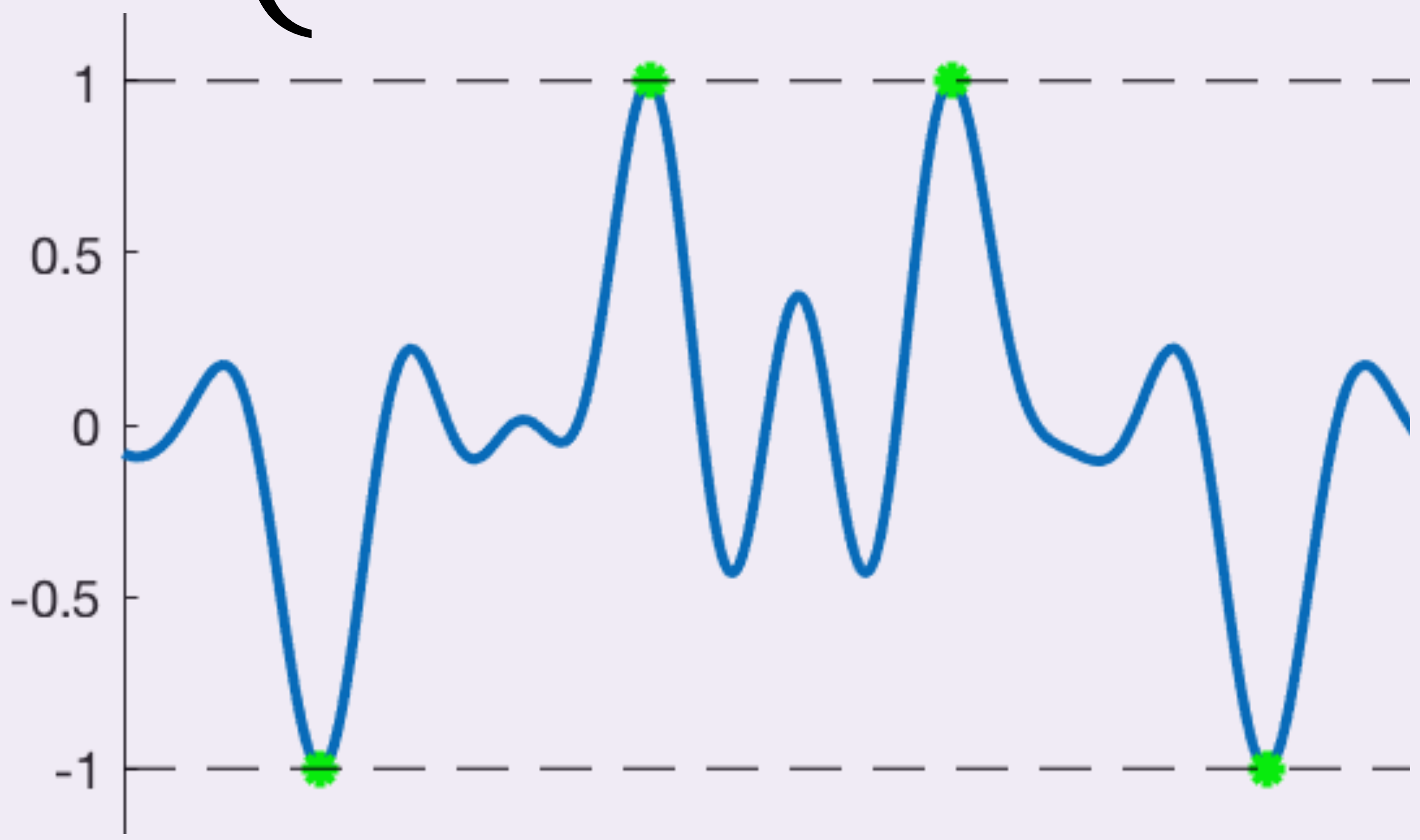
Equivalent characterization for $\|\mu\|_{TV}$: $\partial\|\mu\|_{TV} = \{f \in C(\mathcal{X}) : \ \|f\|_\infty \le 1, \langle f, \mu \rangle = \|\mu\|_{TV}\}$

For sparse measures:

$$\partial\|\mu_{a,x}\|_{TV} = \left\{ f \in C(\mathcal{X}) \ : \ \begin{cases} \|f\|_\infty \le 1 \\ \forall i, \ f(x_i) = \text{sign}(a_i) \end{cases} \right\}$$



$$\mu_{a,x} = \sum_i a_i \, \delta_{x_i}$$



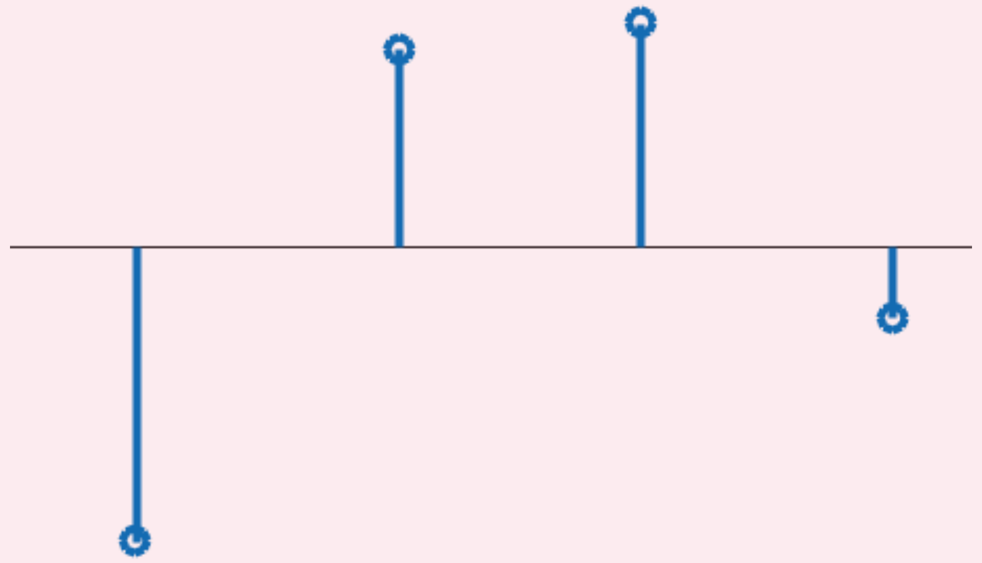
$$f \in \partial\|\mu_{a,x}\|_{TV}$$

Optimality conditions

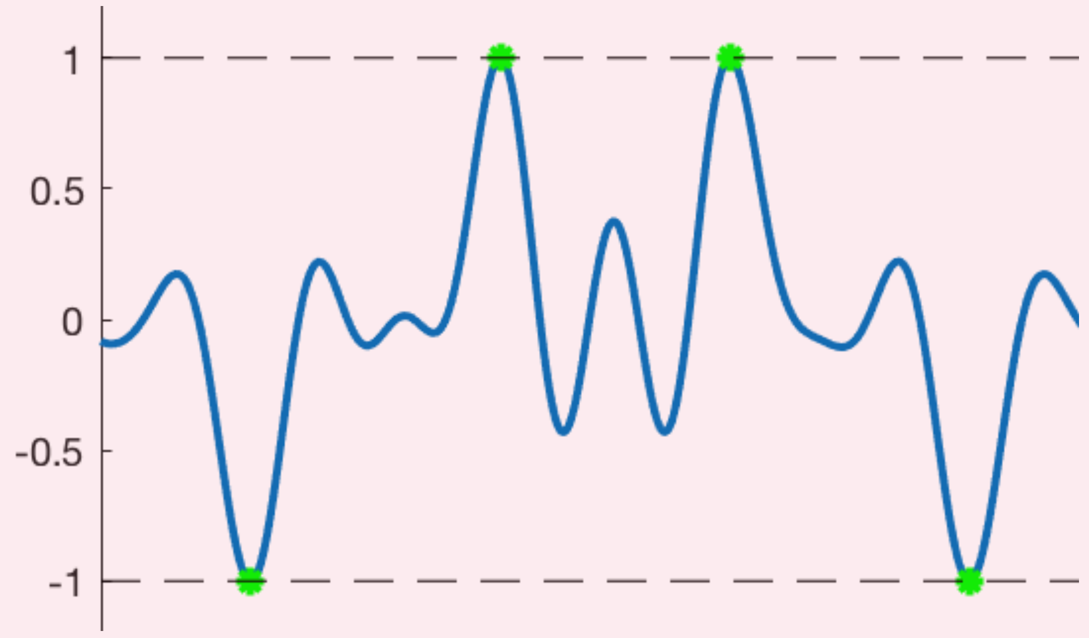
For convex problem $\min_x F(x)$, minimiser iff $0 \in \partial F(x)$

$$\mu_\lambda \in \operatorname*{argmin}_{\mu \in \mathscr{M}(\mathscr{X})} \lambda \|\mu\|_{TV} + \frac{1}{2}\|\Phi\mu - y\|^2 \quad \longleftrightarrow \quad 0 \in \partial \|\mu_\lambda\|_{TV} + \frac{1}{\lambda}\Phi^*(\Phi\mu_\lambda - y)$$

$$\mu_{a,x} = \sum_i a_i \, \delta_{x_i}$$



$$\eta \in \partial \|\mu_{a,x}\|_{TV}$$



$$\eta_\lambda := -\frac{1}{\lambda}\Phi^*(\Phi\mu_\lambda - y) \in \partial \|\mu_\lambda\|_{TV}$$

$$\boxed{\mathrm{Supp}(\mu_\lambda) \subset \{x \;:\; |\eta_\lambda(x)| = 1\}}$$

The *dual certificate* η_λ certifies the support of μ_λ

Convex duality

Primal: $$\min_{\mu \in \mathscr{M}(\mathscr{X})} \|\mu\|_{TV} + \frac{1}{2\lambda}\|\Phi\mu - y\|^2$$

Dual: $$\sup_{\|\Phi^* p\|_\infty \leq 1} \langle p, y \rangle - \frac{\lambda}{2}\|p\|^2 \qquad (D_\lambda(y))$$

Convex duality

Dual: $$\sup_{\|\Phi^* p\|_\infty \le 1} \langle p, y \rangle - \frac{\lambda}{2}\|p\|^2 = -\frac{\lambda}{2}\|p - y/\lambda\|^2 + \frac{1}{\lambda}\|y\|^2$$

Projection onto convex set

- $D_\lambda(y)$ is the projection onto a convex set. So, it has a unique solution.
- If $\mathscr{H} = \mathbb{R}^n$, optimise over finite vector space but with infinite constraints.
- There is strong duality. $\inf P_\lambda(y) = \sup D_\lambda(y)$
- When $\lambda > 0$, solutions to $P_\lambda(y)$ and $D_\lambda(y)$ exist.

The noiseless problem

Primal : $$\min_{\mu \in \mathscr{M}(\mathscr{X})} \|\mu\|_{TV} \quad \text{s.t.} \quad \Phi\mu = y$$

Dual: $$\sup_{\|\Phi^* p\|_\infty \le 1} \langle p, y \rangle$$

- When $\lambda = 0$, only existence of solutions to $P_0(y)$ is guaranteed (unless $\mathscr{H}$ is finite).

Convex duality

μ_λ solves $(P_\lambda(y))$ and p_λ solves $(D_\lambda(y))$

$\Updownarrow$

$\Phi^* p_\lambda \in \partial \|\mu_\lambda\|_{TV}$ and $p_\lambda = -\frac{1}{\lambda}(\Phi\mu_\lambda - y)$

μ_0 solves $P_0(y)$ and p_0 solves $D_0(y)$

$\Updownarrow$

$\Phi^* p_0 \in \partial \|\mu_0\|_{TV}$ and $\Phi\mu_0 = y$

If $p_\lambda = \operatorname{argmax} D_\lambda(y)$ and $\eta_\lambda = \Phi^* p_\lambda$, then $\eta_\lambda \in \partial\|\mu_\lambda\|_{TV}$ means that
$$\mathrm{Supp}(\mu_\lambda) \subset \{x \ : \ |\eta_\lambda(x)| = 1\}$$

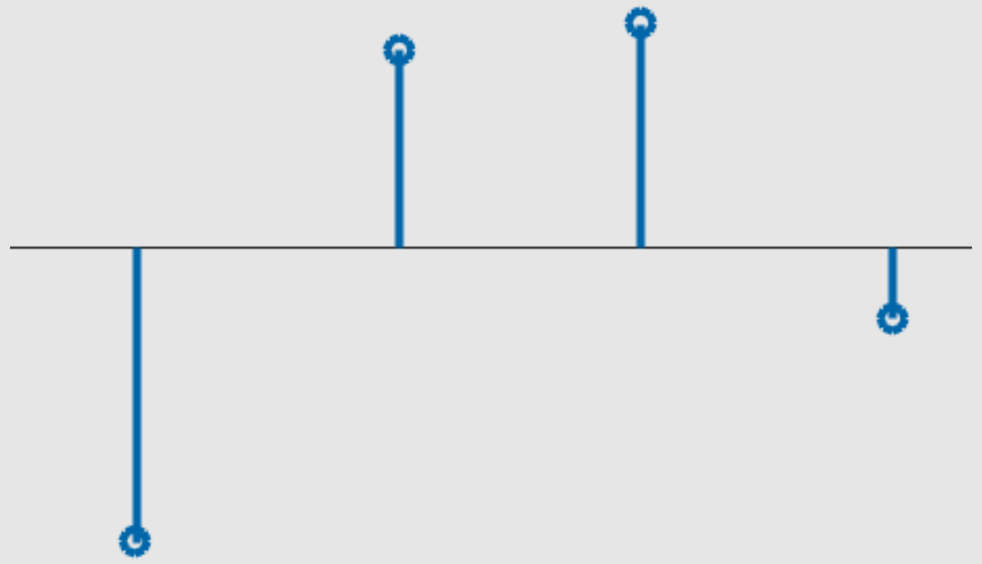
Solutions to $D_0(\Phi\mu_0)$ can tell us about the structure of $\mu_\lambda \in \min P_\lambda(\Phi\mu_0 + w)$

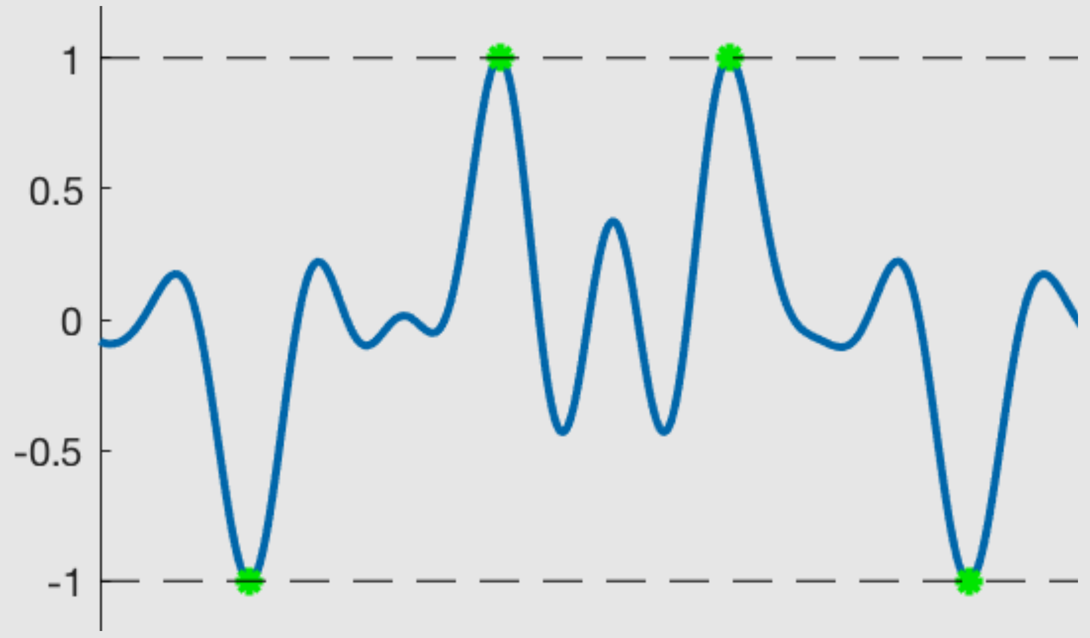
Uniqueness

Theorem:

If $\mu_{a,x} = \sum_j a_j \delta_{x_j}$ and $y = \Phi \mu_{a,x}$ and there exists p such that

- $\eta := \Phi^* p$ satisfies $|\eta(x)| < 1$ for all $x \notin \{x_i\}$
- $\eta(x_i) = \mathrm{sign}(a_i)$ for all i.
- $(\phi(x_i))_i$ are linearly independent.





Then, $\mu_{a,x}$ is the unique solution to $P_0(y)$

Proof: by the primal-dual relationships, any solution has support contained in $\{x_i\}_i$

So, any two solutions take the form: $\mu = \sum_i a_i \delta_{x_i}$ and $\hat{\mu} = \sum_i \hat{a}_i \delta_{x_i}$

We must have $a_i = \hat{a}_i$ since $\Phi\hat{\mu} = \Phi\mu$ and $\phi(x_i)$ are linearly independent.

Stability

Theorem [Azais De Castro & Gamboa (2015)]

Suppose we observe $y = \Phi\mu_{a,x} + w$ with $\|w\| \leq \epsilon$.

In addition to conditions of previous theorem, suppose $\eta = \Phi^* p$ satisfies

i) $|\eta(x)| \leq 1 - c_2\|x - x_i\|^2$ for all $x \in B(x_i, r)$

ii) $|\eta(x)| < 1 - c_o$ for all $x \notin \cup_i B(x_i, r)$

Then, choosing $\lambda \sim \epsilon/\|p\|$, any solution $\hat{\mu}$ to $P_\lambda(y)$ satisfies

$$\underbrace{c_0|\hat{\mu}|(\mathcal{X}\setminus\cup_i B(x_i, r))}_{\text{amplitudes outside neighbourhood of true support is small}} + c_2 \sum_i \int_{B(x_i,r)} \underbrace{\|x - x_i\|^2 \mathrm{d}|\hat{\mu}|(x)}_{\text{Cluster around true support}} \lesssim \epsilon\|p\|$$

Stability

Theorem [Azais De Castro & Gamboa (2015)]

Suppose we observe $y = \Phi\mu_{a,x} + w$ with $\|w\| \leq \epsilon$.

In addition to conditions of previous theorem, suppose $\eta = \Phi^* p$ satisfies

i) $|\eta(x)| \leq 1 - c_2\|x - x_i\|^2$ for all $x \in B(x_i, r)$

ii) $|\eta(x)| < 1 - c_o$ for all $x \notin \cup_i B(x_i, r)$

Then, choosing $\lambda \sim \epsilon/\|p\|$, any solution $\hat{\mu}$ to $P_\lambda(y)$ satisfies

$$c_0|\hat{\mu}|(\mathscr{X}\backslash\cup_i B(x_i, r)) + c_2\sum_i \int_{B(x_i,r)} \|x - x_i\|^2 \mathrm{d}|\hat{\mu}|(x) \lesssim \epsilon\|p\|$$

$$\Rightarrow \quad W_2^2\Big(\sum_j \hat{A}_j \delta_{x_j}, |\hat{\mu}|\Big) \lesssim \epsilon\|p\| \quad \text{and} \quad \max_j |a_j - \hat{a}_j| \lesssim \epsilon\|p\|$$

$$\hat{A}_j = |\hat{\mu}|(B(x_j, r)) \qquad \hat{a}_j = \hat{\mu}(B(x_j, r))$$

If $\eta \in \mathrm{Im}(\Phi^*)$ satisfies (i) and (ii), then we say that it is nondegenerate.

Candidate for a dual certificate

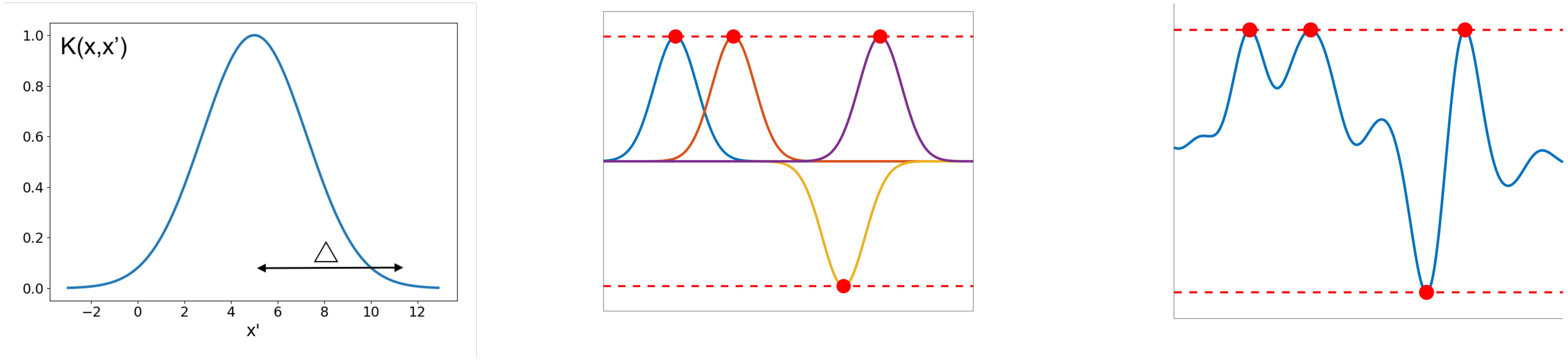
Define:

$$K(x_1, x_2) = \langle \phi(x_1), \phi(x_2) \rangle$$

$$\eta_C(x) = \sum_{i=1}^{n} u_i K(x_i, x) + \sum_{i=1}^{n} v_i \partial_1 K(x_i, x)$$

Want: $\eta(x_i) = \text{sign}(a_i)$ and $\eta'(x_i) = 0$ ← $2n$ equations to solve for $2n$ unknowns in u, v.

Computed η and check if $|\eta(x)| < 1$ for all $x \notin \{x_i\}$.



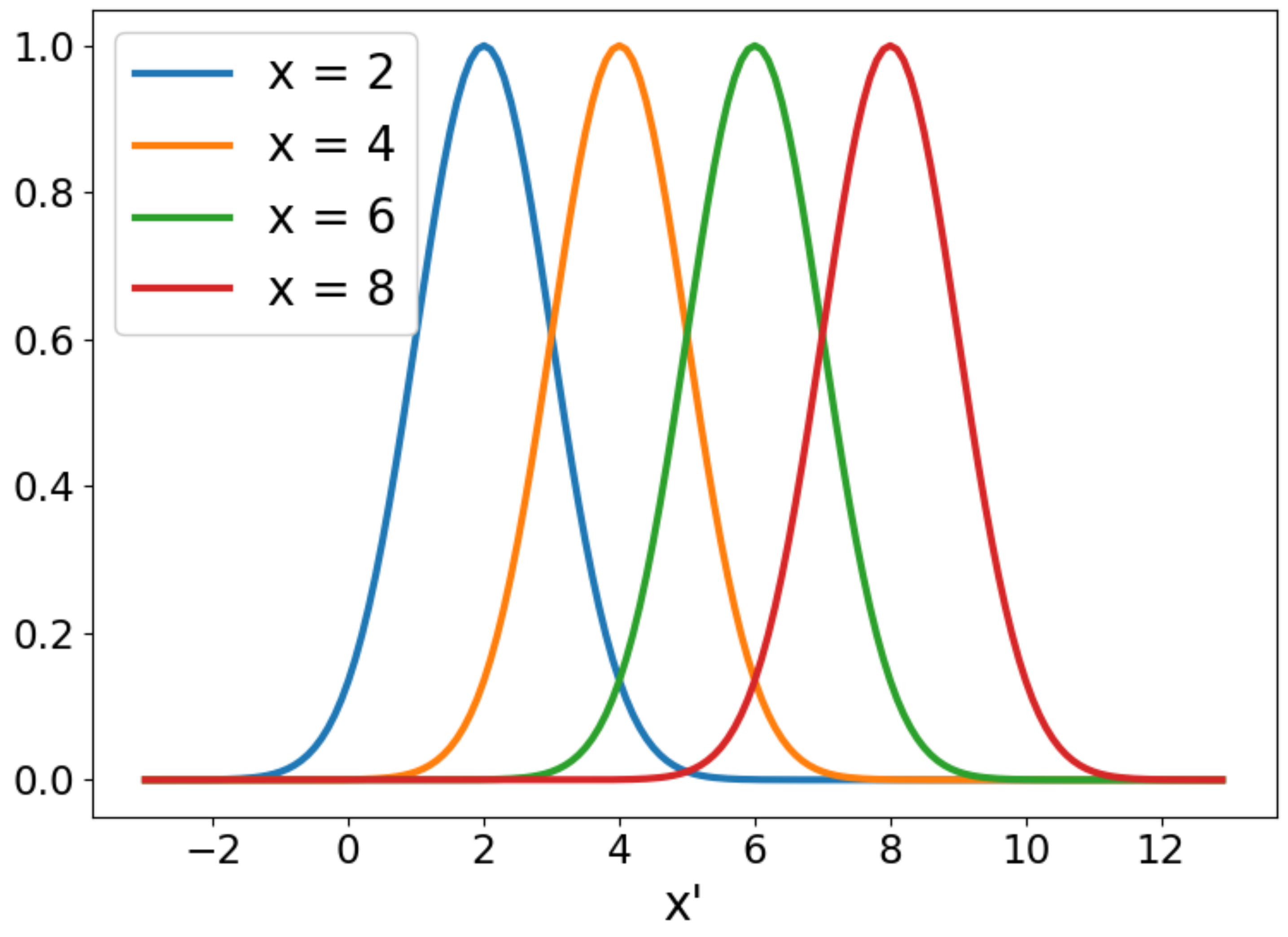
Recovery under minimal separation

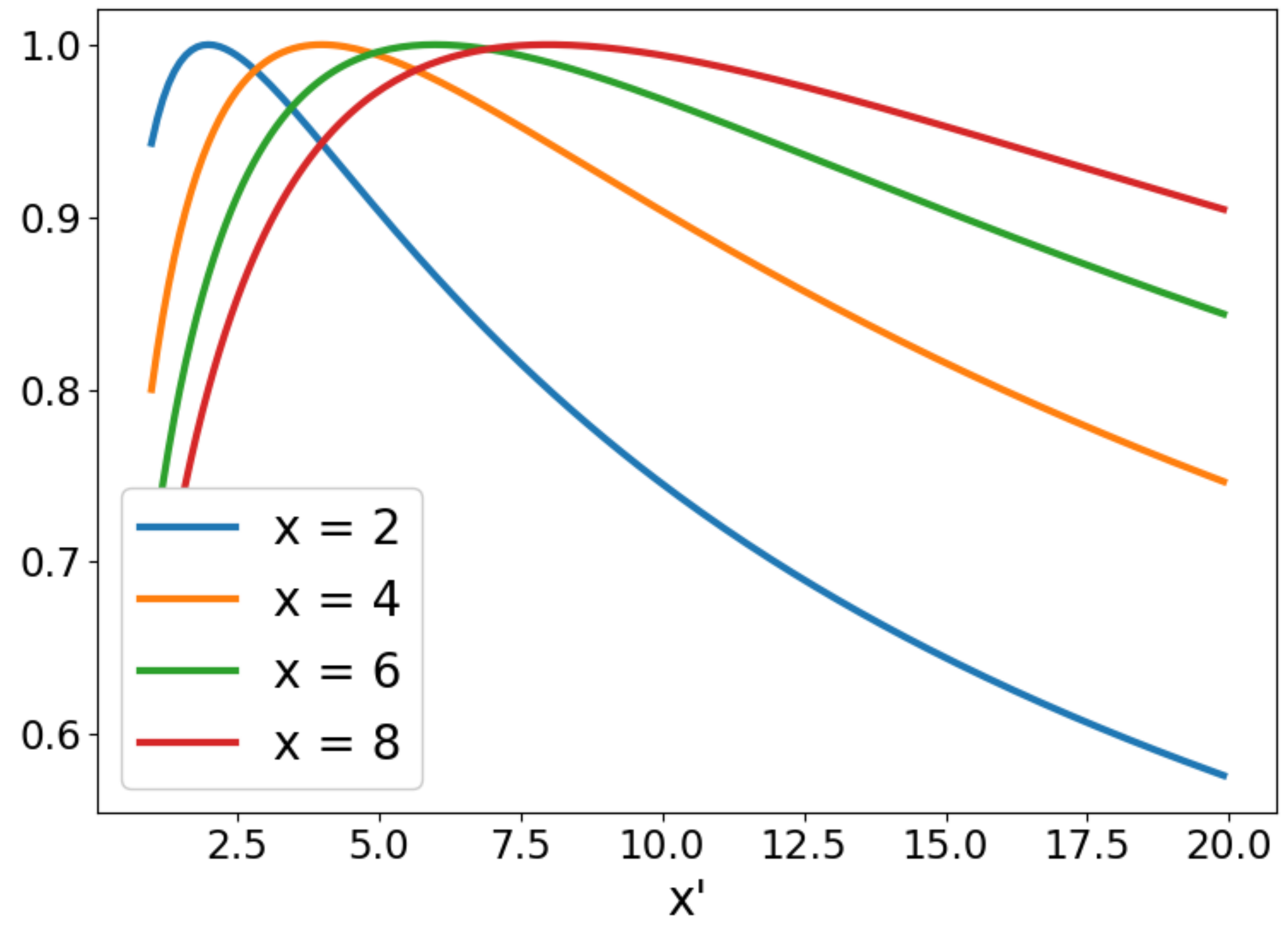
Candès and Fernandez-Granda (2012): Let $\phi(x) = (\exp(2\pi\sqrt{-1}kx)_{|k|\leq f_c}$,

if $\min_{i\neq j} |x_i - x_j| \geq \frac{C}{f_c}$, then η_C is non-degenerate. So, we have stable recovery.

Necessary: If $|x_1 - x_2| < \frac{1}{f_c}$ then $\mu = \delta_{x_1} - \delta_{x_2}$ cannot be recovered by the Blasso

What kind of minimum separation condition to impose for non-translation invariant kernel?

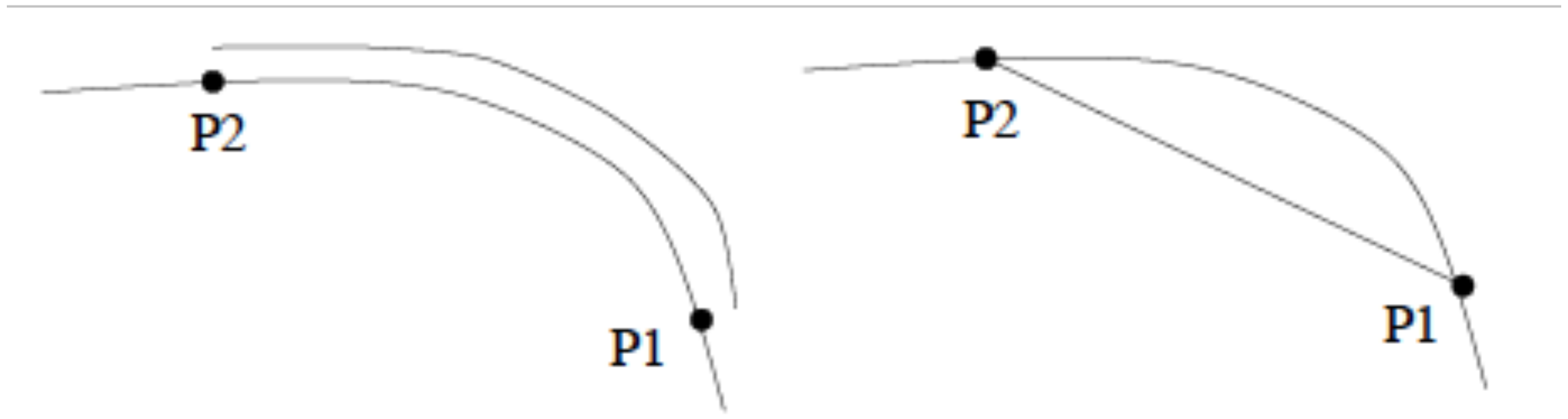




Fisher-Rao distance

Fisher metric: $g_x := \partial_1 \partial_2 K(x, x') = [\nabla \phi(x)][\nabla \phi(x')]^\top \in \mathbb{R}^d$

Fisher-Rao geodesic distance: $d_g(x, x') := \inf_{\gamma: x \to x'} \int_0^1 \sqrt{\langle g_{\gamma(t)} \gamma'(t), \gamma'(t) \rangle} \mathrm{d}t$



Interpretation:

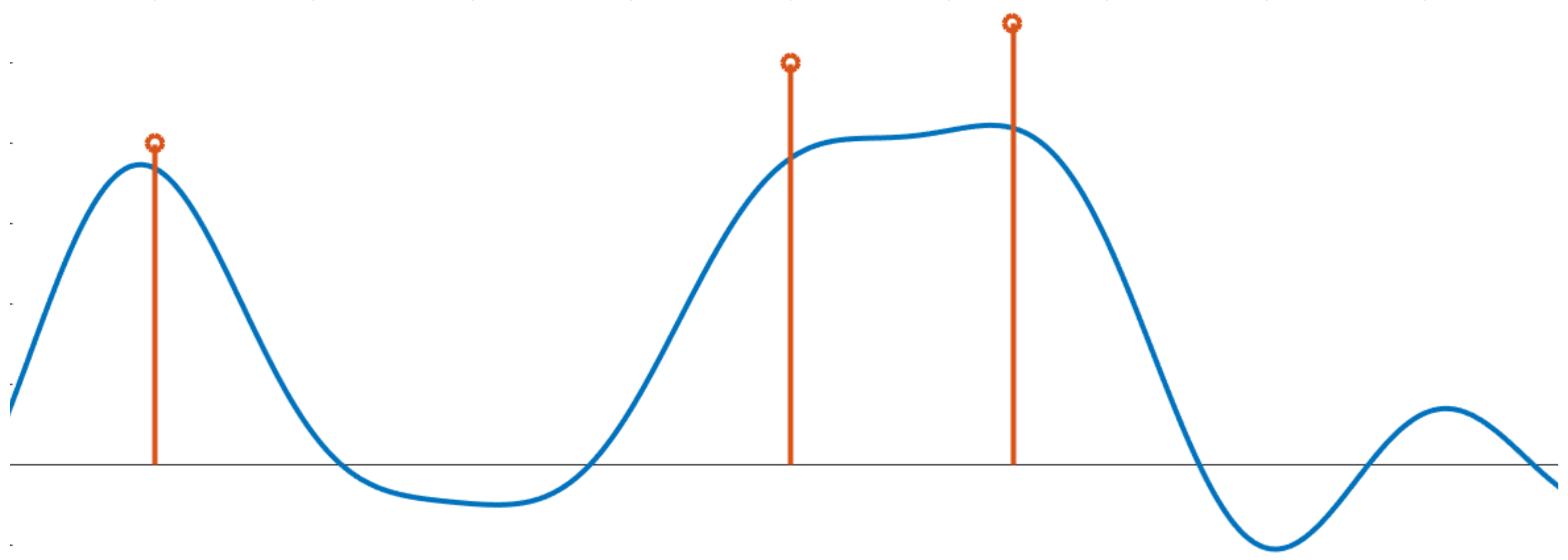
$x \mapsto \phi(x)$ embeds $\mathscr{X}$ into the sphere in $\mathscr{H}$ and

$$d_g(x, x') = \inf_{\gamma: \phi(x) \to \phi(x')} \int_0^1 \|\gamma'(t)\|_{\mathscr{H}} \mathrm{d}t$$

Examples

Poon, Keriven and Peyre *(2019):* If $\min_{i \neq j} d_g(x_i, x_j) \geq \Delta_{s,K}$, then η_C is nondegenerate.

Gaussian	**Fourier**	**Laplace**
$\phi(x) \propto \exp(-\|x - \cdot\|_\Sigma^2)$	$\phi(x) = (\exp(2\pi\sqrt{-1}kx))_{\|k\|_\infty \leq f_c}$	$\phi(x) \propto \exp(-x\cdot)$
$g_x = \Sigma$	$g_x = f_c I$	$g_x = \mathrm{diag}(1/x_i)$
$d_g(x, x') = \|x - x'\|_\Sigma$	$d_g(x, x') \propto f_c \|x - x'\|_2$	$d_g(x, x') = \sqrt{\sum_i \lvert\log(x_i) - \log(x_i')\rvert^2}$
$\Delta = \sqrt{\log(s)}$	$\Delta = \sqrt{d\sqrt{s}}$	$\Delta = d + \log(ds)$



Summary

$$\min_{\mu \in \mathscr{M}(\mathscr{X})} \lambda \|\mu\|_{TV} + \frac{1}{2}\|\Phi\mu - y\|^2$$

Convex relaxation

$$\inf_{a,x} \lambda \sum_{j=1}^{K} |a_j| + \frac{1}{2}\| \sum_{j=1}^{K} \phi(x_j) a_j - y\|^2$$

Non-convex

$$\sup_{\|\Phi^* p\|_\infty \le 1} \langle p, y\rangle - \lambda \|p\|^2$$

Dual

To assess the recovery of $m_{a,x}$,

Find $\eta = \Phi^* p \in C(\mathscr{X})$ such that

$\eta(x_i) = \mathrm{sign}(a_i)$ and $|\eta(x)| < 1$ for all $x \notin \{x_i\}$

Provided that spikes are sufficiently separated:

- Exact recovery in the noiseless setting
- Stable recovery in the noisy setting.

References

General theory for the Blasso:

- Candès, E. J., & Fernandez-Granda, C. (2014). Towards a mathematical theory of super-resolution. *Communications on pure and applied Mathematics*, *67*(6), 906-956.
- Azais, J. M., De Castro, Y., & Gamboa, F. (2015). Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, *38*(2), 177-195.
- Bredies, K., & Pikkarainen, H. K. (2013). Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations*, *19*(1), 190-218.
- Duval, V., & Peyré, G. (2015). Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, *15*(5), 1315-1355.
- Poon, C., Keriven, N., & Peyré, G. (2021). The geometry of off-the-grid compressed sensing. *Foundations of Computational Mathematics*, 1-87

A few references for applications

- Denoyelle, Q., Duval, V., Peyré, G., & Soubies, E. (2019). The sliding Frank–Wolfe algorithm and its application to super-resolution microscopy. *Inverse Problems*, *36*(1), 014001.
- Gribonval, R., Blanchard, G., Keriven, N., & Traonmilin, Y. (2021). Compressive statistical learning with random feature moments. *Mathematical Statistics and Learning*, *3*(2), 113-164
- Bach, F. (2017). Breaking the curse of dimensionality with convex neural networks. *The Journal of Machine Learning Research*, *18*(1), 629-681.
- Golbabaee & Poon (2022). An off-the-grid approach to magnetic resonance fingerprinting. Inverse Problems (to appear).