Mini-course on
Sparse estimation off-the-grid

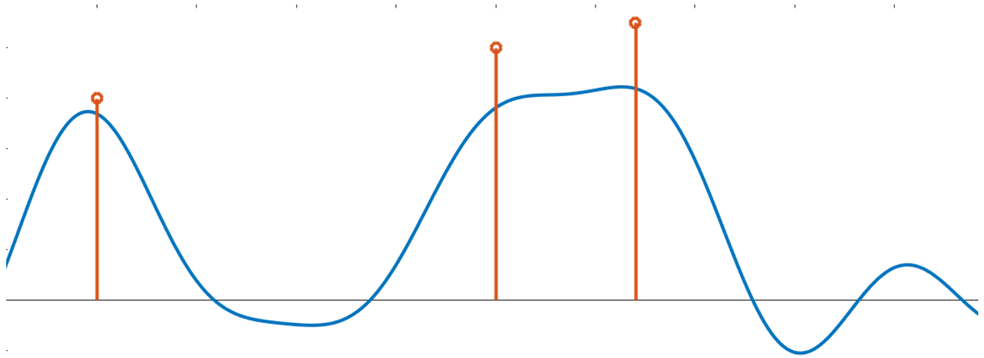
Clarice Poon

Sparse Estimation

Recovering point wise sources from low resolution data

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

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

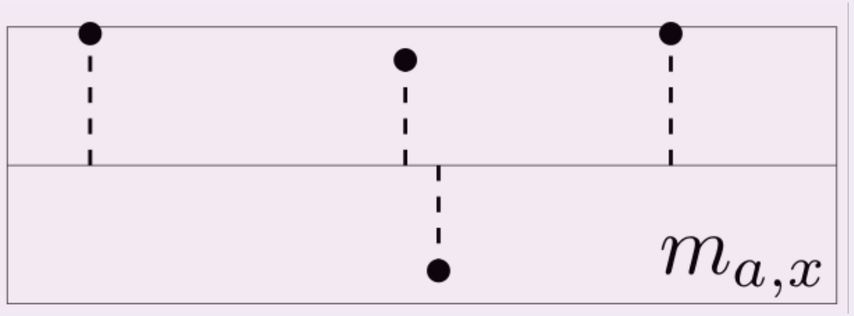


$$y = \sum_{j=1}^{s} a_j [\exp(2\pi\sqrt{-1}\, x_j k)]_k$$

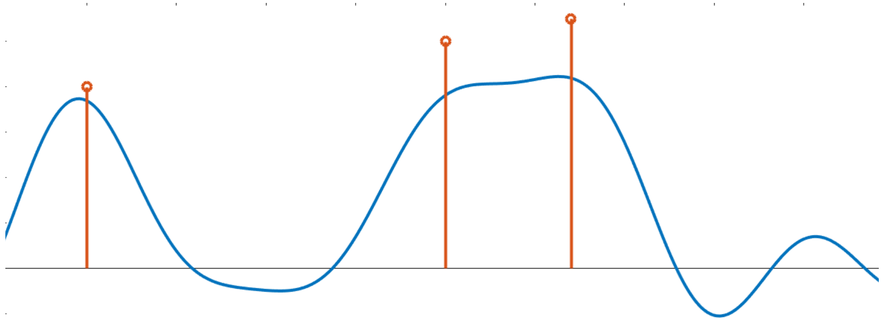
Sparse Estimation

Recovering point wise sources from low resolution data

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



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



Observe linear measurements:

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

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

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

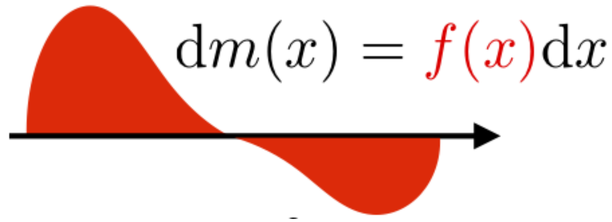
NB: $$\Phi\mu_{a,x} = \sum_{i=1}^{n} a_i \, \phi(x_i)$$

Radon measures

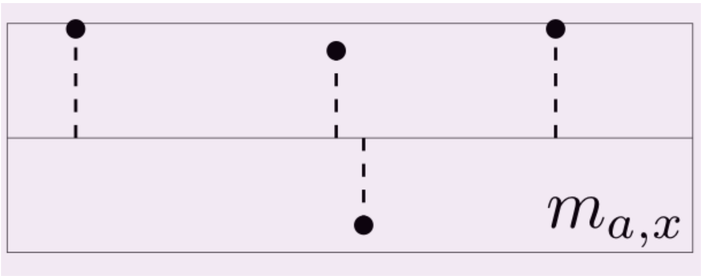
The space of Radon measures $\mathscr{M}(\mathscr{X})$ is the dual of

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

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 $\mu = \sum_j a_j \delta_{x_j}$, $\quad \langle \phi, \mu \rangle = \sum_j \phi(x_j) a_j$

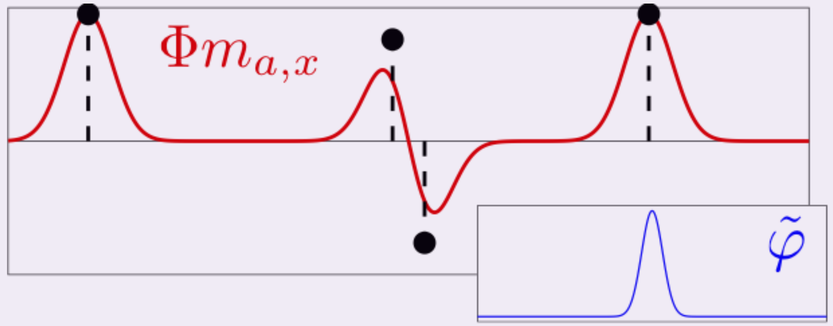


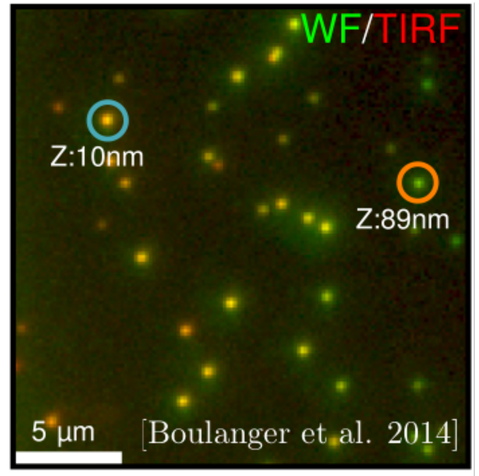
Signal/image processing

Deconvolution:

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

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





Laplace:

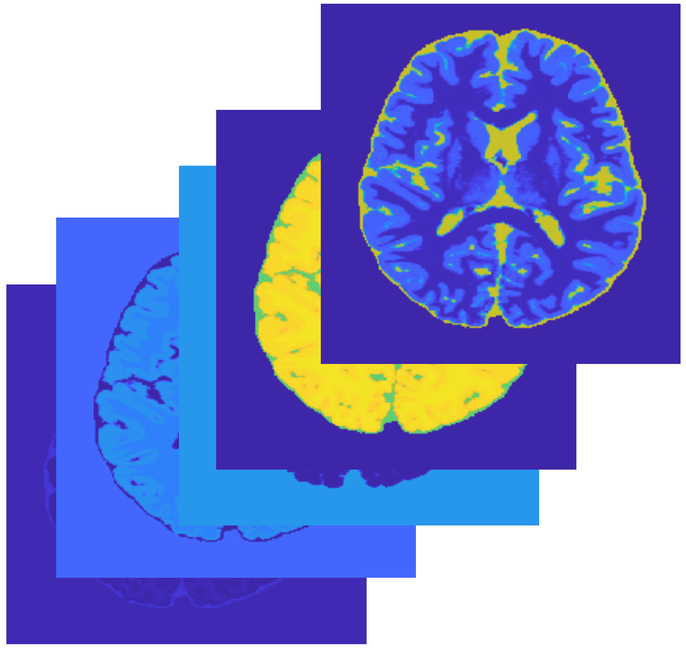
$\phi(x) = \exp(-\langle x, \cdot \rangle) \in L^2(\mathbb{R}^d_+)$

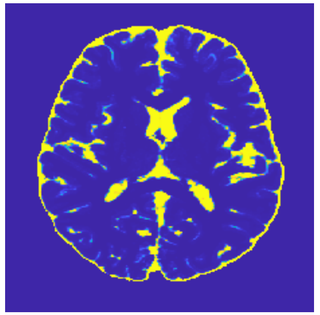
Fourier:

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

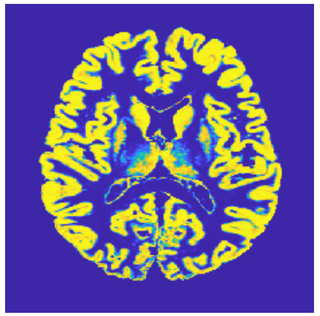
Quantitative MRI

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

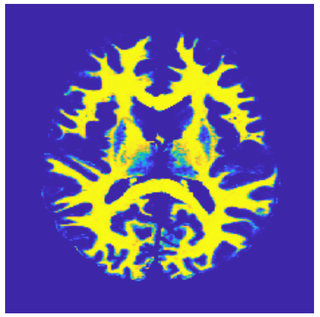




θ_1



θ_2



θ_3

$\theta =$ T1/T2 representing tissue type

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

Time series measurements at voxel v:

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

Recover the NMR properties

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

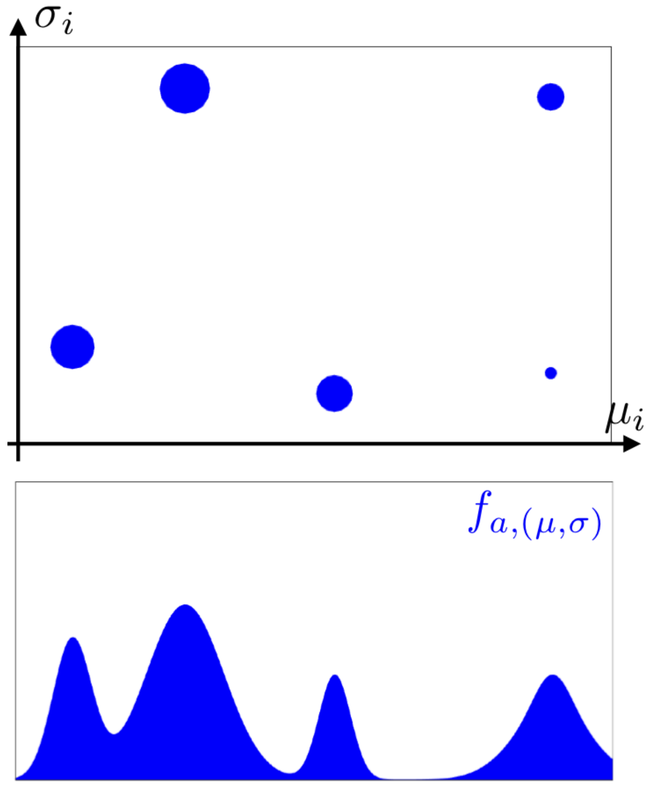
$\mathcal{X}$ are parameters corresponding to different NMR properties.

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

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

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

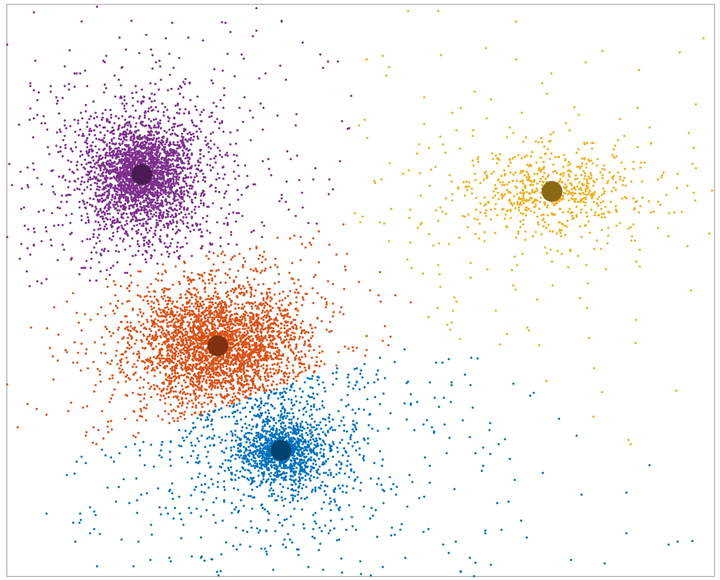
$$f_{a,(z,\sigma)} = \Phi\mu = \int_{\mathbb{R}^{d+1}} \phi(x) \mathrm{d}\mu(x)$$

Linear operator

$$\phi(z, \sigma) = h\left(\frac{\cdot - z}{\sigma}\right)$$

$$m = \sum_{j=1}^{n} a_i \, \delta_{(z_i, \sigma_i)}$$

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 θ_{ω_k}: $y_k = \frac{1}{n} \sum_{j=1}^{n} \theta_{\omega_k}(t_j),\ k \in [m]$

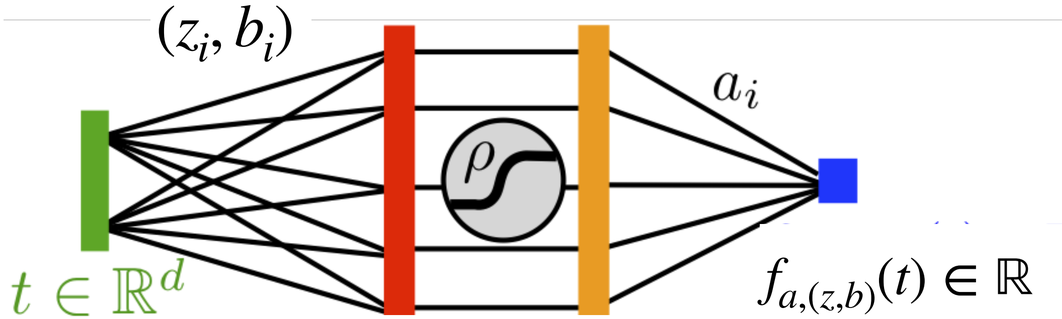
Goal: recover a, x from $y_k \approx \int \theta_{\omega_k}(t) \bar{\xi}(t) \mathrm{d}t = \int_{\mathcal{X}} \underbrace{\int \theta_{\omega_k}(t) \xi(x, t) \mathrm{d}t}_{\phi_{\omega_k}(x)} \mathrm{d}\mu_{a,x}(x)$

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

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



$$(\langle z_i, t\rangle + b_i) \in \mathbb{R}^n$$

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

Convex

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

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

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

Linear operator

Total variation

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

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

- For $f \in L^1(\mathscr{X})$, if $\mathrm{d}\mu(x) = f(x)\mathrm{d}x$ then $\|\mu\|_{TV} = \int |f(x)|\,\mathrm{d}x$
- For $\mu = \sum_j a_j \delta_{x_j}$, $\|\mu\|_{TV} = \sum_j |a_j|$

The Beurling-Lasso

[Beurling (1973)]

[De Castro and Fabrice (2012)]

[Candès and Fernandez-Granda (2012)]

[Duval and Peyré (2015).]

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

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 $K \geq m$!

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

The Beurling-Lasso

[Beurling (1973)]

[De Castro and Fabrice (2012)]

[Candès and Fernandez-Granda (2012)]

[Duval and Peyré (2015).]

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

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!

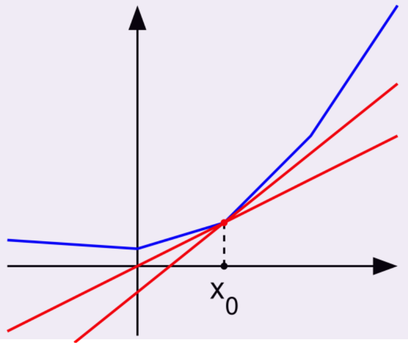
Questions

- 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

$\|\mu\|_{TV}$ is not differentiable, consider its sub differential:

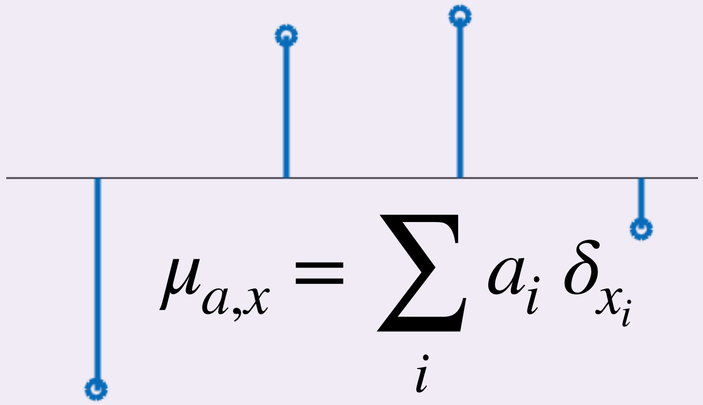
$$\partial\|\mu\|_{TV} = \left\{ f \in C(\mathcal{X}) \;:\; \forall \hat{\mu},\;\; \|\hat{\mu}\|_{TV} \geq \|\mu\|_{TV} + \langle f, \hat{\mu} - \mu \rangle \right\}$$



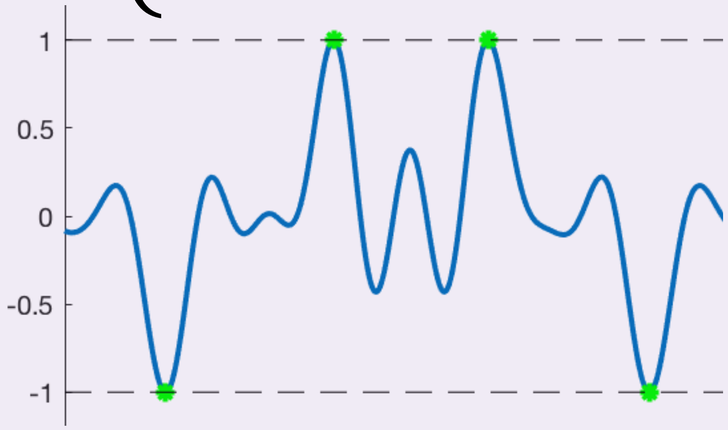
Equivalent characterization: $\partial\|\mu\|_{TV} = \{ f \in \mathscr{C}(\mathcal{X}) :\; \|f\|_\infty \leq 1, \langle f, \mu \rangle = \|\mu\|_{TV} \}$

For sparse measures:

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



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



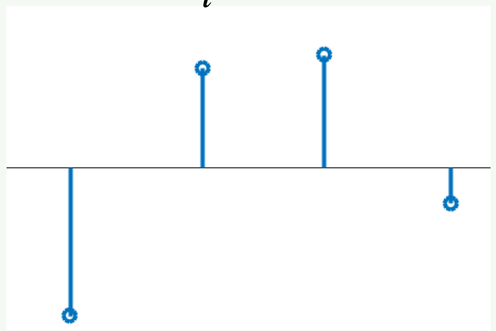
$$f \in \partial\|\mu_{a,x}\|_{TV}$$

Optimality conditions

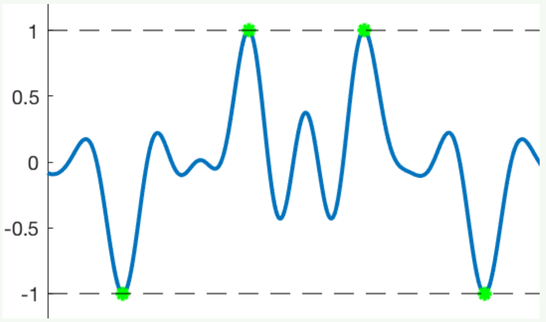
For convex problem $\min_x F(x)$, minimiser if $0 \in \partial F(x)$

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

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



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



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

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

The *dual certificate* η_λ certifies the support of μ_λ

Convex duality

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

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

$$\sup_{\|f\|_\infty \leq 1} \langle p, y \rangle - \frac{\lambda}{2}\|p\|^2 \quad \text{s.t.} \quad f = \Phi^* p$$

$$= \sup_{p, \|f\|_\infty \leq 1} \inf_{\mu \in \mathcal{M}(\mathcal{X})} \langle p, y \rangle - \frac{\lambda}{2}\|p\|^2 + \langle f - \Phi^* p, \mu \rangle \qquad C_0(\mathcal{X})^* = \mathcal{M}(\mathcal{X})$$

$$\leq \inf_{\mu \in \mathcal{M}(\mathcal{X})} \sup_{p, \|f\|_\infty \leq 1} \langle p, y \rangle - \frac{\lambda}{2}\|p\|^2 + \langle f - \Phi^* p, \mu \rangle$$

$$= \inf_{\mu \in \mathcal{M}(\mathcal{X})} \sup_{p} -\frac{\lambda}{2}\|p\|^2 + \|\mu\|_{TV} - \langle p, \Phi\mu - y \rangle \qquad \|\mu\|_{TV} = \sup_{\|f\|_\infty \leq 1} \langle f, \mu \rangle$$

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

Convex duality

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

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

$$\min_{\mu \in \mathscr{M}(\mathscr{X})} \|\mu\|_{TV} \quad \text{s.t.} \quad \Phi\mu = y \qquad\qquad \sup_{\|\Phi^* p\|_\infty \leq 1} \langle p, y \rangle \qquad (D_0(y))$$

- 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 $(D_\lambda(y))$

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

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

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

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

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

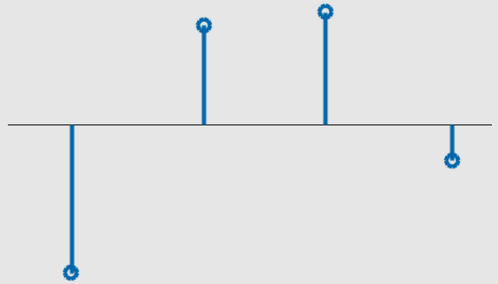
Solutions to $D_0(\Phi\mu_0)$ can tell us about the structure of $\mu_\lambda \in \min P_\lambda(\Phi\mu_0 + w)$

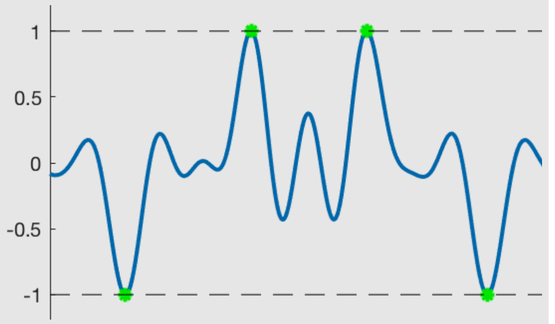
Uniqueness

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 $x \notin \{x_i\}$
- $\eta(x_i) = \operatorname{sign}(a_i)$ 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

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

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

Support outside neighbourhood of true support is small

Cluster around true support

Stability

Theorem [Azais De Castro & Gamboa (2015)]

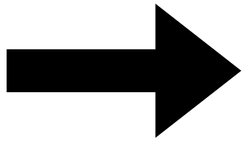
Suppose we observe $y = \Phi\mu_{a,x} + w$ with $\|w\| \leq \epsilon$.

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

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

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



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

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

Candidate for a dual certificate

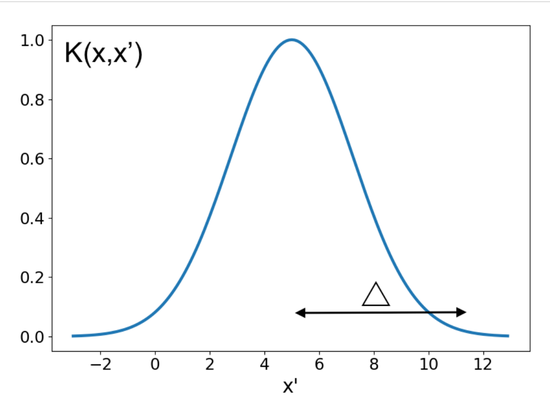
Define:

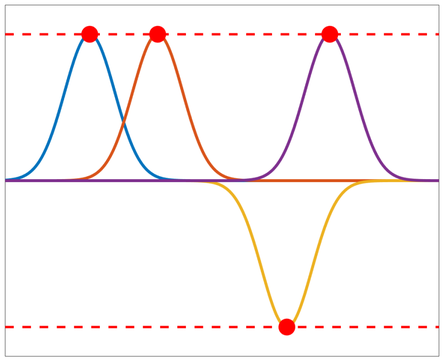
$$K(x_1, x_2) = \langle \phi(x_1), \phi(x_2) \rangle$$

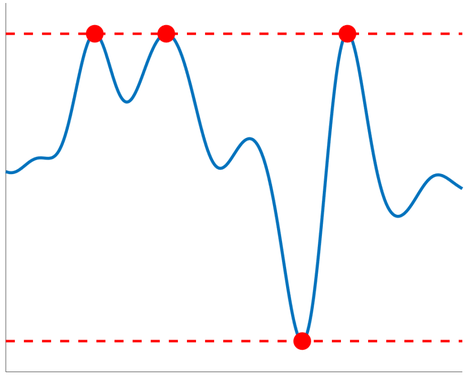
$$\eta_C(x) = \sum_{i=1}^{n} u_i K(x_i, x) + \sum_{i=1}^{n} v_i \partial_1 K(x_i, x)$$

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

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





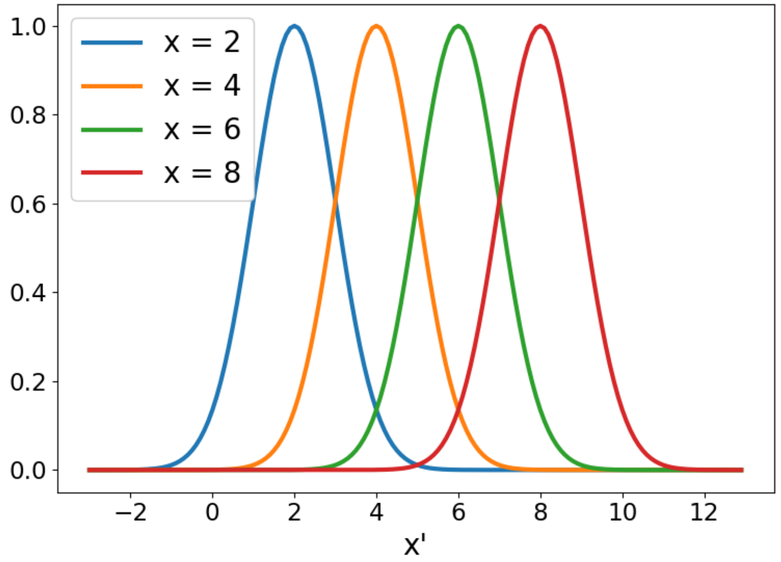


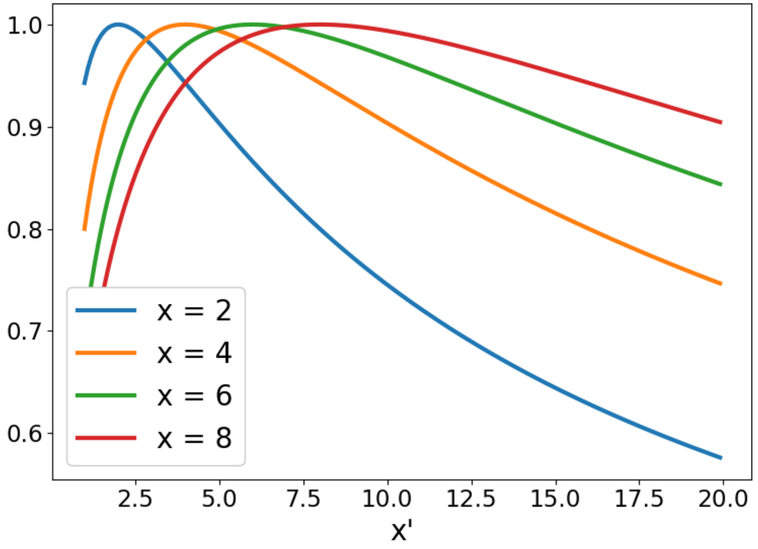
Recovery under minimal separation

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_V is non-degenerate. So, we have stable recovery.

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

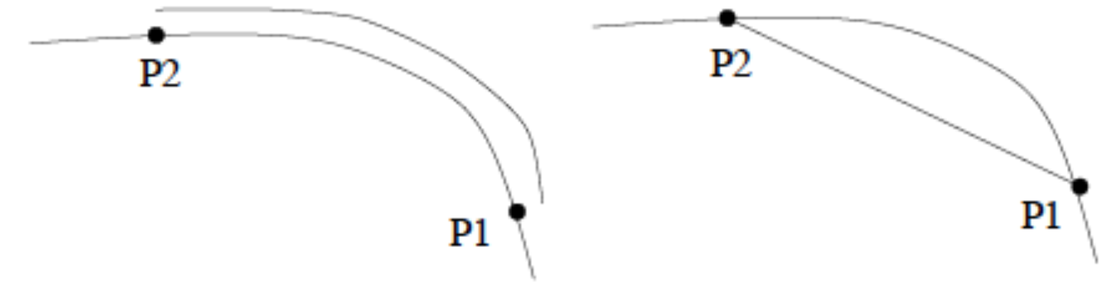




Fisher-Rao distance

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



Interpretation:

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

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

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

Summary

Convex relaxation

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

Non-convex

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

Dual.

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

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

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

$\eta(x_i) = \text{sign}(a_i)$ 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).

Algorithms for the Blasso

Forward-Backward splitting

$$\min_x F(x) := f(x) + g(x)$$

Assume that f is differentiable

Forward-Backward splitting:

$$\begin{cases} \hat{x}_{k+1} &= x_k - \tau \nabla f(x_k) \\ x_{k+1} &= \mathrm{Prox}_{\tau g}(\hat{x}_{k+1}) := \mathrm{argmin}_z \frac{1}{2\tau}\|z - \hat{x}_{k+1}\|^2 + g(z) \end{cases}$$

Assume:

- f, g convex
- $\|\nabla f(x) - \nabla f(x')\| \leq L\|x - x'\|$

→

Convergence rates: If $\tau = 1/L$, then

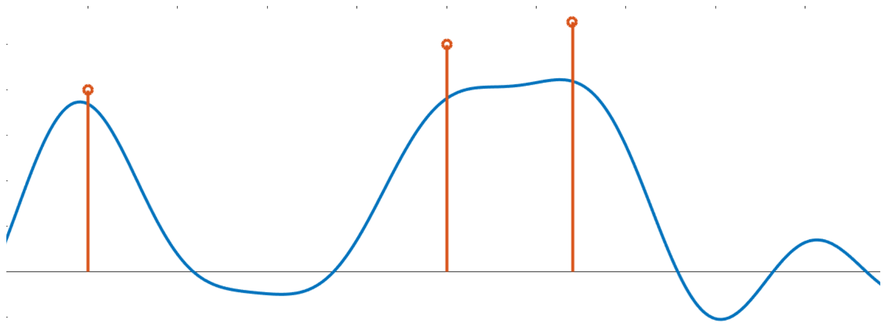
$$F(x_k) - \min_x F(x) \leq \frac{L}{k}$$

Using F-B splitting

Approach: Discrete Φ on a **fine grid**

For $\phi(x) \in \mathbb{R}^m$, define: $A = [\phi(x_1), \phi(x_2), \cdots, \phi(x_N)] \in \mathbb{R}^{m \times N}$

$$\min_{x \in \mathbb{R}^n} F(x) = \frac{1}{2\lambda} \|Ax - y\|^2 + \|x\|_1$$

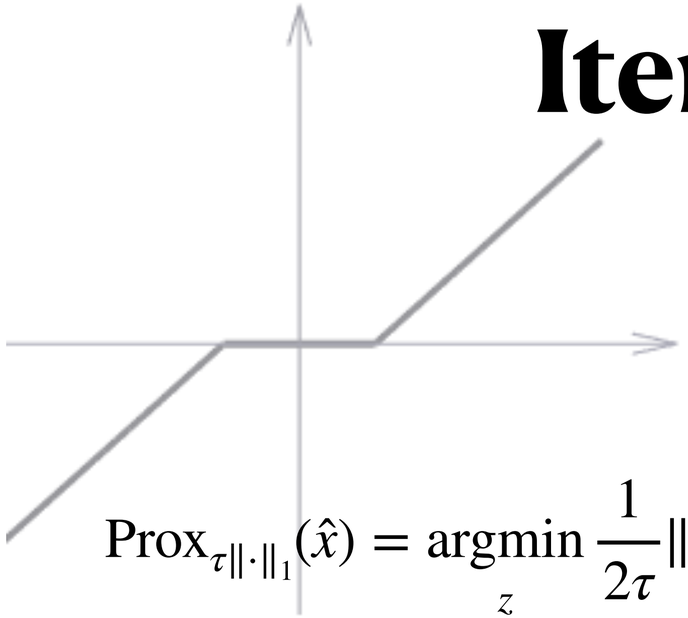


Example: Fourier measurements

Column i : $A_i = \left(\exp(2\pi\sqrt{-1} x_i^\top \omega_k) \right)_k$

$\mathcal{O}(p^{-d})$ grid points if $[x_i] \subseteq [0,1]^d$ spaced p apart

Iterative Soft Thresholding



$$\mathrm{Prox}_{\tau\|\cdot\|_1}(\hat{x}) = \operatorname*{argmin}_z \frac{1}{2\tau}\|z-\hat{x}\|^2 + \|z\|_1$$
$$= \mathrm{sign}(\hat{x})(|\hat{x}| - \tau)_+$$

$$\begin{cases} \hat{x}_{k+1} &= x_k - \tau A^\top (Ax_k - y) \\ x_{k+1} &= \mathrm{Prox}_{\tau\|\cdot\|_1}(\hat{x}_{k+1}) \end{cases}$$

$$f(x) = \frac{1}{2}\|Ax-y\|^2 \text{ is } L\text{-Lipschitz with } L = \|A\|^2$$

Convergence rates: for $\tau = 1/L$

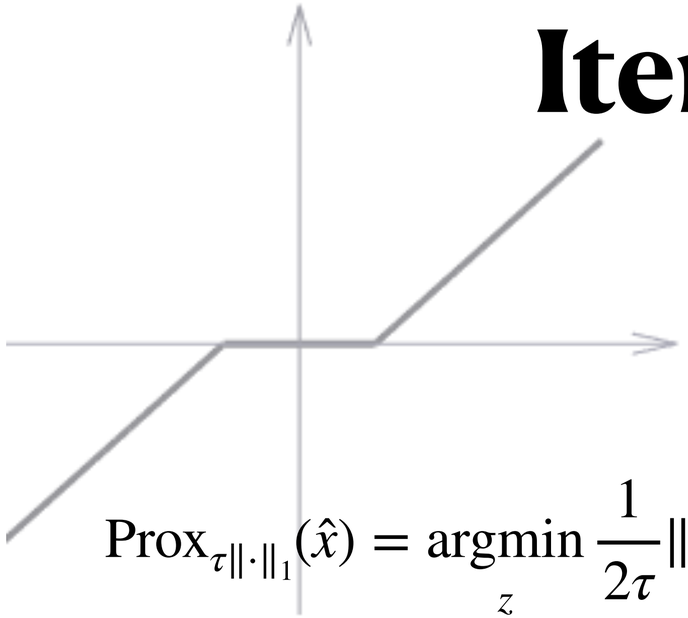
$$F(x_k) - \min_x F(x) \leq \frac{\|A\|^2}{k}$$

Example: Fourier measurements

Column i : $A_i = \left(\exp(2\pi\sqrt{-1}x_i^\top \omega_k)\right)_k$

$\|A\|^2 = \mathcal{O}(p^{-d})$

Iterative Soft Thresholding



$$\min_{x \in \mathbb{R}^n} \Phi(x) = \frac{1}{2\lambda}\|Ax - y\|^2 + \|x\|_1$$

$$\begin{aligned}\mathrm{Prox}_{\tau\|\cdot\|_1}(\hat{x}) &= \operatorname*{argmin}_z \frac{1}{2\tau}\|z - \hat{x}\|^2 + \|z\|_1 \\ &= \mathrm{sign}(\hat{x})(|\hat{x}| - \tau)_+\end{aligned}$$

$$\begin{cases}\hat{x}_{k+1} &= x_k - \tau A^\top(Ax_k - y) \\ x_{k+1} &= \mathrm{Prox}_{\tau\|\cdot\|_1}(\hat{x}_{k+1})\end{cases}$$

Convergence rates (depends on n):

$$F(x_k) - \min_x F(x) \leq \frac{C_n}{k}$$

NB: C_n can grow with n!

Grid-free convergence rates (Chizat 2021):

$$F(x_k) - \min_x F(x) \leq k^{-2/(d+1)}$$

NB: Result is independent of n

Issue with fine grids: exact resolution is not possible [Duval and Peyré 2017]

Finite dimensional formulations

$$\sup_{p\in\mathbb{R}^m} \langle p, y\rangle - \frac{\lambda}{2}\|p\|^2 \quad \text{s.t.} \quad \|\Phi^* p\|_\infty \le 1$$

In general, the constraint is infinite dimensional. However, there are special cases for which one can formulate as a finite dimensional problem.

Fourier setting
[Candés & Fernandez-Granda]:

Minimise over all p such that for all x, $|\sum_{|k|\le f_c} p_k \exp(2\pi\sqrt{-1}kx)| \le 1$

This constraint can be written as a positive semidefinite constraint on matrices.

Quadratic: $\phi(x) = ((u_i^\top x)^2)_i$ for $x \in \mathbb{S}_{n-1}$, then

$$\Phi^* p(x) = \sum_k p_k (u_k^\top x)^2 = \langle (\sum_k p_k u_k u_k^\top)x, x\rangle$$

Constraint: Spectral norm is bounded by 1.

ReLU [Pilanci & Ergen 2020]: $\phi(x) = ((u_i^\top x)_+)_i$ for $x \in \mathbb{S}_{n-1}$. Then. $\Phi^* p(x) = \sum_k p_k (u_k^\top x)_+$

Uses the fact that there are only finitely many support patterns for which (Ux)_+

Semi-definite programming

Special case of Fourier samples:

minimise over all p such that for all x, $\left| \sum_{|k| \leq f_c} p_k \exp(2\pi\sqrt{-1}kx) \right| \leq 1$

Theorem (Dumitrescu): A trigonometric polynomial $f(t) = \sum_{k=0}^{n-1} c_k \exp(\sqrt{-1}2\pi kx)$ with $p \in \mathbb{C}^n$ Iid uniformly bounded by 1 in magnitude if there exists $Q \in \mathbb{C}^{n \times n}$ Hermitian s.t.

$$0 \leq \begin{pmatrix} Q & p \\ p^* & 1 \end{pmatrix} \quad \text{and} \quad \sum_{i=1}^{n-j} Q_{i,i+j} = \delta_{0,j}$$

Semi-definite programming

Equivalent dual formulation: Let $n = 2f_c + 1$:

$$\sup_{p,Q} \langle p, y \rangle - \frac{\lambda}{2}\|p\|^2 \quad \text{s.t.} \quad 0 \leq \begin{pmatrix} Q & p \\ p^* & 1 \end{pmatrix} \quad \text{and} \quad \sum_{i=1}^{n-j} Q_{i,i+j} = \delta_{0,j}$$

Finite dimensional semi-definite program!

1. Solve SDP to find p_λ
2. Find the support of m_λ by finding the roots of the polynomial
$$f_{2n-2}(e^{\sqrt{-1}2\pi x}) = 1 - |\Phi^* p_\lambda(x)|^2$$
3. This has at most $n - 1$ roots (unless identically 0)
4. Solve for amplitudes.

Frank-Wolfe algorithm

$$\min_{x \in C} f(x)$$

C is a weakly compact convex set of a Banach space.

f is a differentiable convex function.

1. $z^k \in \operatorname{argmin}_{z \in C} f(x^k) + \langle \nabla f(x^k), z - x^k \rangle$
2. If $\langle \nabla f(x^k), z^k - x^k \rangle = 0$ then x^k is a solution.
3. $\gamma^k = 2/(k+2)$
4. $x^{k+1} = x^k + \gamma^k (z^k - x^k)$

- f is convex $\Longrightarrow$ $f(z) \geq f(x^k) + \langle \nabla f(x^k), z - x^k \rangle$ $\overset{Step2}{\Longrightarrow}$ $f(z) \geq f(x^k)$ for all z.
- One can replace x^{k+1} in step 4 with any $\hat{x}^{k+1}$ such that $f(\hat{x}^{k+1}) \leq f(x^{k+1})$.

Applying Frank-Wolfe to the Blasso

$$\mu \in \underset{\mu \in \mathcal{M}(\mathcal{X})}{\operatorname{argmin}} f_\lambda(m) := \frac{1}{2}\|\Phi\mu - y\|^2 + \lambda\|\mu\|_{TV}$$

$$(\|\mu\|_{TV}, \mu) \in \underset{(t,\mu)\in C}{\operatorname{argmin}} \hat{f}_\lambda(t,\mu) := \frac{1}{2}\|\Phi\mu - y\|^2 + \lambda t$$

$$C = \{(t,\mu) \in \mathbb{R}_+ \times \mathcal{M}(\mathcal{X}) \ : \ \|\mu\|_{TV} \leq t \leq \|y\|^2/(2\lambda) \ \}$$

$$\hat{f}_\lambda \in C^1 \ : \quad \begin{cases} \partial_t f = \lambda \\ \partial_\mu \hat{f}_\lambda(t,\mu) = \Phi^*(\Phi\mu - y) \end{cases}$$

$m \in \operatorname{argmin} f_\lambda(m)$ implies

$$\lambda\|\mu\|_{TV} \leq \lambda\|\mu\|_{TV} + \frac{1}{2}\|\Phi\mu - y\|^2$$

$$\leq \lambda\|0\|_{TV} + \frac{1}{2}\|\Phi 0 - y\|^2 = \frac{1}{2}\|y\|^2$$

[Bredies and Pikkarainnen 2013, Boyd et al, 2017, Denoyelle et al 2018]

Convergence of Frank-Wolfe

[Jaggi (2011)]: The curvature constant of f over C is

$$R := \max_{\substack{\gamma \in [0,1] \\ x,s,y \in C \\ y=(1-\gamma)x+\gamma s}} \frac{2}{\gamma^2}\left(f(y) - f(x) - \langle \nabla f(x), y - x\right)$$

Convergence rate:

$$f(x^k) - f^* \leq \frac{2R}{k+2}$$

- $R = 0$ if f is linear
- . ∇f is L-Lipschitz $\implies$ $f(y) - f(x) - \langle \nabla f(x), y - x\rangle \leq \frac{L}{2}\|y - x\|^2$

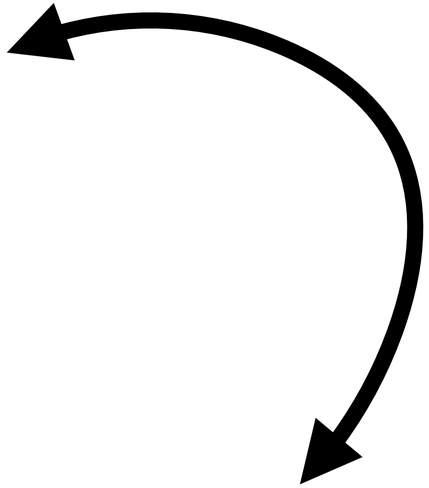
  $\implies$ $R \leq L\text{diam}(C)^2$
- . For $\hat{f}_\lambda$, $R = \frac{1}{2}\sup\{\|\Phi(m - m')\|^2 : \|m\|_{TV}, \|m'\|_{TV} \leq \|y\|^2/2\}$
- If $\|\phi(x)\| = 1$ for all x, then $R \lesssim \|y\|^2$ and the convergence rate is $\mathcal{O}(\|y\|^2/t)$

Applying Frank-Wolfe to the Blasso

Key step. Let $u^k = (t^k, m^k)$.

$$z^k \in \operatorname{argmin}_{z \in C} \hat{f}(u^k) + \langle \nabla \hat{f}(u^k), z - u^k \rangle$$

$$= \operatorname{argmin}_{(t,m) \in C} \langle \Phi^*(\Phi m^k - y), m \rangle + \langle \lambda, t \rangle$$



C is a convex set and minimum is achieved at an extremal point of C.

$$E = \{(M, \pm M\delta_x) \ : \ x \in \mathcal{X}\}$$

$$\Longrightarrow \quad z^k = (M, \pm M\delta_{x_k})$$

$$z^k = \begin{cases} (M, -M\delta_{x_k}) & \text{if} \eta^k(x_k) > 0 \\ (M, M\delta_{x_k}) & \text{if} \eta^k(x_k) < 0 \end{cases}$$

$$x^k \in \operatorname*{argmin}_{x \in \mathcal{X}} \pm [\Phi^*(\Phi m^k - y)](x) + \lambda$$

$$= \operatorname*{argmax}_{x \in \mathcal{X}} \ |\eta^k(x)|, \qquad \eta^k = \frac{1}{\lambda}[\Phi^*(\Phi m^k - y)]$$

Applying Frank-Wolfe to the Blasso

At each iteration k, construct $\mu^k = \sum_{j=1}^{k} a^j \delta_{x^j}$

Define $\eta^k = \frac{1}{\lambda}[\Phi^*(\Phi\mu^k - y)]$ and $\gamma^k = 2/(k+2)$.

1. Add new spike: $x^k = \operatorname{argmax}_{x \in \mathcal{X}} |\eta^k(x)|$
2. $\mu^{k+1} = \mu^k + \gamma^k \left(-\operatorname{sign}(\eta^k(x^k))\, M\, \delta_{x^k} - \mu^k\right)$

Terminate if $\eta^k(x^k) = \pm 1$ and return μ^k

Sliding Frank-Wolfe: *[Denoyelle et al 2018]*

Replace step 2 with any $\hat{\mu}^k$ such that $f_\lambda(\hat{\mu}^{k+1}) \leq f_\lambda(\mu^{k+1})$:

$$\hat{\mu}^{k+1} = \sum_{j=1}^{k+1} \hat{a}_j\, \delta_{x_j} \text{ where}$$

$$(\hat{a}, \hat{x}) \in \min_{a,x} \lambda \|a\|_1 + \frac{1}{2}\|\Phi\mu_{a,x} - y\|^2$$

Theorem (Denoyelle et al): finite convergence if η_V is non-degenerate at the solution.

Remarks

- The sliding Frank-Wolfe is an off-the-grid algorithm, however:

  ➡ the **difficulty** is in the step $x^k = \operatorname{argmax}_{x \in \mathcal{X}} |\eta^k(x)|$

- $\mathcal{X}$ is a continuous space and η^k is a continuous (smooth) function
- In practice, discretize $\mathcal{X}$ and do a local ascent step on η^k
- This is computationally intensive if $\mathcal{X} \subset \mathbb{R}^d$ and d is large.

Particle methods

$$\min_{\mu} \lambda \|\mu\|_{TV} + \frac{1}{2}\|\Phi\mu - y\|^2$$

This has a solution consisting of $m + 1$ Diracs

$$\min_{a,x} \lambda \sum_{i=1}^{k} |a_i| + \frac{1}{2}\| \sum_{i=1}^{k} \phi(x_i)a_i - y\|^2$$

Has the same value as $P_\lambda(y)$ when $k \geq m$

Chizat and Bach (2018): Global convergence results for *sufficiently large k*.

VarPro

[Golub and Pereyra, 1978]

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

$$\Longleftrightarrow$$

$$\min_x f(x) \text{ where}$$

$$f(x) = \min_a \| \sum_j \phi(x_j) a_j - y \|^2$$

Easy to compute gradient:

$$\partial_{x_j} f(x) = \partial_{x_j} \|\Phi_x \bar{a} - y\|^2 = \bar{a}_j \nabla \phi(x_j)^\top (\Phi_x \bar{a} - y)$$

$$\bar{a} = \text{argmin}_a \|\Phi_x a - y\|^2 = \Phi_x^\dagger y$$

Leads to better problem conditioning.

In general, solution is non-sparse, except in special cases, e.g. ReLU

VarPro

A practical approach for sparse solutions (Nonsmooth VarPro):

$$\min_x f(x) \text{ where } f(x) := \min_a \frac{1}{2}\|\Phi_x a - y\|^2 + \lambda\|a\|_1$$

If the inner Lasso problem has a unique solution for x, then f is differentiable at x with gradient

$$\partial_{x_j} f(x) = \partial_{x_j}\|\Phi_x \bar{a} - y\|^2 = \bar{a}_j \nabla\phi(x_j)^\top(\Phi_x \bar{a} - y)$$

$$\bar{a} = \text{argmin}_a \frac{1}{2}\|\Phi_x a - y\|^2 + \lambda\|a\|_1$$

Summary

- For certain settings (Fourier sampling), one can convert to a finite dimensional optimisation problem.
- The Frank-Wolfe algorithm is a versatile algorithm for computations off-the-grid. Works well in low dimensions, but there is a difficulty with finding the argmax of η_k
- Particle methods are effective in practice, but no quantitative rates.

References

SDP/convex finite dimensional formulations

- Candès, E. J., & Fernandez-Granda, C. (2014). Towards a mathematical theory of super-resolution. *Communications on pure and applied Mathematics*, *67*(6), 906-956.
- Catala, Paul, Vincent Duval, and Gabriel Peyré. "A low-rank approach to off-the-grid sparse superresolution." *SIAM Journal on Imaging Sciences* 12.3 (2019): 1464-1500.
- Pilanci, Mert, and Tolga Ergen. "Neural networks are convex regularizers: Exact polynomial-time convex optimization formulations for two-layer networks." *International Conference on Machine Learning*. PMLR, 2020.

Frank-Wolfe

- Bredies, K., & Pikkarainen, H. K. (2013). Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations*, *19*(1), 190-218.
- Boyd, N., Schiebinger, G., & Recht, B. (2017). The alternating descent conditional gradient method for sparse inverse problems. *SIAM Journal on Optimization*, *27*(2), 616-639.
- 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.

“Particle” approaches

- Chizat, L., & Bach, F. (2018). On the global convergence of gradient descent for over-parameterized models using optimal transport. *Advances in neural information processing systems*, *31*.
- Golub, Gene, and Victor Pereyra. "Separable nonlinear least squares: the variable projection method and its applications." *Inverse problems* 19.2 (2003): R1.

Sparsistency

Support stability

Q: Given $y = \Phi\mu_{a,x} + w$***, does the solution to*** $P_\lambda(y)$ ***consist of precisely*** s ***spikes?***

Recall: if $p_\lambda = \operatorname{argmax} D_\lambda(y)$ and $\eta_\lambda = \Phi^* p_\lambda$, then $\operatorname{Supp}(\mu_\lambda) \subset \{x \ : \ |\eta_\lambda(x)| = 1\}$

What is the behaviour of η_λ *when* λ *and* $\|w\|$ *are small?*

Limit of η_λ:

If $D_0(y)$ has a solution, then as $\lambda \to 0$, $\|w\| \to 0$,

$$\|p_\lambda - p_0\| \to 0\,, \qquad p_0 = \operatorname{argmin}\left\{ \|p\| \ : \ p \in \operatorname{argmax} D_0(\Phi\mu_{a,x}) \right\}$$

The limit dual problem

- Recall $p_\lambda = \operatorname{argmax}_{\|\Phi^* p\|_\infty \leq 1} \langle p, y \rangle - \lambda \|p\|^2/2$
- Let p_0 be of minimal norm such that $p_0 \in \operatorname{argmax}_{\|\Phi^* p\|_\infty \leq 1} \langle p, y \rangle$

$$\langle p_\lambda, y \rangle - \lambda \|p_\lambda\|^2/2 \geq \langle p_0, y \rangle - \lambda \|p_0\|^2/2 \geq \langle p_\lambda, y \rangle - \lambda \|p_0\|^2/2$$

- $\|p_\lambda\| \leq \|p_0\|$ for all λ.
- $(p_\lambda)_\lambda$ converges (up to subseq) to $\bar{p}$ with $\|p_0\| \geq \|\bar{p}\|$ and $\|\Phi^* \bar{p}\|_\infty \leq 1$

Take limit $\lambda \to 0$

$\langle \bar{p}, y \rangle \geq \langle p_0, y \rangle$, so $\bar{p} = p_0$

Minimal norm certificate

We say that η is non degenerate if:

- $\eta''(x_i) \neq 0$
- $\eta(x_i) = \operatorname{sign}(a_i)$
- $\forall x \notin \{x_i\}, |\eta(x)| < 1$

Minimal norm certificate

$$\eta_\lambda \overset{L^\infty}{\to} \eta_0 = \Phi^* p_0$$

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

If η_0 is non-degenerate, then η_λ is also non degenerate when λ is sufficiently small.

Theorem (Duval and Peyre, 2015):

If η_0 is non-degenerate, then for $\|w\|/\lambda = \mathcal{O}(1)$ and $\lambda = \mathcal{O}(1)$, the solution to

$P_\lambda(y)$ is unique, $\mu_\lambda = \sum_{i=1}^{n} a_{\lambda,i} \delta_{x_{\lambda,i}}$ and $\|(x_\lambda, a_\lambda) - (x_0, a_0)\| = \mathcal{O}(\|w\|)$

Computing the minimal norm certificate

Minimal norm certificate

$$\eta_0 = \Phi^* p_0 = \underset{\eta = \Phi^* p}{\operatorname{argmin}} \|p\| \quad \text{s.t.} \quad \begin{cases} \forall i,\ \eta(x_i) = \operatorname{sign}(a_i) \\ \|\eta\|_\infty \leq 1 \end{cases}$$

Necessary:

$$\begin{cases} \operatorname{sign}(a_i) = \langle p, \phi(x_i) \rangle \\ 0 = \langle p, \phi'(x_i) \rangle \end{cases}$$

Vanishing derivatives Pre-certificate

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

$$\Gamma = [(\phi(x_i))_i, (\nabla \phi(x_i))_i]$$

$$\Gamma^* p = \begin{pmatrix} \operatorname{sign}(a) \\ 0_{kd} \end{pmatrix}$$

Linear system of $dk + k$ equations

Computing the minimal norm certificate

p_V is the solution to a linear system of $2n$ equations.

$$\begin{pmatrix} [K(x_i, x_j)]_{i,j} & [K^{(1,0)}(x_i, x_j)]_{i,j} \\ [K^{(0,1)}(x_i, x_j)]_{i,j} & [K^{(1,1)}(x_i, x_j)]_{i,j} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \operatorname{sign}(a) \\ 0_n \end{pmatrix}$$

$$\eta_V(x) = \sum_{i=1}^{n} u_i K(x_i, x) + \sum_{i=1}^{n} v_i K^{(10)}(x_i, x) \rangle \qquad K(x, x') = \langle \phi(x), \phi(x') \rangle$$

Useful checks for analysing support stability:

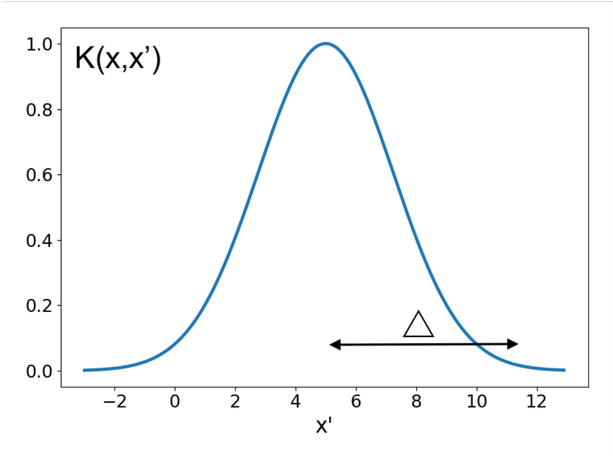
[Necessary cond] η_V *must satisfy* $\|\eta_V\|_\infty \leq 1$ *for support stability.*

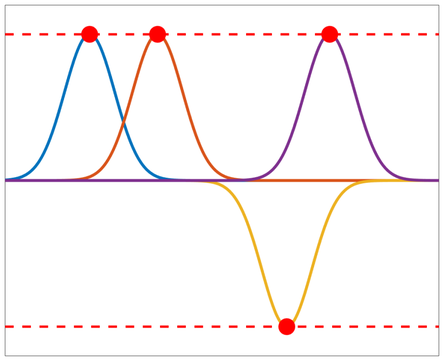
[Sufficient cond] If η_V *is non-degenerate, then support stability is guaranteed*

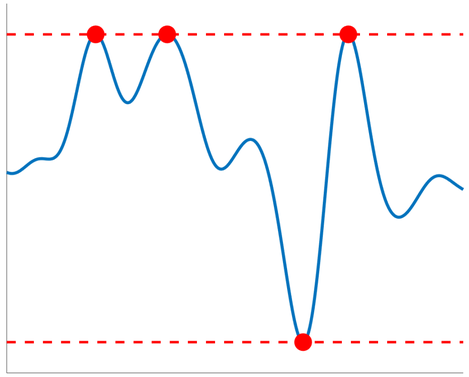
Recovery under minimal separation

Typical analysis strategy to understand sparse identifiability properties of Φ:

Compute η_V and check if it is non-degenerate.







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_V is non-degenerate. So, we have stable recovery.

Super-resolution

No super-resolution for opposite sign spikes:

If $|x - x'| < 1/f_c$, then $\mu := \delta_x - \delta_{x'}$ cannot be recovered from $P_0(\Phi\mu)$

De Castro & Fabrice (2012):

To recover n spikes with positive amplitudes, we need $f_c \geq N$ when there is no noise.

Q: *Given N spikes at distance t apart, how small does the noise level $\|w\|$ need to be to identify N spikes?*

Hint: *Look at the certificate η_{tx} corresponding to positions $tx = (tx_i)_{i=1,\ldots,N}$,*

When is it non-degenerate?

Asymptotic vanishing derivatives precertificate in 1D

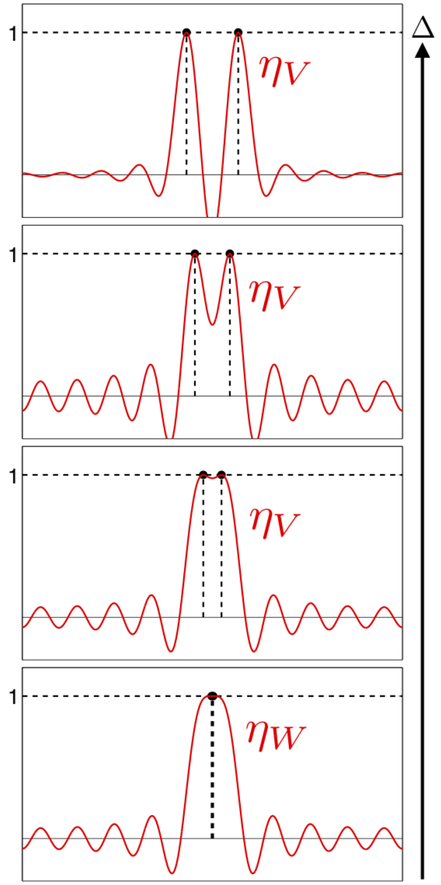
Theorem (Denoyelle et al, 2015):

As $t \to 0$, $\eta_{V,tx} \to \eta_w$ where

$$\eta_W = \operatorname{argmin}_{\eta = \Phi^* p} \|p\| \quad \text{s.t.} \quad \begin{cases} \eta(0) = 1 \\ \eta^{(1)}(0) = \cdots = \eta^{(2N-1)}(0) = 0 \end{cases}$$

This is called non-degenerate if

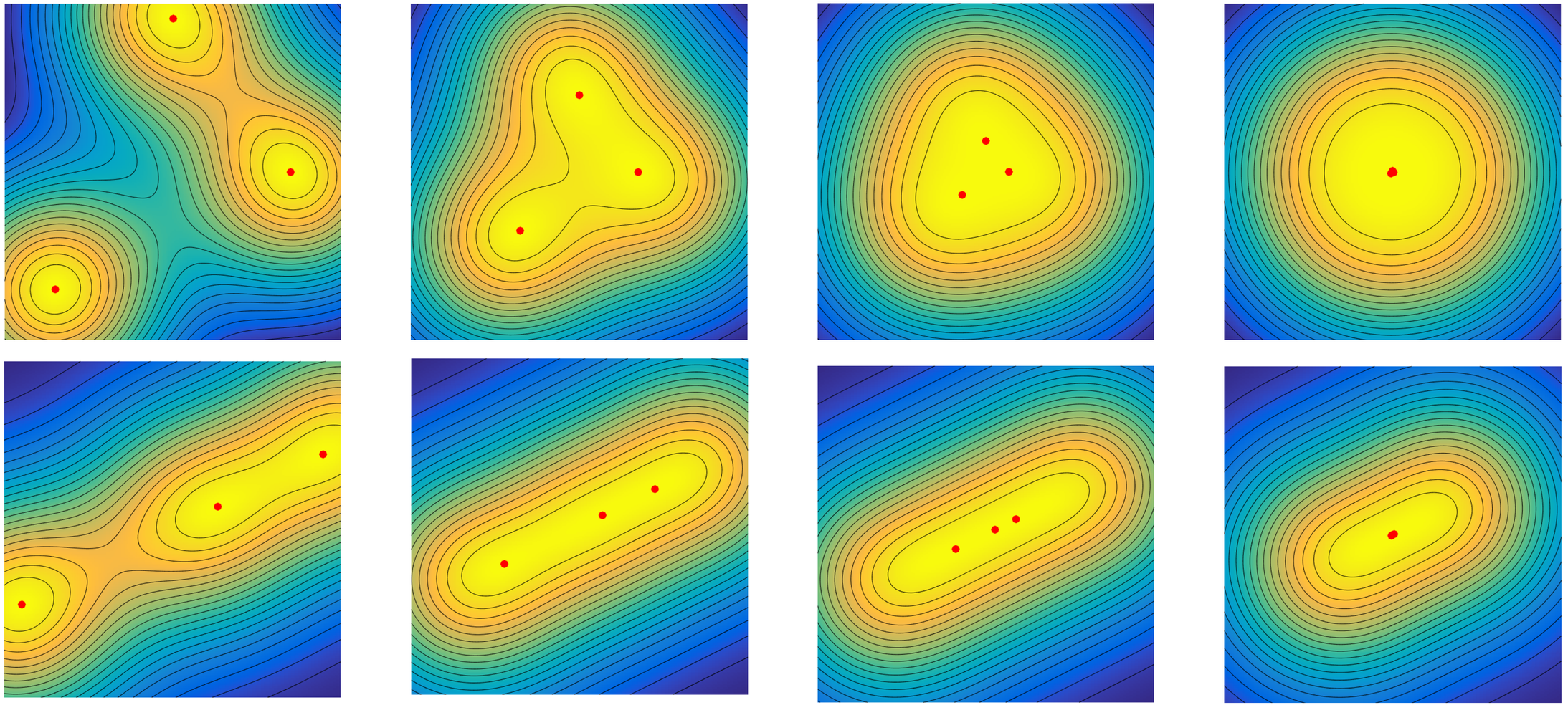
$$\eta_W^{(2N)}(0) < 0 \quad \text{and} \quad \forall z \neq 0, \ |\eta_W(z)| < 1$$



$\eta_{V,tx}$ is non-degenerate for all t sufficiently small.

For $\|w\|/\lambda = \mathcal{O}(1)$, $\lambda = \mathcal{O}(t^{2N-1})$, $P_\lambda(\Phi\mu_{a,tx} + w)$ recovers exactly N spikes.

Asymptotic vanishing derivatives precertificate in higher dimensions



$t = 1$ $t = 0.5$ $t = 0.2$ $t = 0.1$

The limit of η_V depends on the spikes configuration!

The multivariate limiting certificate

Theorem (Poon and Peyré, 2019):

Let $p_{V,tZ}$ be the precertificate associated to support $tz := (tz_i)_{i=1}^N$, then $\|p_{V,tz} - p_{w,z}\| = \mathcal{O}(t)$

where $p_{w,z} = \operatorname{argmin} \left\{ \|p\| \ : \ (\Phi^* p)(0) = 1, \quad P(\partial)(\Phi^* p)(0) = 0, \ P \in \mathcal{S}_z \right\}$

The polynomial space $\mathcal{S}_z$ is the *least interpolant polynomial space associated to* z.

Hermite interpolation problem : Given c_i, d_i ,

find $P \in \mathcal{S}$ such that $\begin{cases} P(z_i) = c_i \\ \nabla P(z_i) = d_i \end{cases}$

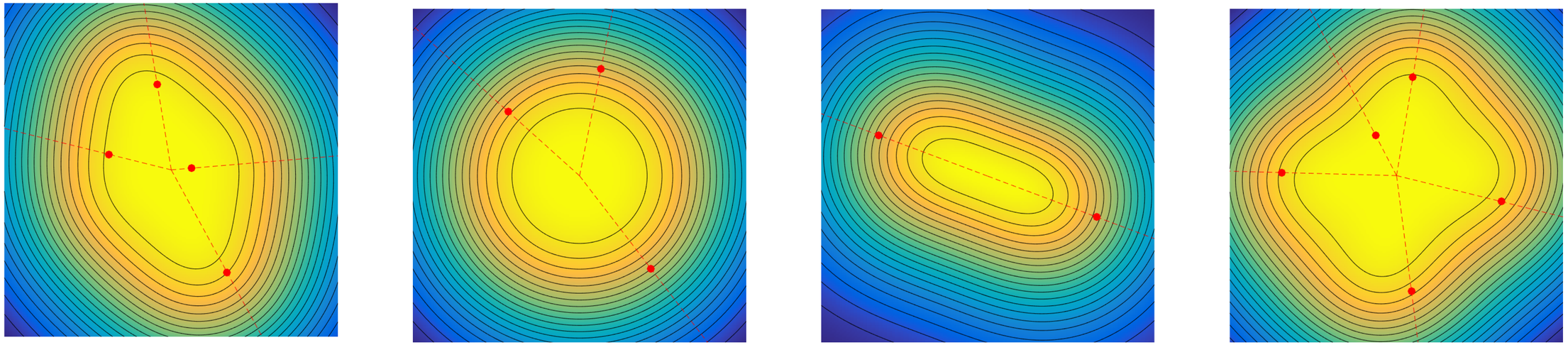
[De Boor and Ron (1990)]:

The least interpolant space is the polynomial space of least degree for which there is a unique solution.

Useful check: For support stability, it is necessary that $\|\eta_{W,z}\|_\infty \leq 1$

Gaussian convolution

$$\phi(x) = \exp(-\|x - \cdot\|^2/(2\sigma^2)) \in L^2(\mathbb{R}^2)$$

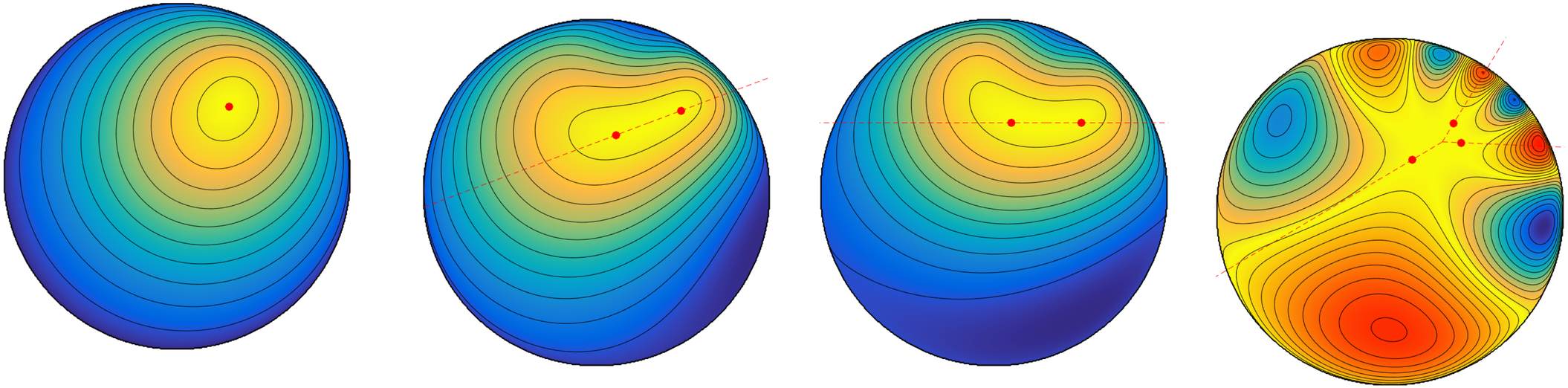


Numerical observation: $\eta_{W,z}$ is always uniformly bounded by 1.

So, we can expect super-resolution when SNR is large enough.

Neuro-imaging

Let $\mathcal{X} = \{x \in \mathbb{R}^2;\ \|x\| \leq 1\}$. To model MEG/EEG, $\phi(x) = u \mapsto \|x-u\|^{-2} \in L^2(\partial\mathcal{X})$



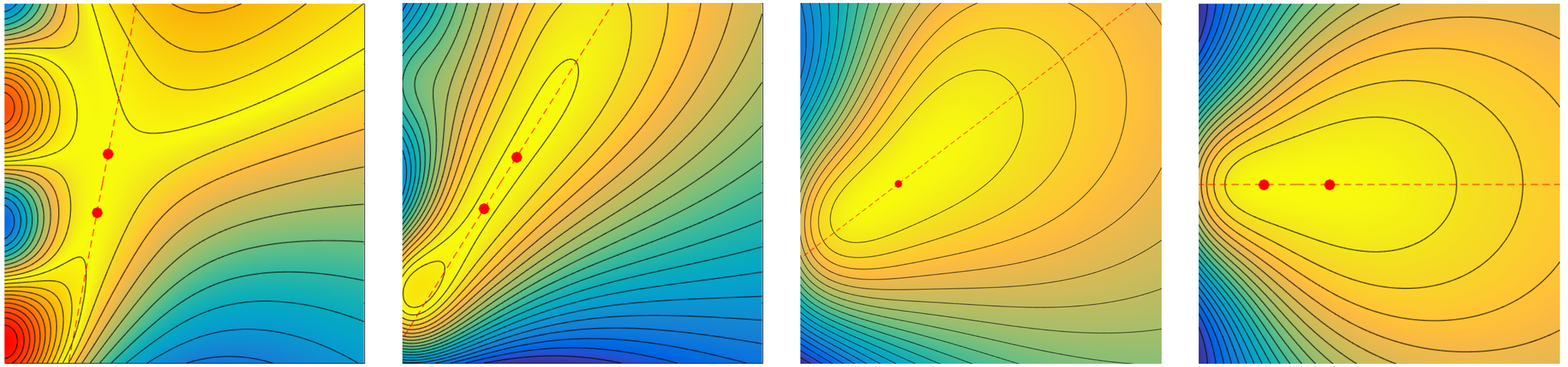
Numerical observation:

- $\eta_{W,z}$ always valid when z consists of aligned spikes
- It is not valid when the spikes are not aligned.

In general, cannot super-resolve 3 close spikes under noise.

Gaussian mixture

For $x = (m, s) \in \mathcal{X} = \mathbb{R} \times \mathbb{R}_+$, $\quad \phi(x) = \frac{1}{s} \exp\left(-\frac{(\cdot - m)^2}{2s^2} \right) \in L^2(\mathbb{R})$



Y-axis = mean, X-axis = standard deviation

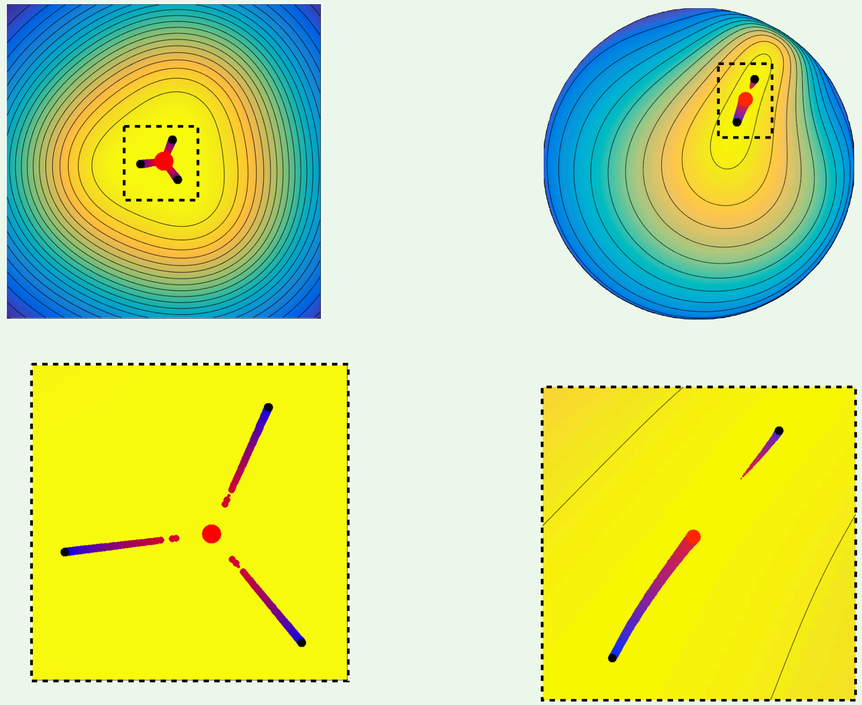
Observation: $\eta_{W,z}$ is a valid certificate if $|m_1 - m_2| \leq |s_1 - s_2|$

One cannot expect to super-resolve a mixture of 2 Gaussians when the variation in means is too large wrt variation in standard deviations.

Evolution of solutions

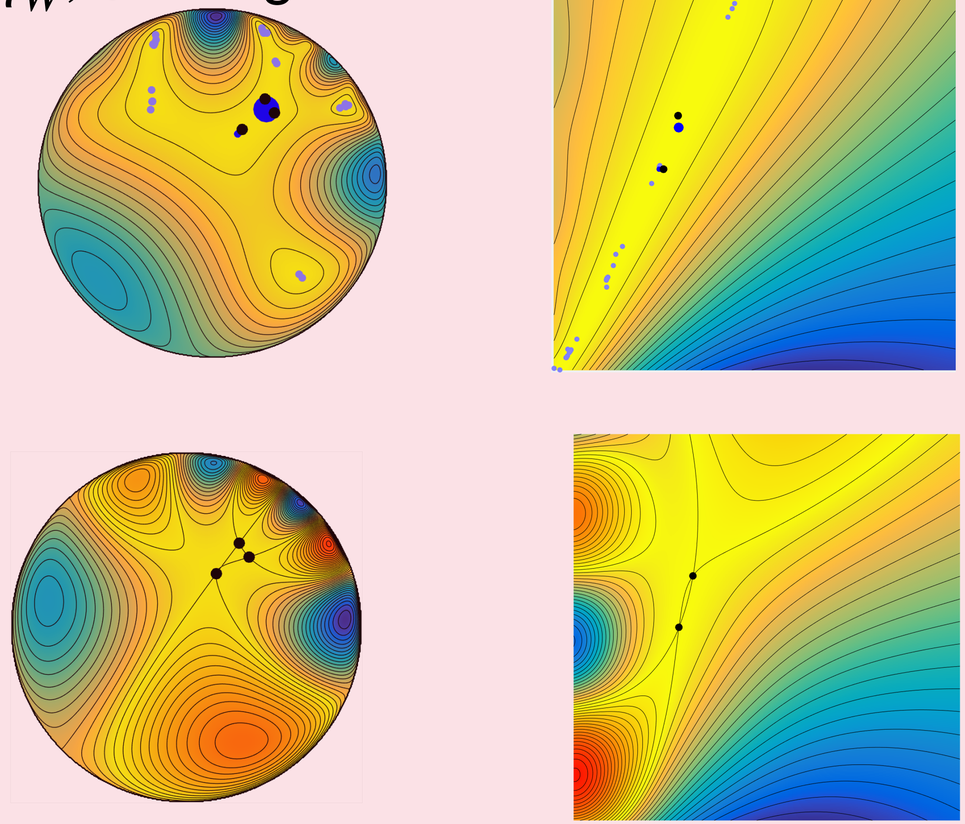
Measurements: $y = \Phi\mu_0 + \lambda w$ where $w = \Phi\hat{\mu}$ with $\hat{\mu} = \sum_{j=1}^{20} b_j \delta_{u_j}$

When η_W, z is non-degenerate



Displaying evolution of solutions from $\lambda_{\max}$ (blue) to 0 (red)

When η_W, z is degenerate



Solution unstable when $\eta_{W,z}$ is degenerate.
Many tiny spikes (light blue) are added!

Compressed sensing for the Blasso

Off-the-grid Compressed sensing

Problem:

- Let $\phi_\omega(x) \in \mathscr{C}(\mathscr{X})$ where $\omega \in \Omega$.
- Suppose we observe $\Phi\mu = \left(\langle \phi_{\omega_k}, \mu \rangle\right)_{k=1}^{m}$ where $\omega_1, \ldots, \omega_m$ are drawn iid from Ω

Example:

- Random Fourier sampling :
  $\phi_\omega(x) = \exp(\sqrt{-1}2\pi\omega x)$ and $\omega \in \{-N, \ldots, N\}$

Question:

If $\mu = \sum_{j=1}^{s} a_j \delta_{x_j}$, how many random samples n do we need to reconstruct m?

Recovery results (random Fourier)

Theorem ***(Tang et al 2013)****: in the case of random Fourier samples.*

If $\min_{i \neq j} |x_i - x_j| \geq C/f_c$, and $\text{sign}(a)$ is distributed uniformly iid on the complex unit circle, then exact recovery is guaranteed with probability at least $1 - \delta$ provided that

$$m = \mathcal{O}(s \log(s/\delta) \log(f_c/\delta))$$

Recovery results (general)

Theorem (Poon et al 2019):

If $\min_{i \neq j} d_g(x_i, x_j) \geq \Delta$, exact recovery is guaranteed with probability at least $1 - \rho$ provided that $$n = \mathcal{O}(s \log(s/\rho)^2 + \log(L/\rho))$$

where Δ depends on s and the kernel and L depends on the bounds on the derivatives of ϕ_ω and the diameter $\sup_{x,x' \in \mathcal{X}} d_g(x, x')$.

Stable recovery: $\lambda = \epsilon/\sqrt{s}$ where ϵ is the noise level. Then,

$$W_2^2\left(\sum_j \hat{A}_j \delta_{x_j}, |\hat{\mu}|\right) \lesssim \epsilon\sqrt{s} \quad \text{and} \quad \max_j |a_j - \hat{a}_j| \lesssim \epsilon\sqrt{s}$$

In practice the bound is:

$s \times$ log factors $\times$ poly(d)

Sketching Gaussian mixtures

- Data samples $z_1, \ldots, z_n \in \mathbb{R}^d$ drawn iid from Gaussian mixture $\xi = \sum_{i=1}^{s} a_i \mathcal{N}(x_i, \Sigma)$.
- Need to find: $a_1, \ldots, a_s > 0$ and $x_1, \ldots, x_s \in \mathbb{R}^d$
- Linear sketch: Draw $\omega_1, \ldots, \omega_n$ iid from $\mathcal{N}(0, \Sigma^{-1}/d)$ and define

$$y = \frac{C}{n} \sum_{i=1}^{n} (\exp(-\sqrt{-1}\omega_k^\top z_i))_{k=1}^{m} \approx \mathbb{E}_z[C \exp(-\sqrt{-1}\omega_k^\top z_i)] = \Phi\mu_0$$

$$\text{with } \mu_0 = \sum_{i=1}^{s} a_i \delta_{x_i} \text{ and } \phi_\omega(x) = \mathbb{E}_{z\sim\mathcal{N}(x,\Sigma)}[C \exp(\sqrt{-1}\omega^\top z)]$$

Provided that $\min_{i\neq j} \|\Sigma^{-1/2}(x_i - x_j)\| \gtrsim \sqrt{d \log(s)}$, stable recovery is guaranteed with with

$$m \gtrsim s\left(d \log(s) \log(s/\rho) + d^2 \log(sdR)^d/\rho\right), \quad \epsilon = \mathcal{O}(n^{-1/2})$$

Summary

- p_λ converges to p_0 the minimal solution to $D_0(y)$
- Support stability is determined by the minimal norm certificate.

One can compute a pre-certificate η_V in closed form and check its properties.

- $\|\eta_V\|_\infty > 1$ implies stability is impossible.
- $|\eta_V(x)| < 1$ outside the support $\{x_i\}_i$ and a pos-def/neg Hessian implies stability

Analysis of η_V has led to theoretical understanding of super-resolution and compressed sensing.

References

Support stability:

- Duval, V., & Peyré, G. (2015). Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, *15*(5), 1315-1355.

Super resolution:

- De Castro, Yohann, and Fabrice Gamboa. "Exact reconstruction using Beurling minimal extrapolation." *Journal of Mathematical Analysis and applications* 395.1 (2012): 336-354.
- Denoyelle, Q., Duval, V., & Peyré, G. (2017). Support recovery for sparse super-resolution of positive measures. *Journal of Fourier Analysis and Applications*, *23*(5), 1153-1194.
- Poon, C., & Peyré, G. (2019). Multidimensional sparse super-resolution. *SIAM Journal on Mathematical Analysis*, *51*(1), 1-44.

Compressed sensing off-the-grid

- Tang, G., Bhaskar, B. N., Shah, P., & Recht, B. (2013). Compressed sensing off the grid. *IEEE transactions on information theory*, *59*(11), 7465-7490.
- Poon, C., Keriven, N., & Peyré, G. (2021). The geometry of off-the-grid compressed sensing. *Foundations of Computational Mathematics*, 1-87.