Mini-course on Sparse estimation off-the-grid

Sparsistency

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

Clarice Poon

Yesterday...

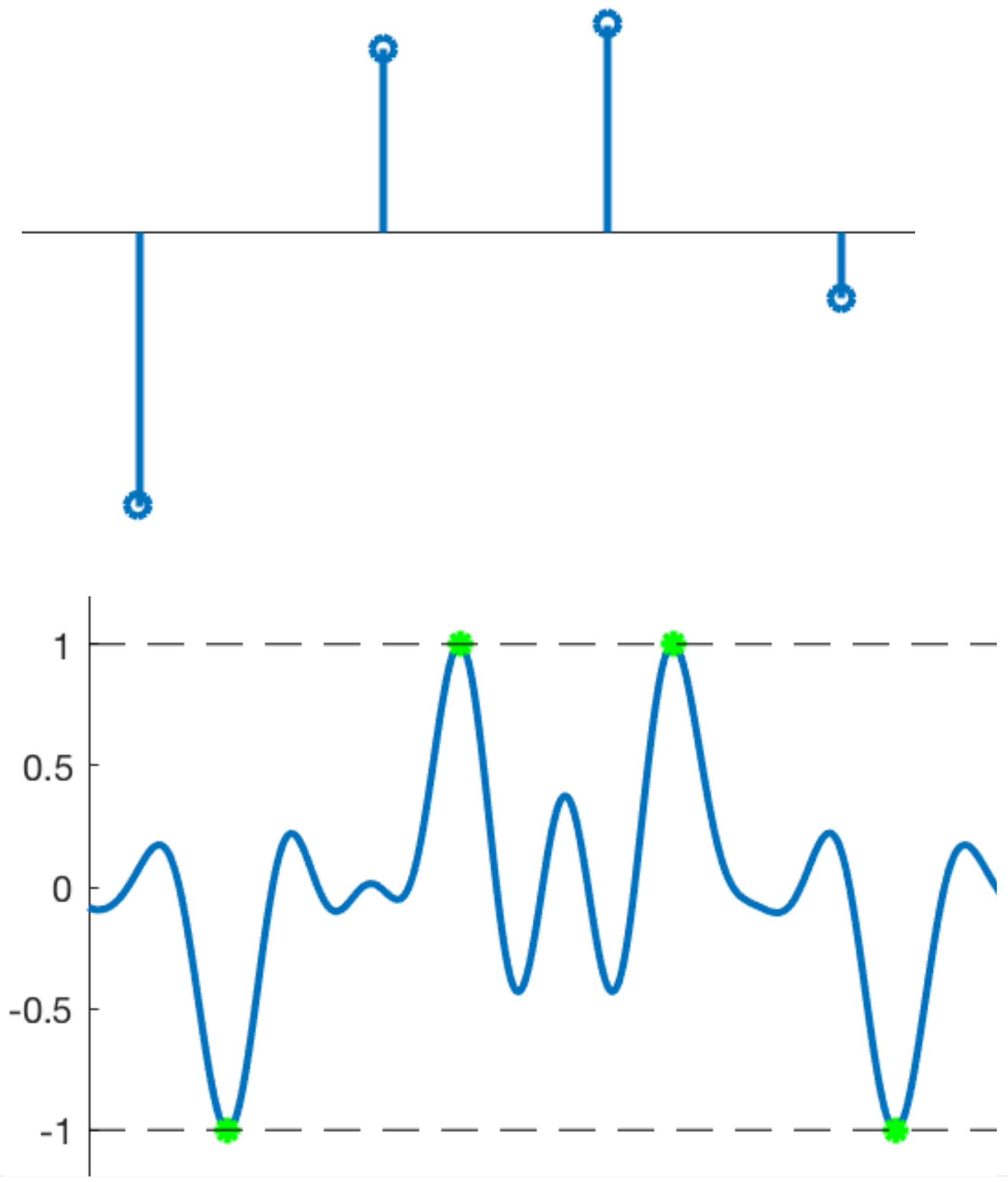
Primal: $\min_{\mu \in \mathcal{M}(\mathcal{X})} \|\mu\|_{TV} + \frac{1}{2\lambda}\|\Phi\mu - y\|^2$

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

If $\mu = \sum_{j=1}^{s} a_j \delta_{x_j}$ and $\eta = \Phi^* p$ is non degenerate:

- $\eta(x_j) = \mathrm{sign}(a_j)$
- $\eta''(x_j) \neq 0$
- $|\eta(x)| < 1$ for $x \notin \{x_1, \ldots, x_s\}$

Then stable and exact recovery is guaranteed.



There exists a non-degenerate η provided that $\min_{i \neq j} d_g(x_i, x_j) \geq \Delta$

η is a solution to $D_0(\Phi\mu)$

Support stability

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

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

Limit of η_λ: Suppose $y = \Phi\mu_{a,x} + w$.

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

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

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 $p_0 \in \operatorname{argmax}_{\|\Phi^* p\|_\infty \leq 1} \langle p, y \rangle$

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

$\implies \quad \|p_\lambda\| \leq \|p_0\|$ for all λ.

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

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:

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

Minimal norm certificate

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

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

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, $\mu_\lambda = \sum_{i=1}^{s} a_{\lambda,i}\delta_{x_{\lambda,i}}$ 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: $\begin{cases} \operatorname{sign}(a_i) = \langle p, \phi(x_i) \rangle \\ 0 = \langle p, \phi'(x_i) \rangle \end{cases}$

Pre-certificate

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

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

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

Linear system of $ds + s$ equations

Computing the minimal norm certificate

η_V can be computed by solving a linear system

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

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

Useful checks for analysing support stability:

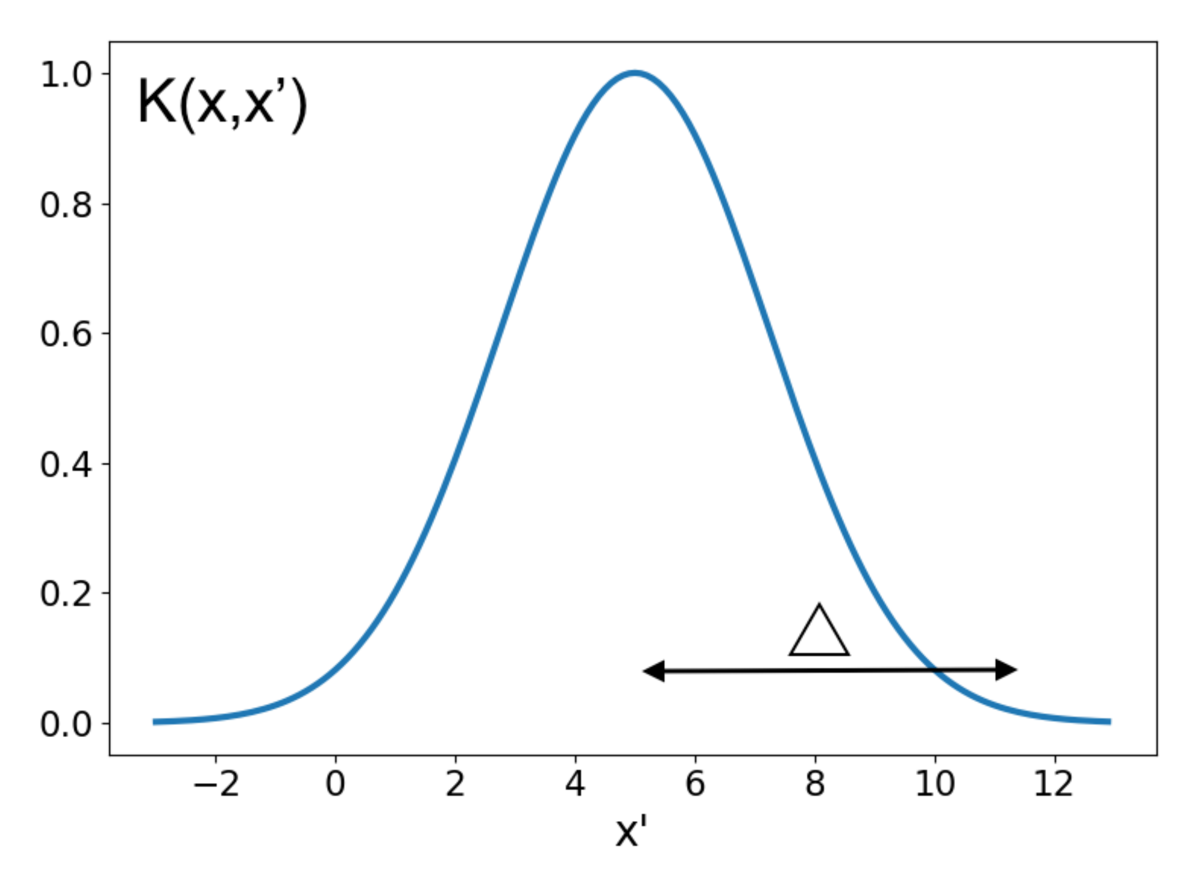
[Necessary cond] η_V *must satisfy* $\|\eta_V\|_\infty \leq 1$ *for support stability.*

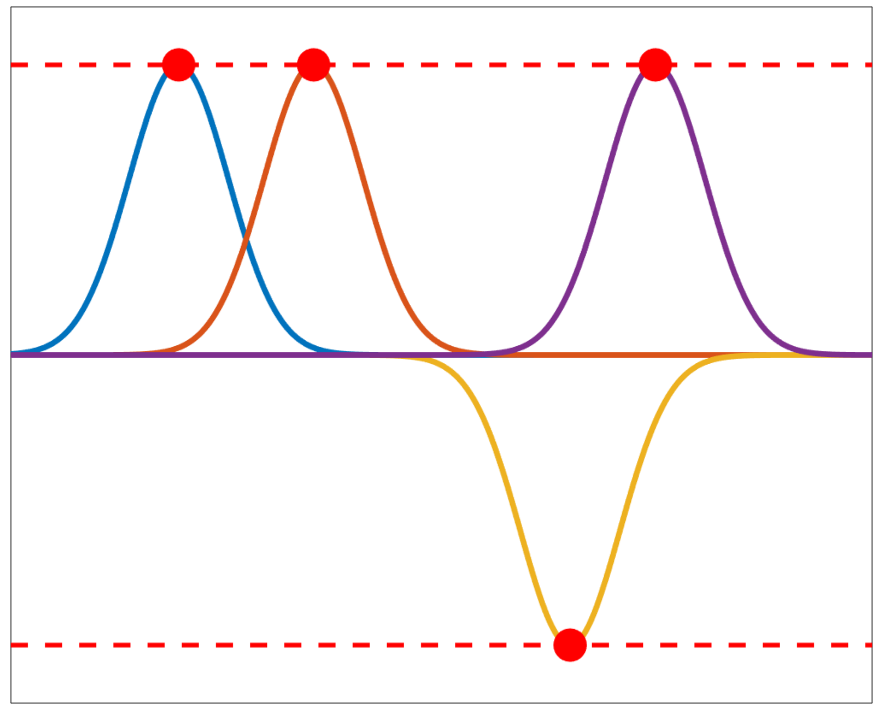
[Sufficient cond] *If* η_V *is non-degenerate, then support stability is guaranteed*

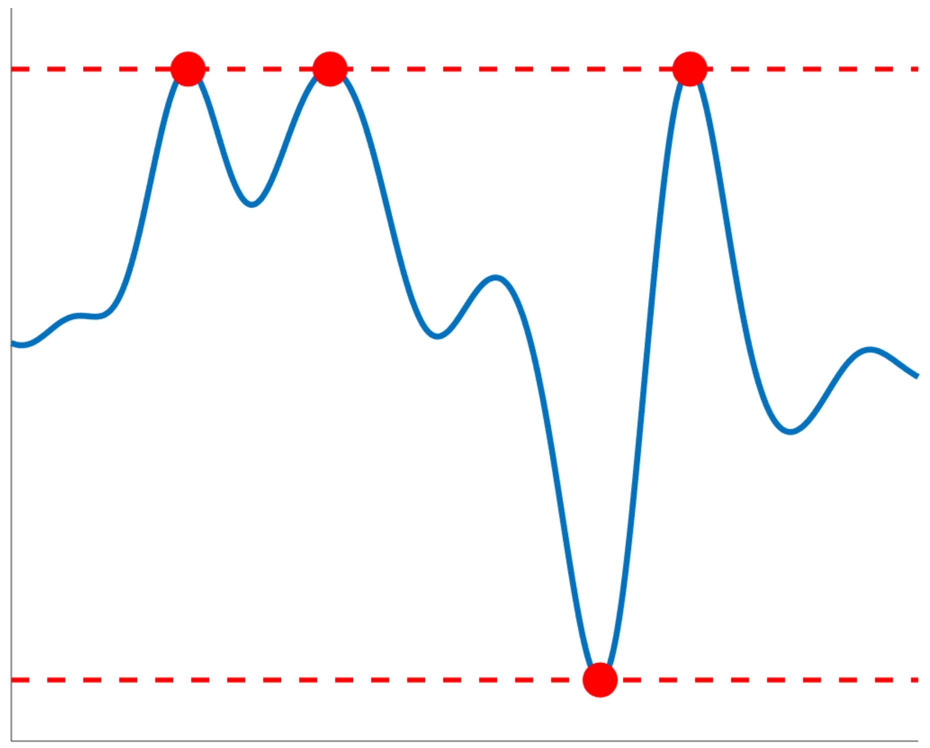
Recovery under minimal separation

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 $tx = (tx_i)_{i=1,\ldots,N}$,*

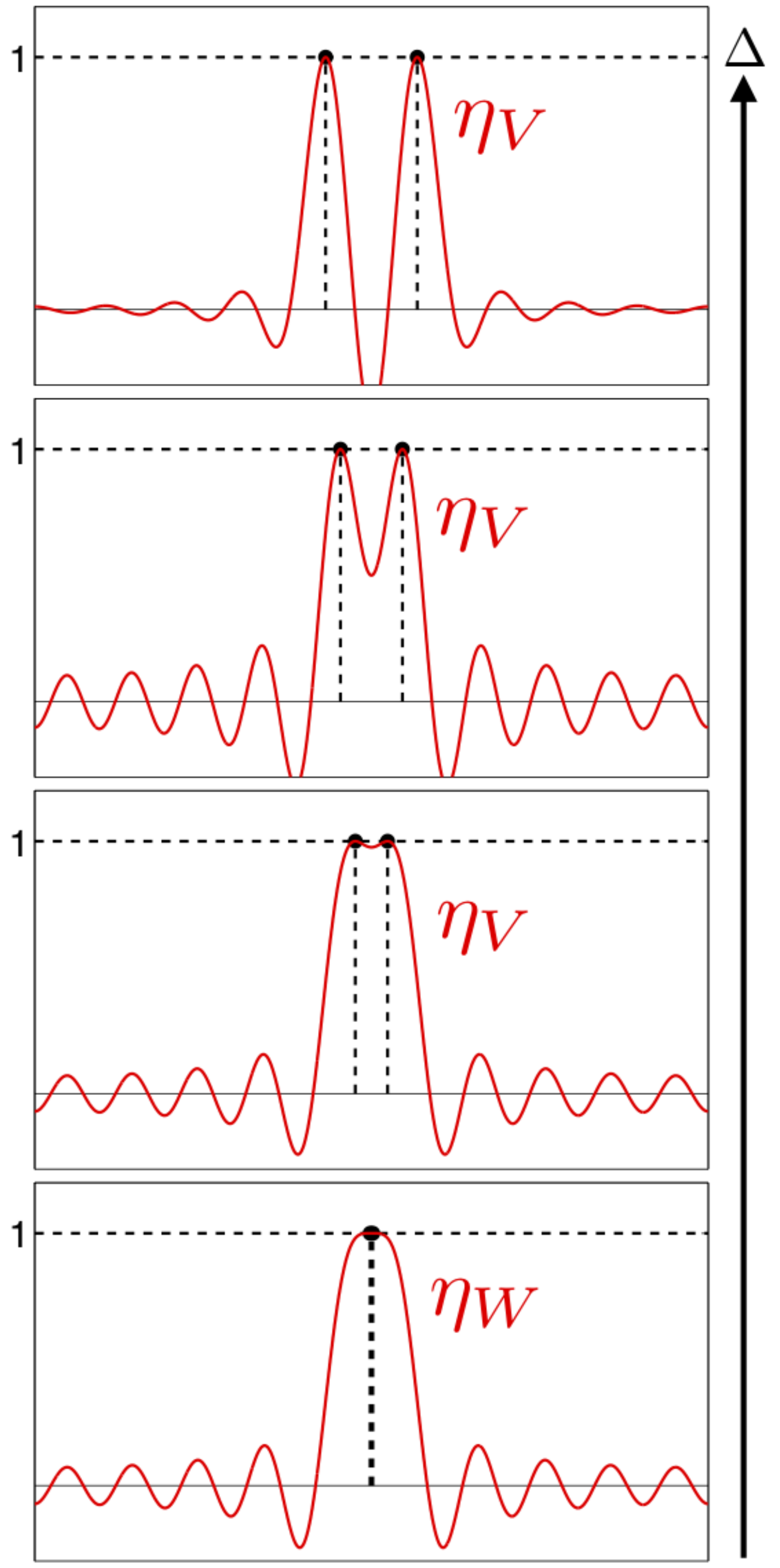
When is it non-degenerate?

Asymptotic vanishing derivatives precertificate in 1D

Theorem (Denoyelle et al, 2015):

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

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



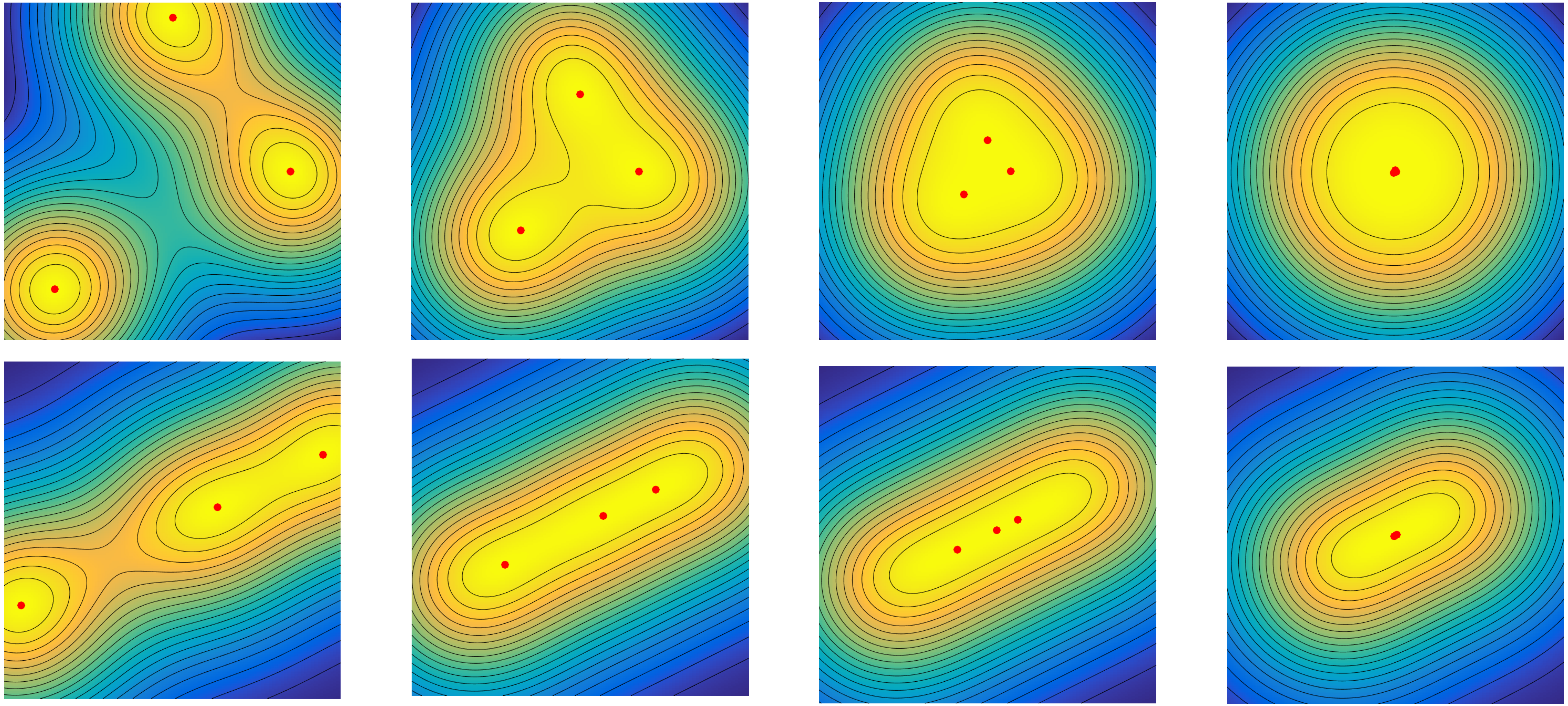
This is called non-degenerate if

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

$\eta_{V,tx}$ 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.

The multivariate limiting certificate

Theorem (sufficiency)

Given 2 spikes spaced t apart, η_W non degenerate and $\|w\|/\lambda = \mathcal{O}(1)$, $\lambda = \mathcal{O}(t^4)$, then $P_\lambda(\Phi\mu_{a,tx} + w)$ recovers exactly 2 spikes and $|(a,x) - (\hat{a},\hat{x})|_\infty \lesssim (\lambda + \|w\|)/t^3$.

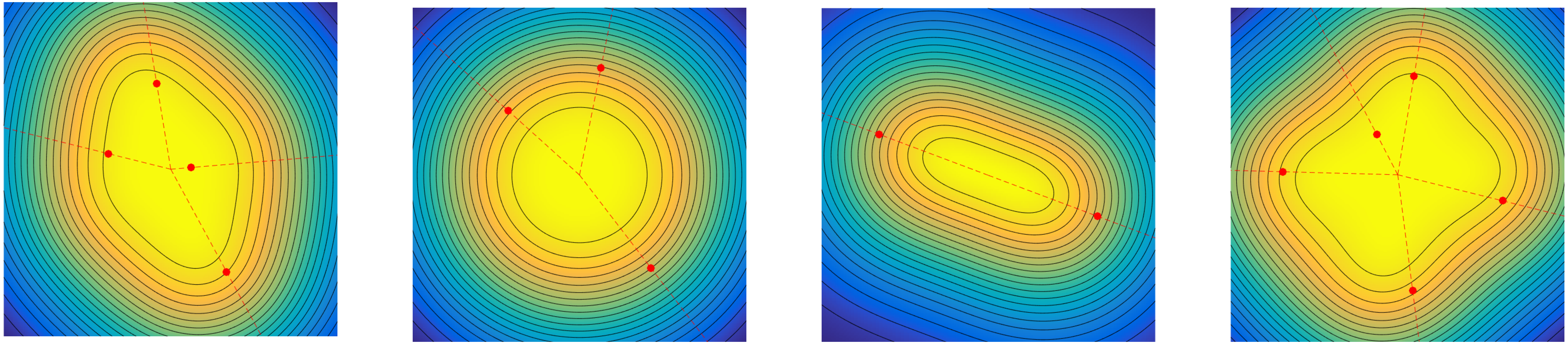
Theorem (necessity):

If there exists $t_n \to 0$ and $(a_n, Z_n) \in \mathbb{R}_+^N \times \mathcal{X}^N$ with $Z_n \to Z_0$ such that μ_{a_n,t_nZ_n} is support stable, then $\|\eta_{W,Z_0}\|_\infty = 1$

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

Gaussian convolution

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

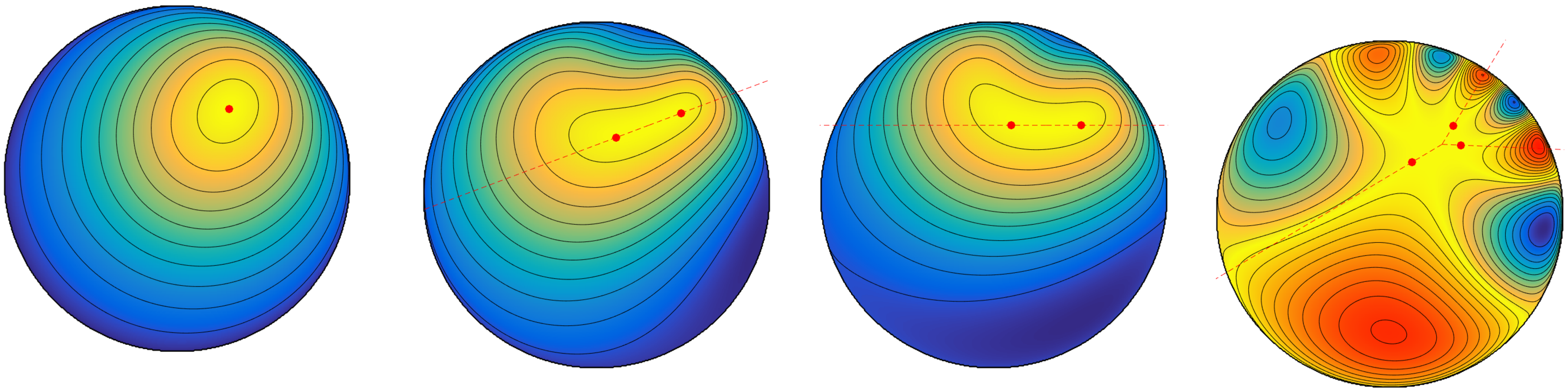


Numerical observation: $\eta_{W,z}$ is always uniformly bounded by 1.

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

Neuro-imaging

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



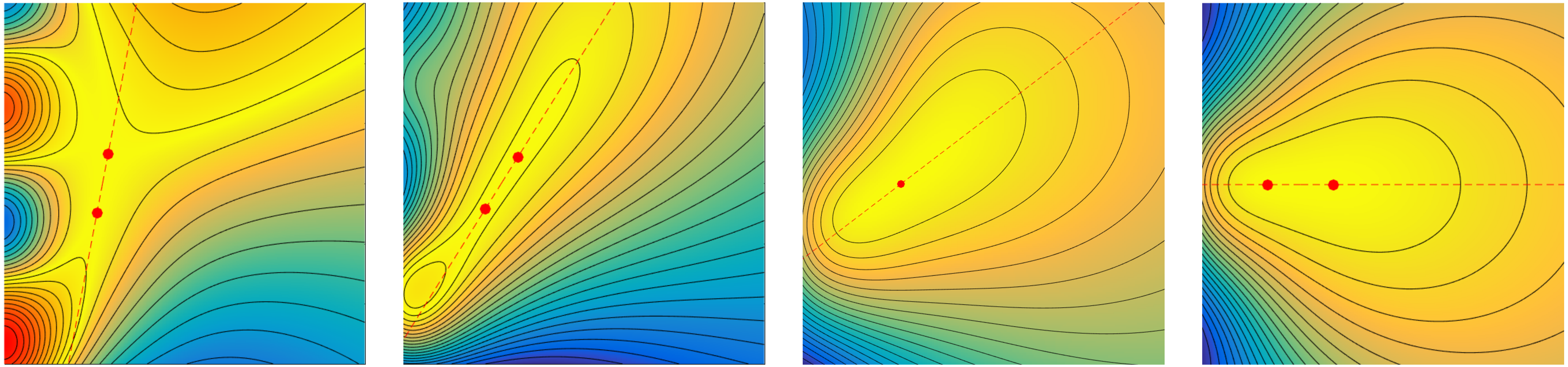
Numerical observation:

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



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.

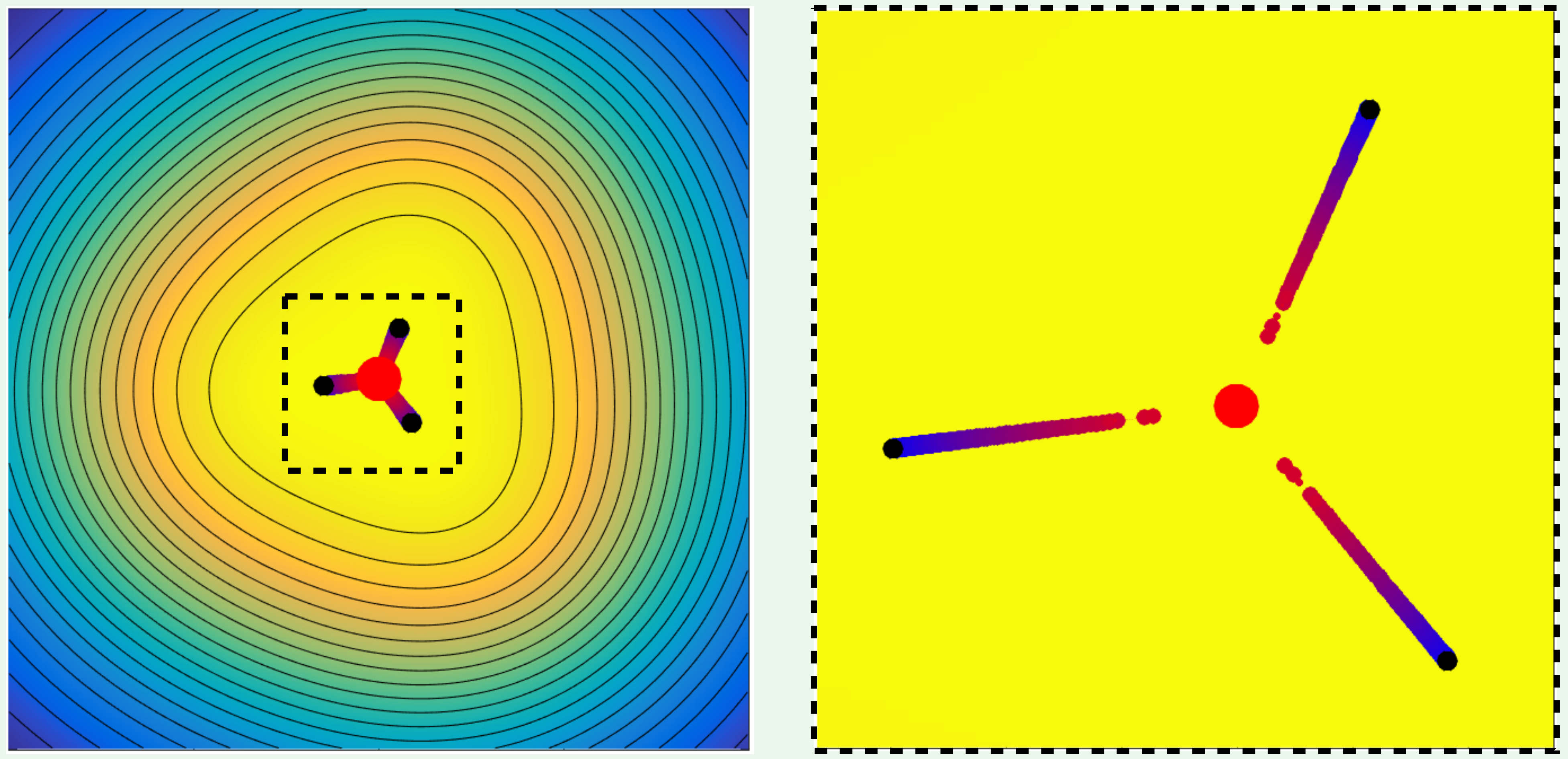
When η_W, z is **non-degenerate**

Evolution of solutions

Measurements:

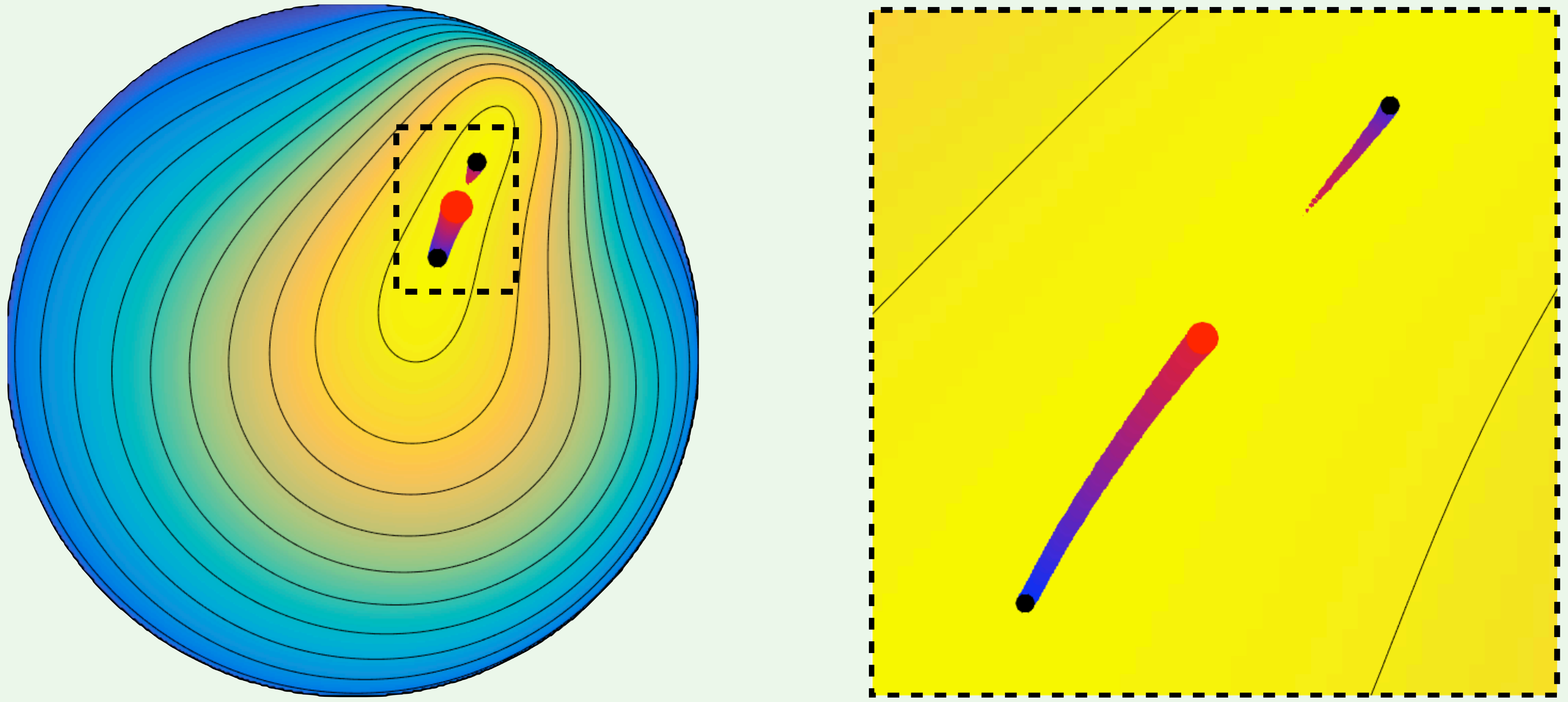
$$y = \Phi\mu_0 + \lambda w \text{ where } w = \Phi\hat{\mu} \text{ with } \hat{\mu} = \sum_{j=1}^{20} b_j \delta_{u_j}, \text{ where } b \in \mathcal{N}(0,\sigma^2) \text{ with } \sigma = 10^{-3}$$

Gaussian deconvolution, $N = 3$



$\eta_{W,Z}$ and μ_λ Zoom

Neuro-imaging, $N = 2$



$\eta_{W,Z}$ and μ_λ Zoom

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

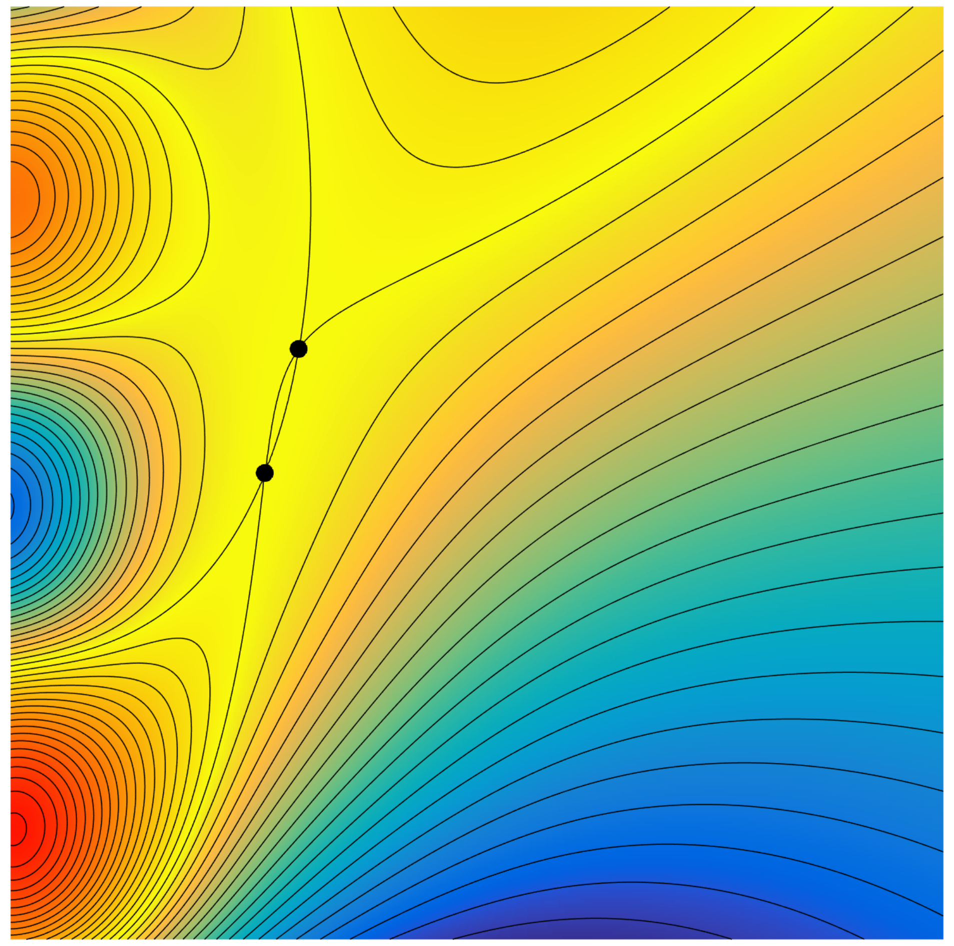
When η_W, z is **degenerate**

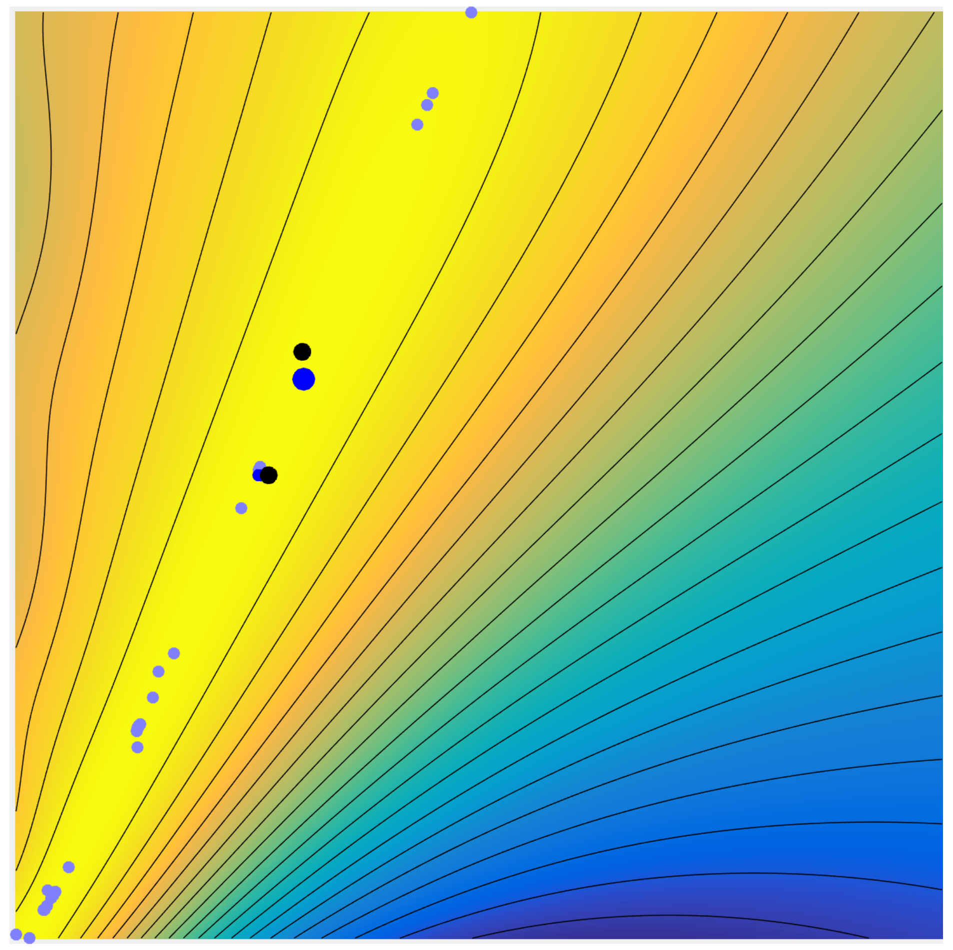
Evolution of solutions

Measurements:

$$y = \Phi\mu_0 + \lambda w \text{ where } w = \Phi\hat{\mu} \text{ with } \hat{\mu} = \sum_{j=1}^{20} b_j \delta_{u_j}, \text{ where } b \in \mathcal{N}(0,\sigma^2) \text{ with } \sigma = 10^{-3}$$

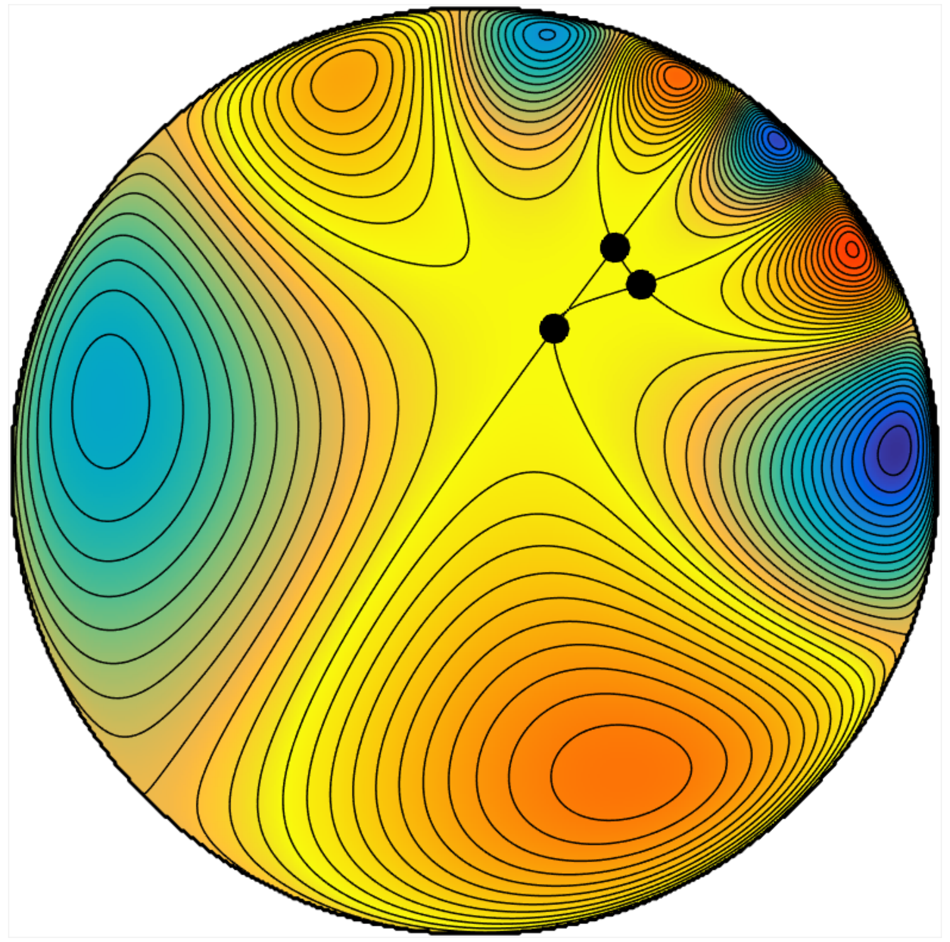
Gaussian mixture, $N = 2$

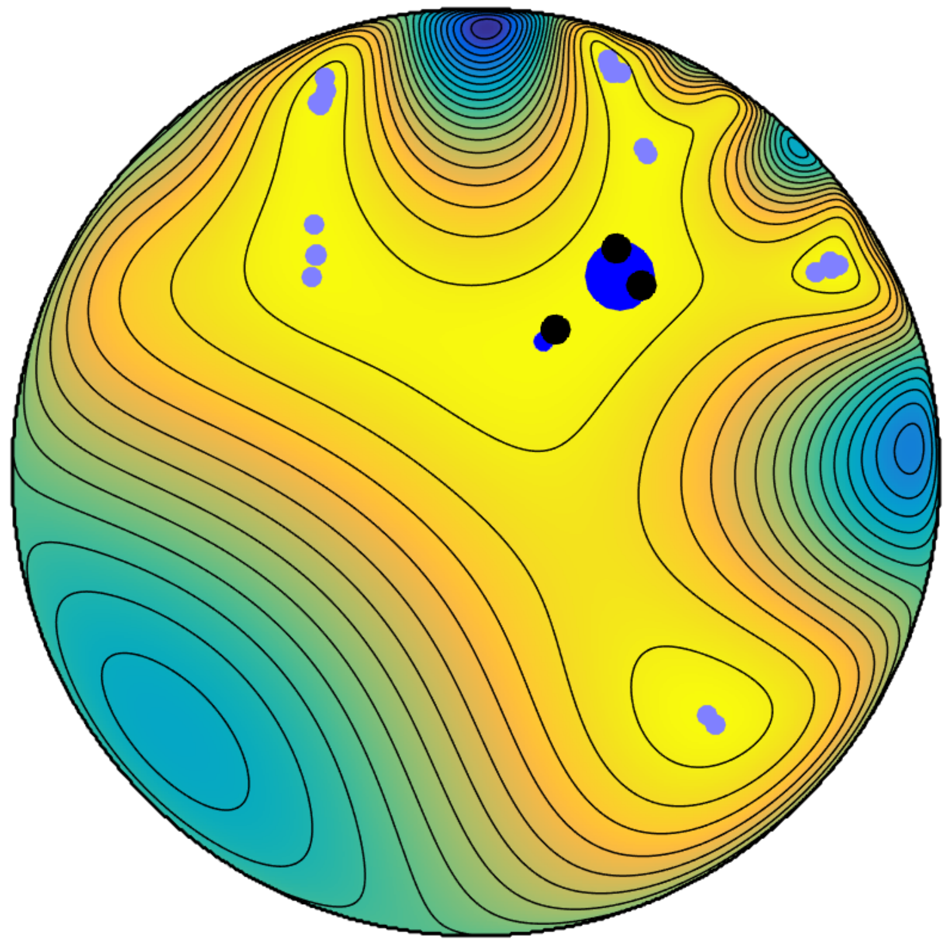




$\eta_{W,Z}$ μ_0 $\eta^{(\ell)}$ $\mu^{(\ell)}$

Neuro-imaging, $N = 3$





$\eta_{W,Z}$ μ_0 $\eta^{(\ell)}$ $\mu^{(\ell)}$

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 $\mathrm{sign}(a)$ 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 $$m = \mathcal{O}(s \log(s/\rho)^2 + \log(L/\rho))$$

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,

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

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 $x_1, \ldots, x_s \in \mathbb{R}^d$

➡ Sketch: Draw $\omega_1, \ldots, \omega_n$ iid from $\mathcal{N}(0, \Sigma^{-1}/d)$, $y := \frac{C}{n} \sum_{i=1}^{n} (\exp(-\sqrt{-1}\omega_k^\top z_i))_{k=1}^m$

$$y \approx \mathbb{E}_z[C \exp(-\sqrt{-1}\omega_k^\top z_i)] = \Phi \mu_0$$

with $\mu_0 = \sum_{i=1}^{s} a_i \delta_{x_i}$ 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

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

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.