Kernel switch & adaptive near regions in continuous sparse regularization: illustration on sketched mixtures

Nicolas Jouvin (MIA PS)

Joint work with Yohann De Castro (ICJ) & Rémi Gribonval (INRIA)

Séminaire de statistiques du LPSM - 17 février 2026

Motivation: sparse measure recovery...

Sparse measures

$$\mu = \sum_{k=1}^{p} a_k \delta_{\boldsymbol{x}_k}, \quad a_k \in \mathbb{R}, \quad \boldsymbol{x}_k \in \mathcal{X} \subset \mathbb{R}^d \text{ compact}$$

Measurement operator / acquisition process

Let $\phi : \mathcal{X} \to \mathcal{F}$ Hilbert, observe s “spikes”

$$y := \sum_{k=1}^{s} a_k^0 \phi(\boldsymbol{x}_k^0) + \text{noise} \in \mathcal{F}$$

Example: Deconvolution

From Clarice Poon slides

$\boldsymbol{x}^0$: star position
$\phi(\boldsymbol{x})$: telescope blur around $\boldsymbol{x}$

Motivation: sparse measure recovery...

Sparse measures

$$\mu = \sum_{k=1}^{p} a_k \delta_{\boldsymbol{x}_k}, \quad a_k \in \mathbb{R}, \quad \boldsymbol{x}_k \in \mathcal{X} \subset \mathbb{R}^d \text{ compact}$$

Measurement operator / acquisition process

Let $\phi : \mathcal{X} \to \mathcal{F}$ Hilbert, observe s "spikes"

$$y := \sum_{k=1}^{s} a_k^0 \phi(\boldsymbol{x}_k^0) + \text{noise} \in \mathcal{F}$$

Example: Deconvolution

From Clarice Poon slides

$\boldsymbol{x}^0$: star position

$\phi(\boldsymbol{x})$: telescope blur around $\boldsymbol{x}$

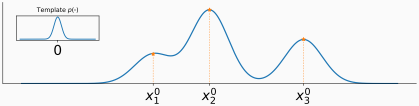
Inverse problem: recover $(a_k^0, \boldsymbol{x}_k^0)_{k=1}^s$ from (noisy) measurements y

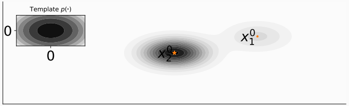
...and its application to mixture modeling

Context: observe sample $Z = \{z_1, \dots, z_n\} \subset \mathbb{R}^d$ from a mixture density

$$z_i \sim f^0(z) = \sum_{k=1}^{s} a_k^0 p(z - \boldsymbol{x}_k^0), \qquad \boldsymbol{x}_k^0 \in \mathcal{X}$$

with $p \in L^1$ some known density function, *e.g.* Gaussian $p = \mathcal{N}(0, \Sigma)$





Goal: estimate parameters $(a_k^0, \boldsymbol{x}_k^0)_{k=1}^s$ of model

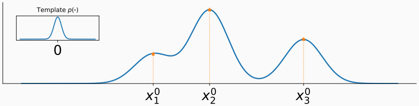
- Usually: MLE, theoretical analysis & algorithm (EM), ...

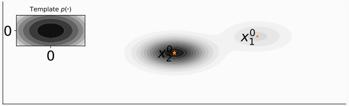
...and its application to mixture modeling

Context: observe sample $Z = \{z_1, \ldots, z_n\} \subset \mathbb{R}^d$ from a mixture density

$$z_i \sim f^0(z) = \sum_{k=1}^{s} a_k^0 p(z - \boldsymbol{x}_k^0), \qquad \boldsymbol{x}_k^0 \in \mathcal{X}$$

with $p \in L^1$ some known density function, *e.g.* Gaussian $p = \mathcal{N}(0, \Sigma)$





Goal: estimate parameters $(a_k^0, \boldsymbol{x}_k^0)_{k=1}^s$ of model

- Usually: MLE, theoretical analysis & algorithm (EM), ...
- **This talk:** linear inverse problem on the space of measure (De Castro et al. 2021)

Today's talk is about

Sparse regularization with BLASSO, model kernel & dual certificates

Dual certificates by pivoting to a new kernel

Illustration on sketched mixture model estimation

Sparse regularization with BLASSO, model kernel & dual certificates

Supermix: mixtures as sparse measure recovery (De Castro et al. 2021)

Observe $\boldsymbol{z}_{1:n} \sim f^0$ from the mixture density. Target sparse measure

$$\mu^0 = \sum_{k=1}^{s} a_k^0 \delta_{\boldsymbol{x}_k^0}$$

Convolution we have $f^0 = \Phi\mu^0$ with the linear operator

$$\Phi : \mu \in \mathcal{M}(\mathcal{X}) \mapsto \Phi\mu := \int p(\cdot - \boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) = p \star \mu$$

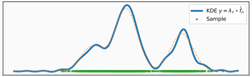
Data-fitting term ? $\rightsquigarrow$ we only have an n-sample $\boldsymbol{z}_{1:n} \sim f^0$. How to compare

$$\hat{f}_n = \frac{1}{n}\sum_{j=1}^{n} \delta_{\boldsymbol{z}_j} \quad \overset{?}{\approx} \quad \Phi\mu \quad \longrightarrow \text{ not in the same space}$$

Supermix (cont'd)

Solution De Castro et al. (2021) embed in an Hilbert $\mathcal{F}$

$$\underbrace{\lambda_\tau \star \hat{f}_n}_{\text{KDE}} \overset{\mathcal{F}}{\approx} \underbrace{\lambda_\tau \star \Phi\mu}_{\text{Smoothed mixture}}$$



using a smoothing kernel λ with bandwidth τ

$$\lambda_\tau(\cdot) = \tau^{-d}\lambda(\cdot/\tau), \quad \text{this talk: } \lambda(x) = \mathrm{sinc}(x) = \prod_{l=1}^{d} \frac{\sin(x_l)}{x_l}$$

- Associated Hilbert $\mathcal{F}_\tau$ = band-limited functions on $[-\delta, \delta]^d$, $\delta = 1/\tau$
- RKHS embedding operator $L_\tau \nu := \lambda_\tau \star \nu$
- Observation $y = L_\tau \hat{f}_n$ is the KDE
- Lifting on measure: find μ such that y and $F\mu = (L_\tau \circ \Phi)\mu$ are close

$$\|\underbrace{L_\tau \hat{f}_n}_{\text{KDE}} - \underbrace{(L_\tau \circ \Phi)\mu}_{\text{Smoothed mixture}}\|_{\mathcal{F}_\tau}$$

General setting: lifting on the space of measures

Measurement process linear operator on measure $\mathcal{M}(\mathcal{X})$

$$F : \begin{array}{ccc} \mathcal{M}(\mathcal{X}) & \to & \mathcal{F} \\ \mu & \mapsto & F\mu \end{array}, \quad \text{Observe } y = F\mu_0 + \Gamma \in \mathcal{F}, \quad \text{noise level } \|\Gamma\|_{\mathcal{F}} \leq \gamma$$

Model kernel $\rightsquigarrow$ F induces a reproducing kernel on $\mathcal{X}$

$$K_{\text{mod}}(\boldsymbol{s}, \boldsymbol{t}) := \langle F\delta_{\boldsymbol{s}}, F\delta_{\boldsymbol{t}} \rangle_{\mathcal{F}}$$

(Assumption) $K_{\text{mod}} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is continuous

- **Mixture example:** $F = L_\tau \circ \Phi$ and $K_{\text{mod}}(\boldsymbol{s}, \boldsymbol{t}) = (\lambda_\tau \star p \star \check{p})(\boldsymbol{t} - \boldsymbol{s})$
- The RKHS $\mathcal{H}_{\text{mod}} \subset \mathcal{F}$ contains continuous functions
- Adjoint $F^\star : \mathcal{F} \to \mathcal{H}_{\text{mod}}$

Lifting on the space of measures (cont'd)

Measurement process

$\boldsymbol{a}^0, \boldsymbol{x}^0$ Parameter space $\mathcal{X}$

Lifting on the space of measures (cont'd)

Measurement process

$\boldsymbol{a}^0, \boldsymbol{x}^0$ Parameter space $\mathcal{X}$

$\downarrow$ Lifting $(\boldsymbol{a}^0, \boldsymbol{x}^0) \mapsto \mu^0$

$$\mu^0 = \sum_{k=1}^{s} a_k^0 \delta_{\boldsymbol{x}_k^0}$$ Measure space $\mathcal{M}(\mathcal{X})$

Lifting on the space of measures (cont'd)

Measurement process

$\boldsymbol{a}^0, \boldsymbol{x}^0$ Parameter space $\mathcal{X}$

$\downarrow$ Lifting $(\boldsymbol{a}^0, \boldsymbol{x}^0) \mapsto \mu^0$

$\mu^0 = \sum_{k=1}^{s} a_k^0 \delta_{\boldsymbol{x}_k^0}$ Measure space $\mathcal{M}(\mathcal{X})$

$\downarrow$ **Linear measurement** $F\mu +$ noise

$y = F\mu^0 + \Gamma$ Hilbert $\mathcal{F}$

Inverse problem

Continuous sparse regression

Beurling Lasso (De Castro and Gamboa 2012; Duval and Peyré 2015)

For $\kappa > 0$, the Beurling LASSO estimator solves

$$\widehat{\mu} \in \underset{\mu \in \mathcal{M}(\mathcal{X})}{\arg\min}\, J_\kappa(\mu) := \frac{1}{2}\|y - F\mu\|_{\mathcal{F}}^2 + \kappa\|\mu\|_{TV} \qquad \text{(BLASSO)}$$

If F is continuous then there exists (at least one) $\widehat{\mu}$

Total variation norm $\|\mu\|_{\mathrm{TV}} \rightsquigarrow$ **continuous analog of l^1-norm**

$$\text{(Discrete)} \quad \mu = \sum_{k=1}^{p} a_k \delta_{\boldsymbol{x}_k} \quad \longrightarrow \quad \|\mu\|_{\mathrm{TV}} = \sum_{k=1}^{p} |a_k| = \|a\|_1$$

$$\text{(Continuous)} \quad \mathrm{d}\mu = f\,\mathrm{d}\lambda \quad \longrightarrow \quad \|\mu\|_{\mathrm{TV}} = \|f\|_{L^1}$$

- Optimization over the space of measures $\mathcal{M}(\mathcal{X})$ (Banach)
- Convex problem in μ with theoretical analysis of $\widehat{\mu}$
- Computational side : Franck-Wolfe (Denoyelle et al. 2019), particle GD (Chizat 2019)

Question: how well does BLASSO estimator $\widehat{\mu}$ localizes around μ^0 ?

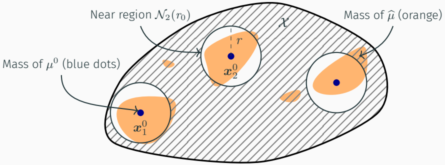
State of the art analysis for support recovery [1] (Poon, Keriven, and Peyré 2023)

- stated in term of near/far region around $\operatorname{Supp}\mu^0$

Radius $r > 0$, distance $\mathfrak{d}(\cdot,\cdot)$

Near: $\mathcal{N}_k(r) := \{\boldsymbol{x}, \mathfrak{d}(\boldsymbol{x}, \boldsymbol{x}_k^0) \leq r\}$

Far: $\mathcal{F}(r) := \mathcal{X} \setminus \cup_k \mathcal{N}_k(r)$



Adapted from Giard, De Castro, and Marteau (2025)

- relies on the construction of a **non-degenerate dual certificate**

[1] See also: De Castro and Gamboa (2012), Bredies and Pikkarainen (2013), and Duval and Peyré (2015)

Dual problem

Convex program The BLASSO problem admit a (pre)-dual

$$\inf_{\mu \in \mathcal{M}(\mathcal{X})} \frac{1}{2}\|y - F\mu\|_{\mathcal{F}}^2 + \kappa\|\mu\|_{TV} \quad \xrightarrow{\text{Dual}} \quad \inf_{c \in \mathcal{F}\,:\,\|F^\star c\|_\infty \leq 1} \langle c, y\rangle_{\mathcal{F}} - \frac{\kappa}{2}\|c\|_{\mathcal{F}}^2 \qquad (\mathcal{D}_\kappa)$$

With the adjoint $F^\star : \mathcal{F} \to \mathcal{H}_{\text{mod}}$

Optimality conditions yield for a BLASSO minimizer $\widehat{\mu}$

$$0 \in \partial J_\kappa(\widehat{\mu}) \iff 0 \in \partial\|\widehat{\mu}\|_{\mathrm{TV}} + \frac{1}{\kappa}F^\star(y - F\widehat{\mu}) \iff \eta \coloneqq -\frac{1}{\kappa}F^\star(y - F\widehat{\mu}) \in \partial\|\widehat{\mu}\|_{\mathrm{TV}}$$

We have

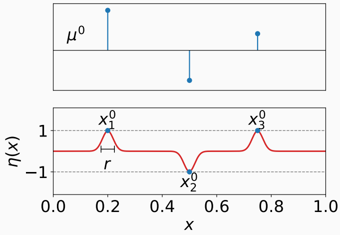
- $\mathrm{Supp}\,\widehat{\mu} \subset \{\boldsymbol{x} \in \mathcal{X},\ |\eta(\boldsymbol{x})| = 1\} \longrightarrow \eta$ *certifies* the support of $\widehat{\mu}$
- Natural to look for certificates $\eta^0 \in \mathrm{Im}(F^\star)$ s.t. $\eta^0 \in \partial\|\mu^0\|_{\mathrm{TV}}$

Non-degenerate dual certificate

Find η^0 subgradient of $\|\mu^0\|_{\mathrm{TV}}$ with additional controls on near/far

Fix radius $r > 0$, distance $\mathfrak{d}(\cdot,\cdot)$ & control $\epsilon_0, \epsilon_2 > 0$.
η^0 is a $(r, \epsilon_0, \epsilon_2)$-**non-degenerate** certificate iff

- $\eta^0 \in \mathrm{Im}(F^\star) = \mathcal{H}_{\text{mod}}$
- $\eta^0(\boldsymbol{x}_k^0) = \text{sign}(a_k^0)$
- $\eta^0(\boldsymbol{x}) \leq 1 - \epsilon_0$, $\forall \boldsymbol{x} \in \mathcal{F}(r)$
- $\eta^0(\boldsymbol{x}) \leq 1 - \epsilon_2 \mathfrak{d}(\boldsymbol{x}, \boldsymbol{x}_k^0)^2$, $\forall \boldsymbol{x} \in \mathcal{N}_k(r)$

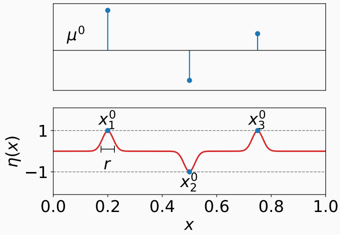


Non-degenerate dual certificate

Find η^0 subgradient of $\|\mu^0\|_{\mathrm{TV}}$ with additional controls on near/far

Fix radius $r > 0$, distance $\mathfrak{d}(\cdot,\cdot)$ & control $\epsilon_0, \epsilon_2 > 0$.
η^0 is a $(r, \epsilon_0, \epsilon_2)$-**non-degenerate** certificate iff

- $\eta^0 \in \mathrm{Im}(F^\star) = \mathcal{H}_{\text{mod}}$
- $\eta^0(\boldsymbol{x}_k^0) = \text{sign}(a_k^0)$
- $\eta^0(\boldsymbol{x}) \leq 1 - \epsilon_0, \forall \boldsymbol{x} \in \mathcal{F}(r)$
- $\eta^0(\boldsymbol{x}) \leq 1 - \epsilon_2 \mathfrak{d}(\boldsymbol{x}, \boldsymbol{x}_k^0)^2, \forall \boldsymbol{x} \in \mathcal{N}_k(r)$



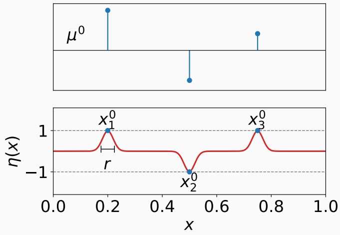
Certificates are the main workhorse: $\exists$ NDC $\implies$ recovery guarantees

Non-degenerate dual certificate

Find η^0 subgradient of $\|\mu^0\|_{\mathrm{TV}}$ with additional controls on near/far

Fix radius $r > 0$, distance $\mathfrak{d}(\cdot,\cdot)$ & control $\epsilon_0, \epsilon_2 > 0$.
η^0 is a $(r, \epsilon_0, \epsilon_2)$-**non-degenerate** certificate iff

- $\eta^0 \in \mathrm{Im}(F^\star) = \mathcal{H}_{\mathsf{mod}}$
- $\eta^0(\boldsymbol{x}_k^0) = \mathsf{sign}(a_k^0)$
- $\eta^0(\boldsymbol{x}) \leq 1 - \epsilon_0$, $\forall \boldsymbol{x} \in \mathcal{F}(r)$
- $\eta^0(\boldsymbol{x}) \leq 1 - \epsilon_2 \mathfrak{d}(\boldsymbol{x}, \boldsymbol{x}_k^0)^2$, $\forall \boldsymbol{x} \in \mathcal{N}_k(r)$



Certificates are the main workhorse: $\exists$ NDC $\implies$ recovery guarantees

How ? Allows to control the **Bregman divergence** associated to η^0

$$D_{\eta^0}(\widehat{\mu}||\mu^0) := \|\widehat{\mu}\|_{\mathrm{TV}} - \|\mu^0\|_{\mathrm{TV}} - \int \eta^0 \,\mathrm{d}(\widehat{\mu} - \mu^0)$$

$$\leq (\dots) \leq \frac{(\gamma + \kappa \|\eta^0\|_{\mathcal{H}_{\mathsf{mod}}})}{2\kappa}$$

Recovery guarantees for BLASSO

Note dependency on noise & $\mathcal{H}_{\text{mod}}$

BLASSO recovery (informal)

Suppose $\eta^0 \in \mathcal{H}_{\text{mod}}$ is a $(r_0, \epsilon_0, \epsilon_2)$-non-degenerate certificate. For a noise $\|\Gamma\|_{\mathcal{F}} \leq \gamma$ and regularization $\kappa \propto \frac{\gamma}{\|\eta^0\|_{\mathcal{H}_{\text{mod}}}}$ we have

1. *Small mass on far region:*

$$|\widehat{\mu}|\big(\mathcal{F}(r_0)\big) \lesssim_d \gamma \|\eta^0\|_{\mathcal{H}_{\text{mod}}} ,$$

2. *Mass of near regions* $\sim a_k^0$*:*

$$\left|\widehat{\mu}(\mathcal{N}_k(r_0)) - a_k^0\right| \lesssim_d \gamma \|\eta^0\|_{\mathcal{H}_{\text{mod}}} ,$$

3. *Detection level:* For all borelian $A \subset \mathcal{X}$ such that $|\widehat{\mu}|(A) \gtrsim_d \gamma \|\eta^0\|_{\mathcal{H}_{\text{mod}}}$,

$$\exists \boldsymbol{x}_k^0, \quad \mathfrak{d}(A, \boldsymbol{x}_k^0) := \min_{\boldsymbol{t} \in A} \mathfrak{d}(t, \boldsymbol{x}_k^0) \lesssim_d r_0 ,$$

Separation is crucial: parameters needs to be "sufficiently separated"

$$\min_{k,l=1,\dots,s} \mathfrak{d}(\boldsymbol{x}_k^0, \boldsymbol{x}_l^0) \geq \Delta$$

A geometric framework: Fisher-Rao distance on $\mathcal{X}$

Separation is crucial: parameters needs to be "sufficiently separated"

$$\min_{k,l=1,\dots,s} \mathfrak{d}(\boldsymbol{x}_k^0, \boldsymbol{x}_l^0) \geq \Delta$$

How to measure separation ? $\rightsquigarrow$ Fisher-Rao distance associated to K_{mod}

For a kernel $K(\cdot,\cdot)$, Poon, Keriven, and Peyré (2023) define

$$\text{Fisher metric: } \mathfrak{g}_{\boldsymbol{x}} = \partial_1 \partial_2 K(\boldsymbol{x}, \boldsymbol{x}) \in \mathbb{R}^{d \times d}$$

$$\rightsquigarrow \text{Geodesic distance: } \mathfrak{d}_{\mathfrak{g}}(\boldsymbol{s}, \boldsymbol{t}) = \inf_{p: \boldsymbol{s} \to \boldsymbol{t}} \int_0^1 \sqrt{p'(u)^\top \mathfrak{g}_{p(t)} p'(u)}\, \mathrm{d}u$$

Amounts to Mahanalobis when $K(\boldsymbol{s}, \boldsymbol{t}) = \rho(\boldsymbol{s} - \boldsymbol{t})$ is translation-invariant (TI)

$$\mathfrak{g}_{\boldsymbol{x}} \equiv \mathfrak{g} = -\nabla^2 \rho(0) \qquad \mathfrak{d}_{\mathfrak{g}}(\boldsymbol{s}, \boldsymbol{t}) = \|\mathfrak{g}^{1/2}(\boldsymbol{s} - \boldsymbol{t})\|_2$$

Riemannian geometry framework with derivatives $K^{(ij)}(\boldsymbol{s}, \boldsymbol{t})$

Local positive curvature (LPC) assumption

Thankfully we've got Poon, Keriven, and Peyré (2023) with the

Local positive curvature (LPC) assumption

If model kernel K_{mod} satisfies

[...Technical condition on K_{mod}...]

Then there exists a $(\bar{\epsilon}_0, \bar{\epsilon}_2, r_0)$-non-degenerate certificate for the Fisher-Rao metric associated to K_{mod}.

Local positive curvature (LPC) assumption

Thankfully we've got Poon, Keriven, and Peyré (2023) with the

Local positive curvature (LPC) assumption

If model kernel K_{mod} satisfies

$$\forall 0 \leq i,j \leq 2, i+j \leq 3, B_{ij} := \sup_{\boldsymbol{s},\boldsymbol{t}\in\mathcal{X}} \left\| K^{(i,j)}(\boldsymbol{s},\boldsymbol{t}) \right\|_{\boldsymbol{s},\boldsymbol{t}} < +\infty \,,$$

$\exists r_0 \in (0, 1/\sqrt{B_{02}})$ such that

$$\bar{\varepsilon}_0 := \sup_{\varepsilon \geq 0} \left\{ \varepsilon \,:\, K(\boldsymbol{s},\boldsymbol{t}) \leq 1-\varepsilon \,,\, \forall \boldsymbol{s},\boldsymbol{t} \in \mathcal{X} \text{ s.t. } \mathfrak{d}_{\mathfrak{g}}(\boldsymbol{s},\boldsymbol{t}) \geq r_0 \right\} < +\infty$$

$$\bar{\varepsilon}_2 := \sup_{\varepsilon \geq 0} \left\{ \varepsilon \,:\, -K^{(0,2)}(\boldsymbol{s},\boldsymbol{t})[\boldsymbol{v},\boldsymbol{v}] \geq \varepsilon \|\boldsymbol{v}\|_{\boldsymbol{t}}^2 \,,\, \forall \boldsymbol{v} \in \mathbb{T}_{\boldsymbol{t}} \,,\, \forall \boldsymbol{s},\boldsymbol{t} \in \mathcal{X} \text{ s.t. } \mathfrak{d}_{\mathfrak{g}}(\boldsymbol{s},\boldsymbol{t}) < r_0 \right\} < +\infty$$

$\Delta_0 < +\infty,$ with: $\Delta_0 = \inf \left\{ \Delta \,:\, \sum_{l=2}^{s} \|K^{(i,j)}(\boldsymbol{x}_1,\boldsymbol{x}_l)\|_{\boldsymbol{x}_1,\boldsymbol{x}_l} \leq \min\left(\frac{\bar{\varepsilon}_0}{B_0}, \frac{2\bar{\varepsilon}_2}{B_2}\right),\, i,j=0,\ldots,2,\, \min_{k,l=1,\ldots,s} \mathfrak{d}_{\mathfrak{g}}(\boldsymbol{x}_k^0,\boldsymbol{x}_l^0) \geq \Delta \right\}$

Then there exists a $(\bar{\epsilon}_0, \bar{\epsilon}_2, r_0)$-non-degenerate certificate for Δ_0-separated target μ^0. With $\mathfrak{d}_{\mathfrak{g}}$ the Fisher-Rao metric associated to K_{mod}.

Local positive curvature (LPC) assumption

Thankfully we've got Poon, Keriven, and Peyré (2023) with the

Local positive curvature (LPC) assumption

If model kernel K_{mod} satisfies

[...Technical condition on K_{mod}...]

Then there exists a $(\bar{\epsilon}_0, \bar{\epsilon}_2, r_0)$-non-degenerate certificate for the Fisher-Rao metric associated to K_{mod}.

So, problem solved right ?

Problem not solved...

1. New measurement operator F ? You need to prove LPC for your new K_{mod}...
   $\rightsquigarrow$ ... or leverage on known LPC-kernel: *pivot kernels*

2. Radius r of near/far is **fixed** but in statistical applications $\gamma \sim n^{-1/2}$
   $\rightsquigarrow$ make the radius r_γ adaptive to noise

Dual certificates by pivoting to a new kernel

Few known LPC kernel

LPC has been proven for the following kernels (Poon, Keriven, and Peyré 2023)

	Jackson	Gaussian	Laplace
$\mathcal{X}$	$(\mathbb{R}\backslash\mathbb{Z})^d$	$\mathbb{R}^d$	$\mathbb{R}^d$
$K(\boldsymbol{s},\boldsymbol{t})$	$\prod_{i=1}^d \left(\frac{\sin\left(f_c\pi(s_i-t_i)\right)}{f_c\sin(\pi(s_i-t_i))}\right)^4$	$\exp\left(-\frac{1}{2}\lVert\boldsymbol{s}-\boldsymbol{t}\rVert_\Omega^2\right)$	$\prod_{i=1}^d \frac{2\sqrt{(s_i+\alpha_i)(t_i+\alpha_i)}}{s_i+t_i+2\alpha_i}$
$\mathcal{O}(\Delta_0)$	$\sqrt{d\sqrt{s}}$	$\log(s)$	$d+\log(d^{3/2}s)$
$\mathcal{O}(r_0)$	1	1	1

Computation heavy and no general recipes

Increment the list with sinc-4 kernel (De Castro et al. 2021)

Sinc-4 kernel: 4-th power of the sinus cardinal filter with bandwidth τ

$$K_\tau(\boldsymbol{s}, \boldsymbol{t}) = \Psi_\tau(\boldsymbol{s} - \boldsymbol{t}) = \operatorname{sinc}\left(\frac{\boldsymbol{s} - \boldsymbol{t}}{4\tau}\right)^4,$$

De Castro, Gribonval, and Jouvin (Theorem 4, 2025)

The sinc-4 kernel Ψ_τ satisfies the LPC with

$$\Delta_0 \propto s^{1/4} d^{7/4}, \ r_0 \sim \frac{1}{d}, \ \bar{\varepsilon}_0 \sim \frac{1}{d^3} \text{ and } \bar{\varepsilon}_2 \sim 1.$$

Fisher-Rao is a re-scaled euclidean metric:

$$\mathfrak{g} = -\nabla^2 \Psi_\tau(0) \propto \frac{1}{\tau} \mathrm{Id}_d \implies \mathfrak{d}_{\mathfrak{g},\tau}(\boldsymbol{s}, \boldsymbol{t}) \propto \frac{1}{\tau} \|\boldsymbol{s} - \boldsymbol{t}\|_2$$

Sufficient separation ? Just choose a bandwidth τ

$$\min_{k,l} \mathfrak{d}_{\mathfrak{g},\tau}(\boldsymbol{x}_k^0, \boldsymbol{x}_l^0) \geq \Delta_0 \iff \tau \lesssim \frac{\min_{k,l} \|\boldsymbol{x}_k^0 - \boldsymbol{x}_l^0\|_2}{\Delta_0}$$

Kernel switch: RKHS inclusion

Necessary condition the certificate η^0 is a continuous function in $\mathcal{H}_{\text{mod}}$

$$D_{\eta_0}(\widehat{\mu}, \mu^0) \lesssim \gamma \|\eta^0\|_{\mathcal{H}_{\text{mod}}}$$

Kernel switch: RKHS inclusion

Necessary condition the certificate η^0 is a continuous function in $\mathcal{H}_{\text{mod}}$

$$D_{\eta_0}(\widehat{\mu}, \mu^0) \lesssim \gamma \|\eta^0\|_{\mathcal{H}_{\text{mod}}}$$

Central idea switch to kernel K_{pivot} with RKHS $\mathcal{H}_{\text{pivot}} \subset \mathcal{H}_{\text{mod}}$

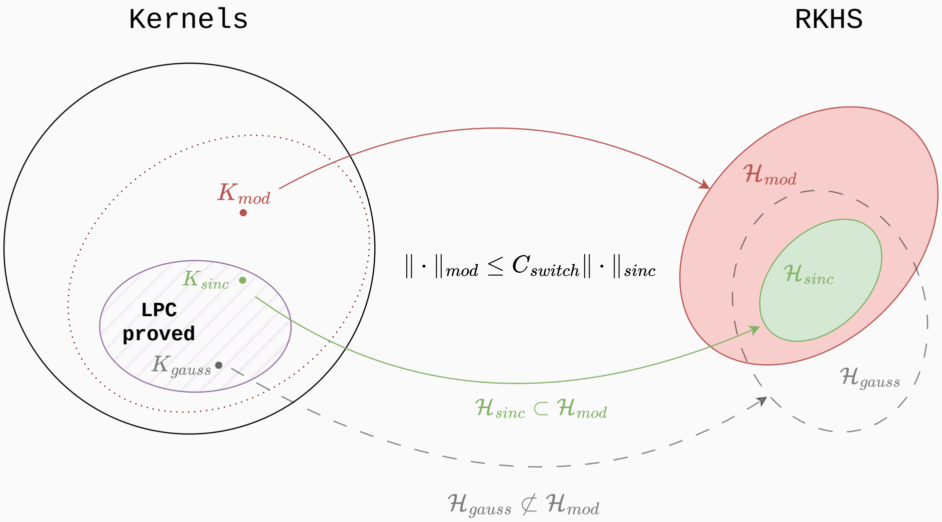
$$\|\eta\|_{\mathcal{H}_{\text{mod}}} \leq C_{\text{switch}} \|\eta\|_{\mathcal{H}_{\text{pivot}}}$$

$\rightsquigarrow$ *i.e.* the inclusion map $\mathrm{Id}_{\text{pivot,mod}} : \mathcal{H}_{\text{pivot}} \to \mathcal{H}_{\text{mod}}$ is continuous and

$$C_{\text{switch}}(K_{\text{mod}}, K_{\text{pivot}}) := \sup_{\eta \in \mathcal{H}_{\text{pivot}}} \frac{\|\eta\|_{\mathcal{H}_{\text{mod}}}}{\|\eta\|_{\mathcal{H}_{\text{pivot}}}} < \infty$$

is the *kernel switch constant.*

Kernel switch: an illustration



Recovery guarantees under kernel switch

De Castro, Gribonval, and Jouvin (2025, Theorem 1, informal)

Let K_{pivot} be a *pivot kernel* for K_{mod} with $C_{\text{switch}} < +\infty$. Suppose K_{pivot} satisfies LPC with parameters $(r_0, \epsilon_0, \epsilon_2)$. For a noise $\|\Gamma\|_{\mathcal{F}} \leq \gamma$, and $\kappa \propto \frac{\gamma}{C_{\text{switch}} \|\eta^0\|_{\mathcal{H}_{\text{pivot}}}}$

1. *Small mass on far region:*

$$|\widehat{\mu}|(\mathcal{F}(r_0)) \lesssim_d \gamma C_{\text{switch}} \|\eta^0\|_{\mathcal{H}_{\text{pivot}}} ,$$

2. *Mass of near regions* $\approx a_k^0$*:*

$$\left|\widehat{\mu}(\mathcal{N}_k(r_0)) - a_k^0\right| \lesssim_d \gamma C_{\text{switch}} \|\eta^0\|_{\mathcal{H}_{\text{pivot}}} ,$$

3. *Detection level:* For all borelian $A \subset \mathcal{X}$ such that $|\widehat{\mu}|(A) \gtrsim_d \gamma C_{\text{switch}} \|\eta^0\|_{\mathcal{H}_{\text{pivot}}}$,

$$\exists \boldsymbol{x}_k^0, \quad \mathfrak{d}_{\text{pivot}}(A, \boldsymbol{x}_k^0) \lesssim_d r_0 ,$$

Additionally, under LPC, we have $\|\eta^0\|_{\mathcal{H}_{\text{pivot}}} \leq \sqrt{s}$

The case of translation-invariant kernels

Natural question: how to choose the pivot and compute C_{switch} ?

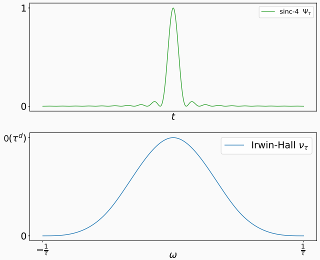
- General case is hard (Zhang and Zhao 2013)
- Special cases: **translation invariant**, radial, Hilbert-Schmidt

Bochner's Theorem TI kernels are (inverse) Fourier transforms

$$\exists \text{ spectral density } \nu, \quad K(\boldsymbol{s}, \boldsymbol{t}) = \rho(\boldsymbol{s} - \boldsymbol{t}) = \int e^{+\imath \boldsymbol{\omega}^\top (\boldsymbol{s} - \boldsymbol{t})} \nu(\boldsymbol{\omega}) \, \mathrm{d}\boldsymbol{\omega} \tag{1}$$

Example: The sinc-4 $\Psi_\tau(\boldsymbol{s} - \boldsymbol{t})$ is TI with

$\nu_{\text{pivot},\tau}$ the Irwin-Hall p.d.f. on $\mathbb{B}_\tau = \left[-\frac{1}{\tau}, \frac{1}{\tau}\right]^d$



The case of translation-invariant kernels (cont'd)

Switch constant If both $K_{\text{pivot}} = \rho_{\text{pivot}}$ and $K_{\text{mod}} = \rho_{\text{mod}}$ then

$$C_{\text{switch}}(K_{\text{mod}}, K_{\text{pivot}}) = \operatorname*{ess\,sup}_{\boldsymbol{\omega} \in \operatorname{Supp} \nu_{\text{pivot}}} \sqrt{\frac{\nu_{\text{pivot}}}{\nu_{\text{mod}}}(\boldsymbol{\omega})}$$

$\rightsquigarrow$ "simple" ratio of Fourier transforms

Example: sinc-4 pivot - Proposition 4.2 (De Castro, Gribonval, and Jouvin 2025)

- **Pivot:** Sinc-4 $K_{\text{pivot}} = \Psi_\tau$
- **Model:** K_{mod} such that $\mathbb{B}_\tau \subset \operatorname{Supp} \nu_{\text{mod}}$ (*e.g.* Gaussian)

$$C_{\text{switch}}(K_{\text{mod}}, \texttt{sinc-4}) \lesssim \sqrt{\frac{\tau^d}{\inf_{\boldsymbol{\omega} \in \mathbb{B}_\tau} \nu_{\text{mod}}(\boldsymbol{\omega})}}$$

ν_{mod} lower-bounded on $\mathbb{B}_\tau (\nu_{\text{mod}} > C_{d,\tau} > 0$ a.e. $) \implies C_{\text{switch}} < +\infty$

$\rightsquigarrow$ in practice kernel switch has been implicitly used in BLASSO literature

Illustration on sketched mixture model estimation

Reminder on Supermix

Solve the following BLASSO problem

$$\arg\min_{\mu} \frac{1}{2}\|\underbrace{L_\tau \hat{f}_n}_{y} - \underbrace{(L_\tau \circ \Phi)}_{F_\tau}\mu\|^2_{\mathcal{F}_\tau} + \kappa\|\mu\|_{TV} \qquad \text{(Supermix)}$$

- Analytic & TI $K_{\text{mod}} = \rho_{\text{mod}} = \lambda_\tau \star p \star \check{p}$
- **However:** LPC needs to be checked for each template p ($\mathfrak{d}_{\mathfrak{g}}$, Δ_0, r_0, etc.) $\rightsquigarrow$ **switch!**

To use sinc-4 as pivot, we need

1. *RKHS inclusion:* if $\inf_{\mathbb{B}_\tau} \mathfrak{F}[\lambda_\tau] \big| \mathfrak{F}[p](\boldsymbol{\omega})\big|^2 = C_{d,\tau} > 0$

$$C_{\text{switch}} = C_{\text{switch}}(\tau, p) := \sqrt{\frac{\tau^d}{C_{d,\tau}}} < +\infty, \qquad (\mathbf{H}_p)$$

2. *Separation condition:* for a target μ^0 choose τ s.t.

$$\tau \leq \tau_{\max}(\mu^0) := \frac{\min_{1\leq k\neq \ell \leq s} \left\|\boldsymbol{x}_k^0 - \boldsymbol{x}_\ell^0\right\|_2}{s^{1/4} d^{7/4}} . \qquad (\mathbf{H}_\tau)$$

Effective near region: beyond fixed radius

Guarantees hold under LPC with a radius r_0

What about noise ? We can show

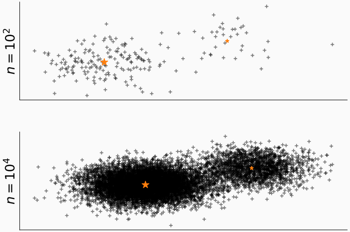
$$\gamma_n = \|y - F\mu^0\|_{\mathcal{F}} \leq \frac{1}{\sqrt{n}}$$

Effective near region: beyond fixed radius

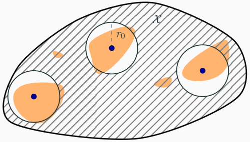
Guarantees hold under LPC with a radius r_0

What about noise ? We can show

$$\gamma_n = \|y - F\mu^0\|_{\mathcal{F}} \leq \frac{1}{\sqrt{n}}$$



Same radius r_0

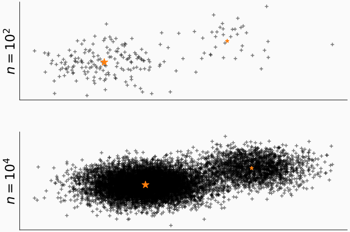


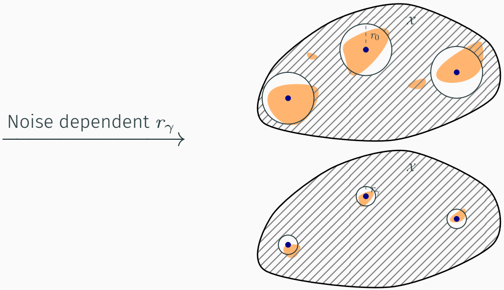
Effective near region: beyond fixed radius

Guarantees hold under LPC with a radius r_0

What about noise ? We can show

$$\gamma_n = \|y - F\mu^0\|_{\mathcal{F}} \leq \frac{1}{\sqrt{n}}$$





Sketching the kernel: random Fourier features (Rahimi and Recht 2007)

K_{mod} admits a weighted *random Fourier features* representation

$$K_{\text{mod}}(\boldsymbol{s},\boldsymbol{t}) = \mathbb{E}_{\boldsymbol{\omega}\sim\Lambda}\left[\varphi_{\boldsymbol{\omega}}(\boldsymbol{s})\overline{\varphi_{\boldsymbol{\omega}}(\boldsymbol{t})}\right], \qquad \begin{aligned} \varphi_{\boldsymbol{\omega}}(\boldsymbol{t}) &:= W(\omega)e^{-\imath\boldsymbol{\omega}^\top\boldsymbol{t}} \\ W(\omega) &= \sqrt{\frac{\nu_{\lambda_\tau}}{\Lambda}}\mathfrak{F}[p](\omega) \end{aligned}$$

- Computing the d-dimensional integral can be challenging
- Sketching $\rightsquigarrow$ Monte-Carlo approximation draw $\boldsymbol{\omega}_{1:m}\sim\Lambda$
- *Sketched model kernel*

$$K_{\text{sketch,mod}}(\boldsymbol{s},\boldsymbol{t}) := \frac{1}{m}\sum_{i=1}^{m}\varphi_{\boldsymbol{\omega}_i}(\boldsymbol{s})\overline{\varphi_{\boldsymbol{\omega}_i}(\boldsymbol{t})} \quad \xrightarrow[m\to+\infty]{\text{a.s.}} \quad K_{\text{mod}}$$

Sketched Supermix (S2Mix)

$$\widehat{\mu}_{\text{sketch}} \in \arg\min_{\mu} \frac{1}{2}\|\boldsymbol{y}_{\text{sketch}} - F_{\text{sketch}}\mu\|_{\mathbb{C}^m}^2 + \kappa\|\mu\|_{\mathrm{TV}} \tag{S2mix}$$

Random measurements m random draws $\boldsymbol{\omega}_{1:m}$

- Sketched operator:

$$F_{\text{sketch}}\mu := \frac{1}{\sqrt{m}}\left(\int_{\mathcal{X}} \varphi_{\boldsymbol{\omega}_i}(\boldsymbol{t})\mathrm{d}\mu(\boldsymbol{t})\right)_{i=1}^{m} \in \mathbb{C}^m =: \mathcal{F}_{\text{sketch}}$$

- Sketch vector = compressed dataset (Gribonval et al. 2020)

$$\boldsymbol{y}_{\text{sketch}} = \frac{1}{\sqrt{m}}\left(\sqrt{\frac{\nu_{\lambda_\tau}}{\Lambda}}\,\mathfrak{F}[\hat{f}_n](\boldsymbol{\omega}_i)\right)_{i=1}^{m}$$

Question How many measurement m to ensure recovery with high $\mathbb{P}_\Lambda$?

Guarantees for sketched Blasso

Informal: De Castro, Gribonval, and Jouvin (2025, Proposition 3.2)

If the sketch size

$$m \gtrsim \texttt{cte}(d, \tau) \cdot s \cdot \log\left(\frac{s|\mathcal{X}|}{\alpha}\right)$$

Then, with proba $1 - \alpha$, the recovery guarantee hold for S2Mix under the kernel switch with $K_{\text{pivot}} = \Psi_\tau$

Guarantees for sketched Blasso

Informal: De Castro, Gribonval, and Jouvin (2025, Proposition 3.2)

If the sketch size

$$m \gtrsim \texttt{cte}(d, \tau) \cdot s \cdot \log\left(\frac{s|\mathcal{X}|}{\alpha}\right)$$

Then, with proba $1-\alpha$, the recovery guarantee hold for S2Mix under the kernel switch with $K_{\text{pivot}} = \Psi_\tau$

+ **effective near region** guarantees hold for radius $r_n = n^{-1/4}\delta_n$, with any $\delta_n \to +\infty$

$$n >> 1, \qquad \widehat{\mu}(\mathcal{F}(r_n)) \lesssim \frac{1}{\delta_n^2} C_{\text{switch}} \|\eta^0\|_{\mathcal{H}_{\text{pivot}}}$$

$$\text{Trade-off} \quad \begin{cases} \text{Radius scale as } r_n = n^{-1/4}\delta_n \\ \text{Recovery rate } \delta_n^{-2} \end{cases}$$

$\rightsquigarrow$ δ_n cannot diverge faster than $n^{1/4}$

Conclusion

Pre-print: ArXiV link

Contributions

- **Kernel switch:** general principle to build certificate without proving LPC
- **Effective near regions** noise adaptive radius r_γ

Perspective

- Beyond TI mixtures $\rightsquigarrow$ GMM $\boldsymbol{x}_k^0 = (\mu_k, \Sigma_k)$ (Giard, De Castro, and Marteau 2025)
- Practical algorithms to compute $\widehat{\mu}$ (Chizat 2019; De Castro, Gadat, and Marteau 2023)

Thank you for your attention !

References

Bredies, Kristian and Hanna Katriina Pikkarainen (2013). **"Inverse problems in spaces of measures".** In: *ESAIM: Control, Optimisation and Calculus of Variations* 19.1, pp. 190–218.

Chizat, Lenaic (2019). **"Sparse optimization on measures with over-parameterized gradient descent".** In: *arXiv preprint arXiv:1907.10300*.

De Castro, Yohann, Sébastien Gadat, and Clément Marteau (2023). **"FastPart: Over-Parameterized Stochastic Gradient Descent for Sparse optimisation on Measures".** In: *arXiv preprint arXiv:2312.05993*.

De Castro, Yohann and Fabrice Gamboa (2012). **"Exact reconstruction using Beurling minimal extrapolation".** In: *Journal of Mathematical Analysis and applications* 395.1, pp. 336–354.

- De Castro, Yohann, Rémi Gribonval, and Nicolas Jouvin (2025). **"Effective regions and kernels in continuous sparse regularisation, with application to sketched mixtures".** In: *arXiv preprint arXiv:2507.08444*.
- De Castro, Yohann et al. (2021). **"SuperMix: Sparse regularization for mixtures".** In: *The Annals of Statistics* 49.3, pp. 1779 –1809.
- Denoyelle, Quentin et al. (2019). **"The sliding frank-wolfe algorithm and its application to super-resolution microscopy".** In: *Inverse Problems*.
- Duval, V. and G. Peyré (2015). **"Exact support recovery for sparse spikes deconvolution".** In: *Foundations of Computational Mathematics*, pp. 1–41.
- Giard, Romane, Yohann De Castro, and Clément Marteau (2025). **"Gaussian Mixture Model with unknown diagonal covariances via continuous sparse regularization".** In: *arXiv preprint arXiv:2509.12889*.
- Gribonval, Rémi et al. (2020). **"Sketching datasets for large-scale learning (long version)".** In: *arXiv preprint arXiv:2008.01839*.

Bibliography iii

- Poon, Clarice, Nicolas Keriven, and Gabriel Peyré (2023). **"The geometry of off-the-grid compressed sensing".** In: *Foundations of Computational Mathematics* 23.1, pp. 241–327.
- Rahimi, Ali and Benjamin Recht (2007). **"Random features for large-scale kernel machines".** In: *Advances in neural information processing systems* 20.
- Zhang, Haizhang and Liang Zhao (2013). **"On the inclusion relation of reproducing kernel Hilbert spaces".** In: *Analysis and Applications* 11.02, p. 1350014.

Building the dual certificate 101

Candidate certificate: For a kernel K satisfying LPC build

$$\eta^0 = \sum_{k=1}^{s} \alpha_k K(\boldsymbol{x}_k^0, \cdot) + \sum_{k=1}^{s} \langle \beta_k, \nabla_1 K(\boldsymbol{x}_k^0, \cdot)\rangle \in \mathcal{H}_K$$

$\rightsquigarrow$ linear system in $(\alpha_k, \beta_k)_{k=1}^s$ for the constraints $\eta^0(\boldsymbol{x}_k^0) = 1$ and $\nabla\eta^0(\boldsymbol{x}_k^0) = 0$

- *Invertibility ?* The system is invertible if spikes are "sufficiently separated"
- *Non-degeneracy* $\Longleftarrow$ K statisfies LPC
- Under LPC we also have $\|\eta^0|_{\mathcal{H}_K} \leq \sqrt{s}$

Different options

1. Prove K_{mod} satisfy LPC + $K = K_{\text{mod}}$ for certificate $\rightsquigarrow$ cumbersome
2. *Kernel switch:* use a pivot kernel $K = K_{\text{pivot}} \longrightarrow \eta^0 \in \mathcal{H}_{K_{\text{pivot}}} \subset \mathcal{H}_{\text{mod}}$

From Bregman to near/far region

We can always bound the Bregman divergence

$$D_{\eta^0}(\widehat{\mu}||\mu^0) := \|\widehat{\mu}\|_{\mathrm{TV}} - \|\mu^0\|_{\mathrm{TV}} - \int \eta^0 \,\mathrm{d}(\widehat{\mu} - \mu^0)$$
$$\leq (\dots) \leq \frac{\left(\gamma + \kappa\|\eta^0\|_{\mathcal{H}_{\mathsf{mod}}}\right)}{2\kappa}$$

Choosing $\kappa = \gamma / \|\eta^0\|_{\mathcal{H}_{\mathsf{mod}}}$ + using the control of η^0 we get for any radius $r \leq r_0$,

$$\mathcal{D}_{\eta^0}(\mu \,||\, \mu^0) = \|\mu\|_{\mathrm{TV}} - \langle \eta^0, \mu \rangle_{\mathcal{C}(\mathcal{X}) \times \mathcal{M}(\mathcal{X})}$$
$$\geq \|\mu\|_{\mathrm{TV}} - \sum_{l=1}^{s} \int_{\mathcal{N}_l(r)} |\eta^0| \,\mathrm{d}|\mu| - \int_{\mathcal{F}(r)} |\eta^0| \,\mathrm{d}|\mu|$$
$$\dots$$

$$\mathcal{D}_{\eta^0}(\mu \,||\, \mu^0) \geq \bar{\varepsilon}_2 r^2 |\mu|(\mathcal{F}(r)) + \bar{\varepsilon}_2 \sum_{l=1}^{s} \int_{\mathcal{N}_l(r)} \mathfrak{d}_{\mathfrak{g}}(x, t_i^0)^2 \,\mathrm{d}|\mu|(x) \,. \qquad (2)$$

Sketch vector = compressed dataset

$$\boldsymbol{y}_{\text{sketch}} = \frac{1}{\sqrt{m}} \left(\sqrt{\frac{\nu_{\lambda_\tau}}{\Lambda}}\, \mathfrak{F}[\hat{f}_n](\boldsymbol{\omega}_i) \right)_{i=1}^{m}$$

Analogies with compressed sensing

$$\boldsymbol{y}_{\text{sketch}} = \frac{1}{n} \sum_{j=1}^{n} \underbrace{\xi(\boldsymbol{z}_j)}_{\in \mathbb{C}^m} \xrightarrow[n \to +\infty]{\text{a.s.}} \int \xi(\boldsymbol{z}) f^0(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z} = F_{\text{sketch}} \mu^0$$

with $\xi(\boldsymbol{z}) \in \mathbb{C}^m$ the (weighted) random Fourier measurements of the dataset

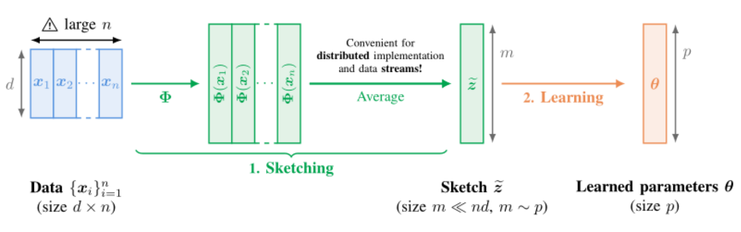


Illustration of sketching as dataset compression from Gribonval et al. (2020)

Model RKHS & functional analysis

$$
\begin{aligned}
K_{\text{mod}}(\boldsymbol{s}, \boldsymbol{t}) &:= \langle F\delta_{\boldsymbol{s}}, F\delta_{\boldsymbol{t}} \rangle_{\mathcal{F}}, \\
&= \langle K_{\text{mod}}(\boldsymbol{s}, \cdot), K_{\text{mod}}(\boldsymbol{t}, \cdot) \rangle_{\mathcal{H}_{\text{mod}}}
\end{aligned}
$$

Adjoint $F^\star : \mathcal{F} \to \mathcal{H}_{\text{mod}}$ and $(F^\star c)(\boldsymbol{t}) = \eta_c(\boldsymbol{t}) := \langle c, F\delta_{\boldsymbol{t}} \rangle_{\mathcal{F}}$

- The unique RKHS of K_{mod} is given by

$$
\mathcal{H}_{\text{mod}} = \left\{ \eta : \mathcal{X} \to \mathbb{R} \;\middle|\; \exists c \in \mathcal{F} \,,\; \eta = \eta_c \right\} = \left\{ \eta : \mathcal{X} \to \mathbb{R} \;\middle|\; \exists ! c \in \overline{\mathrm{Im}(F)} \,,\; \eta = \eta_c \right\},
$$

  and for all c orthogonal to $\overline{\mathrm{Im}(F)}$ in $\mathcal{F}$, $\eta_c = 0$.
- The isometry is given by the mapping $c \in \overline{\mathrm{Im}(F)} \mapsto \eta_c \in \mathcal{H}_{\text{mod}}$.
- The norms satisfy

$$
\forall c \in \overline{\mathrm{Im}(F)} \,, \quad \|\eta_c\|_{\mathcal{H}_{\text{mod}}} = \|c\|_{\mathcal{F}} \,.
$$

Mixture models: the kernel switch constant

the case of *supersmooth* densities

$$\exists p \in [1, +\infty],\ \alpha, \beta > 0, \quad \mathfrak{F}[p](\boldsymbol{\omega}) \propto e^{-\alpha \|\boldsymbol{\omega}\|_p^\beta}$$

leads to a scaling in

$$C_{\text{switch}} = \mathcal{O}_d\left(\tau^{d/2} e^{\alpha\left(\frac{d^{1/p}}{\tau}\right)^\beta}\right),$$

$\mathcal{O}_d$ may depend exponentially in the dimension d but not on τ. This case encompasses

- Gaussian $\rightsquigarrow$ $C_{\text{switch}} = \mathcal{O}_d(\tau^{d/2} e^{d/2\tau^2})$
- multivariate Cauchy
- product of univariate Cauchy
- centered stable distributions with known scale parameter and zero skewness.