# Mixture model and sparse regularization on measure

Nicolas Jouvin (MIA PS)

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

*Statistiques au sommet - Rochebrune 2026*

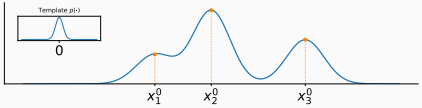

# Mixture models (not this one)

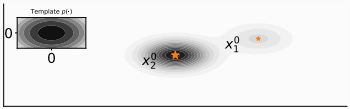
# Mixture models

**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

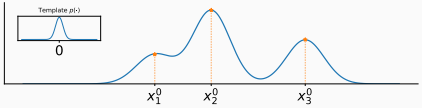
- Usually: MLE, theoretical analysis & algorithm (EM), ...

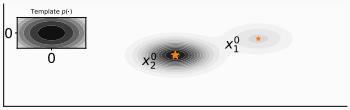
# Mixture models

**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), ...
- **This talk:** linear inverse problem on the space of measure (De Castro et al. 2021)

# Spoiler: lifting the problem to measures

**Encode** Target **sparse measure**[1] (sum of diracs)

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

**Goal** recover **sparse** $\mu^0$ from $n$-sample $Z = \{z_1, \ldots, z_n\} \sim f^0 = \Phi\mu^0$

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

**How ?** Sparse regression on measures $\mathcal{M}(\mathcal{X})$

$$\hat{\mu} \in \arg\min_{\mu} L(Z, \mu) + R(\mu)$$

[1] Bayesians call $\mu^0$ a *mixing distribution*. When it is continuous, $\Phi\mu^0$ is a continuous mixture.

# The Beurling-LASSO (BLASSO)

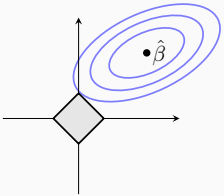
# Reminder on the LASSO...

Linear combination of $p$ covariates

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{\beta}^0 + \boldsymbol{\Gamma},\ \boldsymbol{A} \in \mathbb{R}^{n \times p}$$

**Goal:** recover $s$-sparse $\boldsymbol{\beta}^0$ from noisy $\boldsymbol{y}$
**How ?** solve the LASSO problem

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min} \frac{1}{2} \underbrace{\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\beta}\|^2}_{\text{data-fitting}} + \underbrace{\kappa \|\boldsymbol{\beta}\|_1}_{\text{sparsity promoting}}$$




Guarantees under conditions[2] on the measurement $\boldsymbol{A}$

- recovery results $\operatorname{Supp} \hat{\boldsymbol{\beta}} = \operatorname{Supp} \boldsymbol{\beta}^0$
- control on $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|$

[2] irrepresentability, RIP, REV, etc.

# .. and its extension to measures (BLASSO)

Linear measurements of a measure through

$$F : \begin{array}{ccc} \mathcal{M}(\mathcal{X}) & \to & \text{Hilbert } \mathcal{F} \\ \mu & \mapsto & F\mu \end{array}, \qquad \text{Observe } y = F\mu_0 + \Gamma \in \mathcal{F}$$

Goal: recover $s$-sparse $\mu^0$ from noisy $y$ by solving

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

$$\widehat{\mu} \in \operatorname*{arg\,min}_{\mu \in \mathcal{M}(\mathcal{X})} J_\kappa(\mu) \coloneqq \frac{1}{2}\left\|y - F\mu\right\|_{\mathcal{F}}^2 + \kappa \|\mu\|_{\mathrm{TV}} \qquad \text{(BLASSO)}$$

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**

$$\begin{array}{lccc} \text{(Discrete)} & \mu = \displaystyle\sum_{k=1}^{p} a_k \delta_{\boldsymbol{x}_k} & \longrightarrow & \|\mu\|_{\mathrm{TV}} = \displaystyle\sum_{k=1}^{p} |a_k| = \|a\|_1 \\ \text{(Continuous)} & \mathrm{d}\mu = f\,\mathrm{d}\lambda & \longrightarrow & \|\mu\|_{\mathrm{TV}} = \|f\|_{L^1} \end{array}$$

# Theoretical analysis for the BLASSO

$$\widehat{\mu} \in \operatorname*{arg\,min}_{\mu \in \mathcal{M}(\mathcal{X})} J_\kappa(\mu) \coloneqq \frac{1}{2}\left\|y - F\mu\right\|_{\mathcal{F}}^2 + \kappa \|\mu\|_{\mathrm{TV}} \tag{BLASSO}$$

- Convex problem over the space of measures $\mathcal{M}(\mathcal{X})$
- How well does $\widehat{\mu}$ recovers $\mu^0$ ? Dependency on noise $\gamma = \|\Gamma\|_{\mathcal{F}}$
- Requires technical conditions on

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

$$K_{\mathrm{mod}}(\boldsymbol{s}, \boldsymbol{t}) \coloneqq \langle F\delta_{\boldsymbol{s}}, F\delta_{\boldsymbol{t}} \rangle_{\mathcal{F}}$$

with associated RKHS $\mathcal{H}_{\mathrm{mod}}$ which contains continuous functions

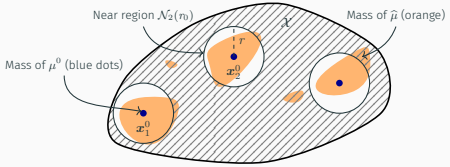
# Near and far regions

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

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

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

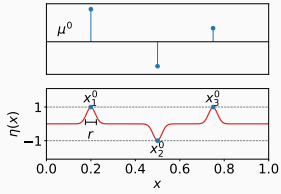
$$\text{Near:} \quad \mathcal{N}_k(r) \coloneqq \{\boldsymbol{x}, \mathfrak{d}(\boldsymbol{x}, \boldsymbol{x}_k^0) \leq r\}$$
$$\text{Far:} \quad \mathcal{F}(r) \coloneqq \mathcal{X} \setminus \cup_k \mathcal{N}_k(r)$$




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

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

1. $\eta^0 \in \mathcal{H}_{\text{mod}}$
2. $\eta^0(\boldsymbol{x}_k^0) = \pm 1 \rightsquigarrow \eta^0$ in $\partial \|\mu^0\|_{\text{TV}}$
3. + additional controls on near/far



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* $\sim a_k^0$*:*

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

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

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

# Proving dual certificate is hard

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

**Local positive curvature (LPC) assumption**

If model kernel $K_{\text{mod}}$ satisfies

[...Technical condition on $K_{\text{mod}}$...]

Then there exists a $(\bar{\epsilon}_0, \bar{\epsilon}_2, r_0)$-non-degenerate certificate.

The distance $\mathfrak{d}$ is the *Fisher-Rao* metric associated to $K_{\text{mod}}$.

# Proving dual certificate is hard

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

**Local positive curvature (LPC) assumption**

If model kernel $K_{\text{mod}}$ satisfies

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

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

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

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

$\Delta_0 < +\infty$, 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 $(\bar{\varepsilon}_0, \bar{\varepsilon}_2, r_0)$-non-degenerate certificate for $\Delta_0$-separated target $\mu^0$.
With $\mathfrak{d}_{\mathfrak{g}}$ the Fisher-Rao metric associated to $K_{\text{mod}}$.

# Proving dual certificate is hard

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

**Local positive curvature (LPC) assumption**

If model kernel $K_{\text{mod}}$ satisfies

[...Technical condition on $K_{\text{mod}}$...]

Then there exists a $(\bar{\epsilon}_0, \bar{\epsilon}_2, r_0)$-non-degenerate certificate.

The distance $\mathfrak{d}$ is the *Fisher-Rao* metric associated to $K_{\text{mod}}$.

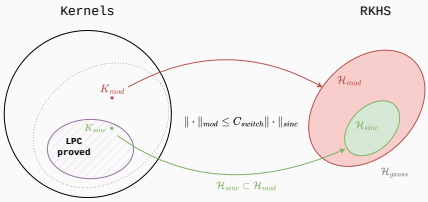
Our contribution (De Castro, Gribonval, and Jouvin 2025)

1. New measurement operator $F$ ? You need to prove LPC for your new $K_{\text{mod}}$...
   $\rightsquigarrow$ ... 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 $r_\gamma$ adaptive to noise

# The "kernel switch" principle in a nutshell

- Theory says $\eta^0$ must $\in \mathcal{H}_{\text{mod}}$
- **Switch** LPC kernel $K_{\text{pivot}}$ with RKHS $\mathcal{H}_{\text{pivot}}$
- **We need**
  1. Inclusion: $\mathcal{H}_{\text{pivot}} \subset \mathcal{H}_{\text{mod}}$
  2. Control $\|\eta\|_{\mathcal{H}_{\text{mod}}} \leq C_{\text{switch}} \|\eta\|_{\mathcal{H}_{\text{pivot}}}$

**The price you "pay" is** $C_{\text{switch}}$




Explicit $C_{\text{switch}}$ for translation-invariant $K \rightsquigarrow$ spectral representation $\nu(\boldsymbol{\omega})$

$$K(\boldsymbol{s},\boldsymbol{t}) = \rho(\boldsymbol{s}-\boldsymbol{t}) \overset{\text{(Bochner)}}{=} \int e^{+\imath\boldsymbol{\omega}^\top(\boldsymbol{s}-\boldsymbol{t})}\nu(\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega} \quad \longrightarrow C_{\text{switch}}(K_1,K_2) = \inf_{\boldsymbol{\omega}} \sqrt{\frac{\nu_2}{\nu_1}(\boldsymbol{\omega})}$$

## But, you promessed to talk about mixtures ?

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

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

# 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 $\mathcal{F}$

**For mixtures:** what is $y$, $F$, $\mathcal{F}$ ?

Inverse problem

# 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

$$\mu^0 = \sum_{k=1}^{s} a_k^0 \delta_{\boldsymbol{x}_k^0} \qquad \xrightarrow{\text{convolution}} \qquad f^0 = \Phi\mu^0 = p \star \mu^0$$

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

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

# 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

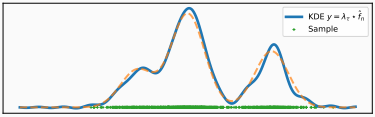
$$\mu^0 = \sum_{k=1}^{s} a_k^0 \delta_{\boldsymbol{x}_k^0} \quad \xrightarrow{\text{convolution}} \quad f^0 = \Phi\mu^0 = p \star \mu^0$$

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

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

**Embed into the same RKHS** Smoothing kernel $\lambda_\tau(\cdot) = \tau^{-d}\lambda(\cdot/\tau)$

- RKHS embedding $L_\tau : \nu \mapsto \lambda_\tau \star \nu \in \mathcal{F}_\tau$
- Associated Hilbert $\mathcal{F}_\tau$ depends on $\lambda_\tau$
- Observation is the KDE
  $y = L_\tau \hat{f}_n = \lambda_\tau \star \hat{f}_n$




KDE using sinus cardinal $\lambda(x) = \sin(x)/x$

# Supermix (cont'd)

Solve the following BLASSO problem

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

**Model kernel** is explicit & translation-invariant $K_{\mathsf{mod}} = \rho_{\mathsf{mod}} = \lambda_\tau \star p \star \check{p}$

**However**

1. *Theory:* LPC needs to be checked for each template $p \rightsquigarrow$ **switch!**

$$C_{\mathsf{switch}} = C_{\mathsf{switch}}(\tau, p) \coloneqq \sqrt{\frac{\tau^d}{\inf_{\mathbb{B}_\tau} \nu_{\lambda_\tau} \left| \nu_p(\boldsymbol{\omega}) \right|^2}} < +\infty, \tag{$\mathbf{H}_p$}$$

# Supermix (cont'd)

Solve the following BLASSO problem

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

**Model kernel** is explicit & translation-invariant $K_{\mathsf{mod}} = \rho_{\mathsf{mod}} = \lambda_\tau \star p \star \check{p}$

**However**

1. *Theory:* LPC needs to be checked for each template $p \rightsquigarrow$ **switch!**

$$C_{\mathsf{switch}} = C_{\mathsf{switch}}(\tau, p) \coloneqq \sqrt{\frac{\tau^d}{\inf_{\mathbb{B}_\tau} \nu_{\lambda_\tau} \left|\nu_p(\boldsymbol{\omega})\right|^2}} < +\infty, \qquad (\mathbf{H}_p)$$

2. *Practice:* Every algorithm computing $\widehat{\mu}$ requires evaluation of $K_{\mathsf{mod}}$ (and $\nabla K_{\mathsf{mod}}$)

Costly $d$-dimensional integrals $K_{\mathsf{mod}}(\boldsymbol{s} - \boldsymbol{t}) = \int_{\mathbb{R}^d} (\dots)$

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

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

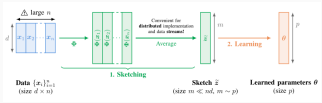
$$K_{\text{mod}}(\boldsymbol{s}, \boldsymbol{t}) = \mathbb{E}_{\boldsymbol{\omega} \sim \Lambda}\left[\varphi_{\boldsymbol{\omega}}(\boldsymbol{s})\overline{\varphi_{\boldsymbol{\omega}}(\boldsymbol{t})}\right], \qquad \varphi_{\boldsymbol{\omega}}(\boldsymbol{t}) \coloneqq W(\boldsymbol{\omega}) e^{-\imath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

- **Sketching** $\rightsquigarrow$ Monte-Carlo approximation draw $\boldsymbol{\omega}_{1:m} \sim \Lambda$
- *Sketched model kernel*

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

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

$$\boldsymbol{y}_{\text{sketch}} = \frac{1}{\sqrt{m}} \left( \sqrt{\frac{\nu_{\lambda_\tau}}{\Lambda}}\, \mathfrak{F}[\hat{f}_n](\boldsymbol{\omega}_i) \right)_{i=1}^{m} = \frac{1}{n} \sum_{j=1}^{n} \underbrace{\xi(\boldsymbol{z}_j)}_{\in \mathbb{C}^m}$$




Sketching as dataset compression from Gribonval et al. (2020)

# S2Mix : sketching the Supermix problem (De Castro, Gribonval, and Jouvin 2025)

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

**Random measurements** $m$ random draws $\boldsymbol{\omega}_{1:m}$ with sketched operator:

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

# S2Mix : sketching the Supermix problem (De Castro, Gribonval, and Jouvin 2025)

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

**Random measurements** $m$ random draws $\boldsymbol{\omega}_{1:m}$ with sketched operator:

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

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

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

If the sketch size

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

Then, with proba $1-\alpha$, the recovery guarantee hold for S2Mix with kernel switch

# Effective near regions: beyond fixed radius

**What about noise ?** We can show $\gamma_n = \|y - F\mu^0\|_{\mathcal{F}} \leq \frac{1}{\sqrt{n}}$

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

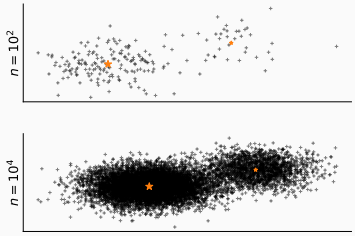
$$n >> 1, \qquad \widehat{\mu}(\mathcal{F}(r_n)) \lesssim \frac{1}{\delta_n^2} C_{\text{switch}} \|\eta^0\|_{\mathcal{H}_{\text{pivot}}}$$

# Effective near regions: beyond fixed radius

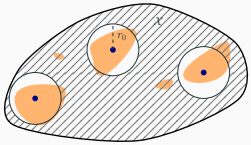
**What about noise ?** We can show $\gamma_n = \|y - F\mu^0\|_{\mathcal{F}} \leq \frac{1}{\sqrt{n}}$

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

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




Same radius $r_0$

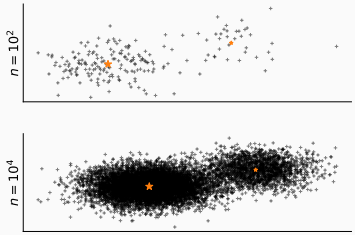


# Effective near regions: beyond fixed radius

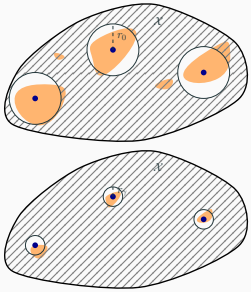
**What about noise ?** We can show $\gamma_n = \|y - F\mu^0\|_{\mathcal{F}} \leq \frac{1}{\sqrt{n}}$

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

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




Noise dependent $r_\gamma$



# Conclusion

Pre-print: ArXiV link

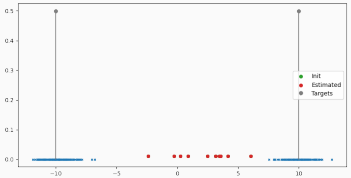
## Contributions

- **Kernel switch:** general principle to build certificate without proving LPC
- **Effective near regions** noise adaptive radius $r_\gamma$

## Perspective

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

Gradient descent for $\min_{\text{discrete } \mu_p} J_\kappa(\mu_p) \xrightarrow[p\to+\infty]{\text{approximate Wasserstein flow}} \frac{\partial \mu_t}{\partial t} = -\text{div}\left(\mu_t \nabla^W J_\kappa(\mu_t)\right)$



Thank you for your attention !

# References


- 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*.

# Bibliography ii


- De Castro, Yohann et al. (2021). **“SuperMix: Sparse regularization for mixtures”.** In: *The Annals of Statistics* 49.3, pp. 1779 –1809.
- 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*.
- 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.

# Why is it called a *dual* certificate ? Duality

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

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

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

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

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

We have

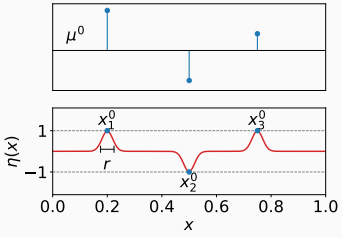
- $\operatorname{Supp} \widehat{\mu} \subset \{\boldsymbol{x} \in \mathcal{X},\ |\eta(\boldsymbol{x})| = 1\} \longrightarrow \eta$ *certifies* the support of $\widehat{\mu}$
- Natural to look for certificates $\eta^0 \in \operatorname{Im}(F^\star)$ s.t. $\eta^0 \in \partial \|\mu^0\|_{\mathrm{TV}}$

# Non-degeneracy condition for the dual certificate

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

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

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

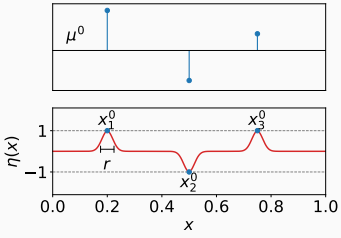


# Non-degeneracy condition for the dual certificate

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

Fix radius $r > 0$, distance $\mathfrak{d}(\cdot, \cdot)$ & control $\epsilon_0, \epsilon_2 > 0$.
$\eta^0$ 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

# 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 $K_{\text{mod}}$ 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

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

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

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

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

# Sketching

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

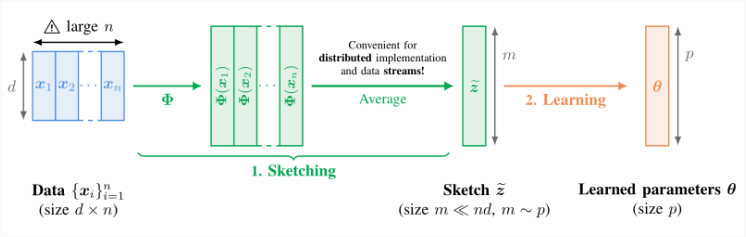



Illustration of sketching as dataset compression from Gribonval et al. (2020)

# Model RKHS & functional analysis

$$K_{\text{mod}}(\boldsymbol{s}, \boldsymbol{t}) \coloneqq \langle F\delta_{\boldsymbol{s}}, F\delta_{\boldsymbol{t}} \rangle_{\mathcal{F}},$$
$$= \langle K_{\text{mod}}(\boldsymbol{s}, \cdot), K_{\text{mod}}(\boldsymbol{t}, \cdot) \rangle_{\mathcal{H}_{\text{mod}}}$$

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

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

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

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

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

# 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

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

$\mathcal{O}_d$ may depend exponentially in the dimension $d$ but not on $\tau$. 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.