Estimation of Gaussian Mixture Models via the BLASSO

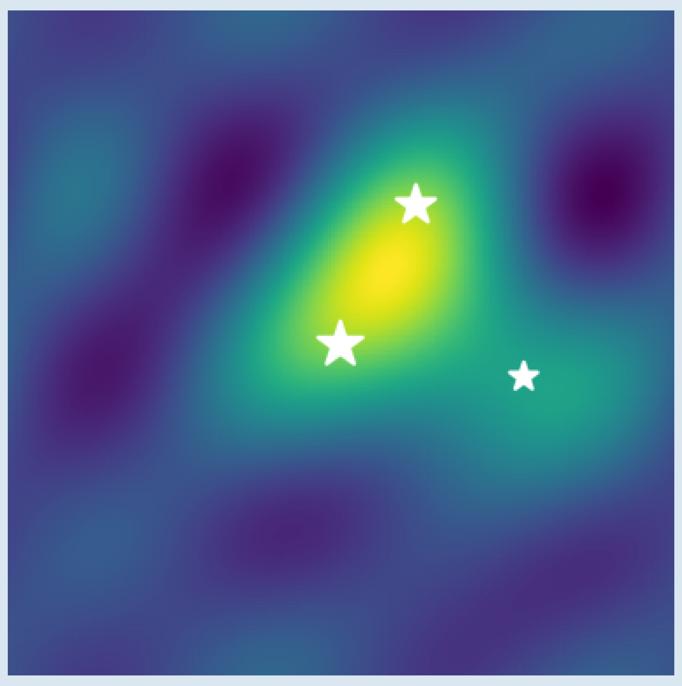
Romane Giard

Joint work with Yohann De Castro & Clément Marteau

Institut Camille Jordan, Centrale Lyon

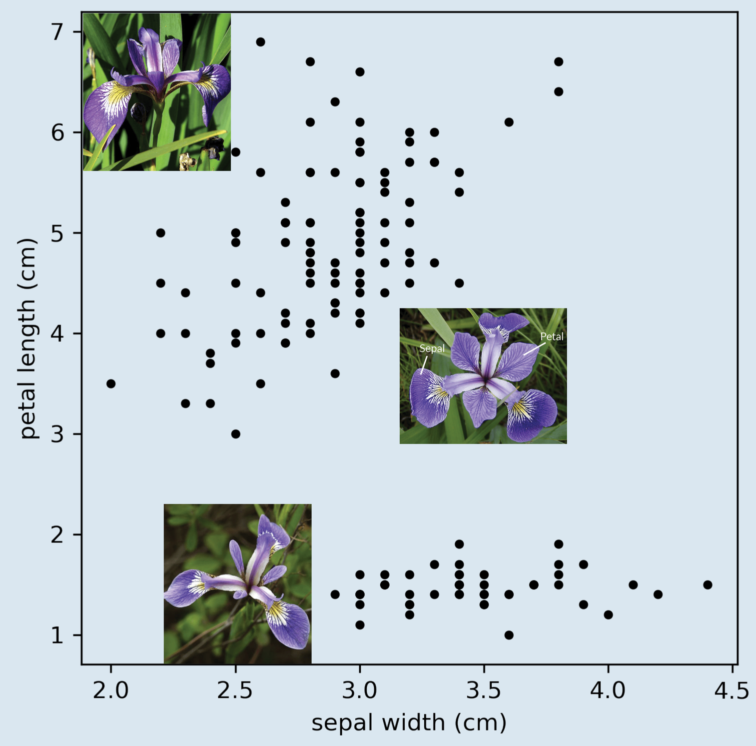
Inverse problem

Recover a discrete object from obs.

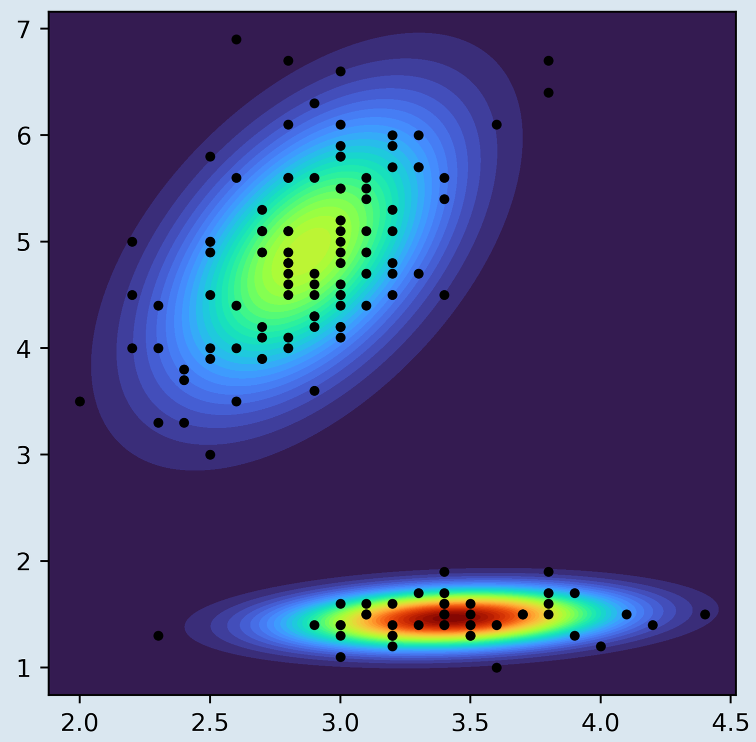


Statistical estimation

$$X_i \in \mathbb{R}^d \overset{i.i.d.}{\sim} f^0$$



Gaussian mixture distribution



Assumption: f^0 is the density of a Gaussian mixture distribution.

$$f^0 = \sum_{j=1}^{s} a_j^0 \varphi_{(t_j^0, U_j^0)}$$

where $a_j^0 > 0$ and $\sum_{j=1}^s a_j^0 = 1$.

$\varphi_{(t,U)}$: density of a d-dimensional Gaussian distribution, with mean $t \in \mathbb{R}^d$ and square root of covariance $U \in \mathbb{S}_{++}^d$.

Sample from f^0: $j \in \{1, \ldots, s\}$ is chosen with prob. a_j^0; then $X_i | j \sim \varphi_{(t_j^0, U_j^0)}$.

Goal: Estimation of a Gaussian mixture model

Goal: Recover the parameters of the mixture from the data $X_1, \ldots, X_n$.
Parameters: Number of components p (unknown), $(a_j)_{j=1}^p$, $(t_j, U_j)_{j=1}^p$.

	Convex	Do not require knowledge of p
Log-likelihood maximization (EM algo.)	✗	✗
BLASSO	✓	✓

The Beurling-LASSO (BLASSO)

Main idea: Embed the Gaussian mixture parameters into a measure:

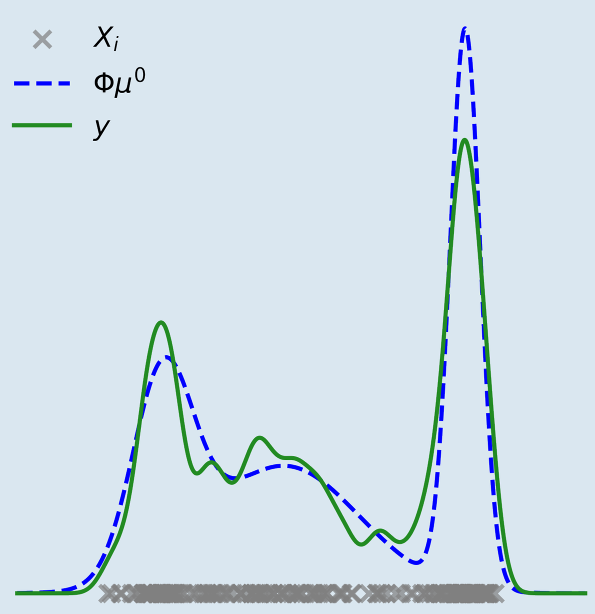
$$(a_j, t_j, U_j)_{j=1}^p \mapsto \mu = \sum_{j=1}^{p} a_j \delta_{(t_j, U_j)} .$$

The goal is to recover the target measure $\mu^0 = \sum_{j=1}^s a_j^0 \delta_{(t_j^0, U_j^0)}$.

Loss function: Combines a *data fidelity term* and a *regularization term*.

Data fidelity term: Ensures the estimate fits the observations.

- Rewrite $f^0 = \Phi\mu^0$ with Φ a linear map from measures to functions.
- Construct an approximation y of f^0 from the observations $X_1, \ldots, X_n$ (convolution with Gaussian kernel).
- For any candidate measure μ, evaluate $\|\Psi\mu - y\|_{\mathbb{L}}^2$ ($\mathbb{L}$: Hilbert space).



Ψ: smoothing of Φ

Regularization term: Promotes sparsity.
Total variation norm: for a discrete measure $\mu = \sum_{j=1}^p a_j \delta_{(t_j, U_j)}$,

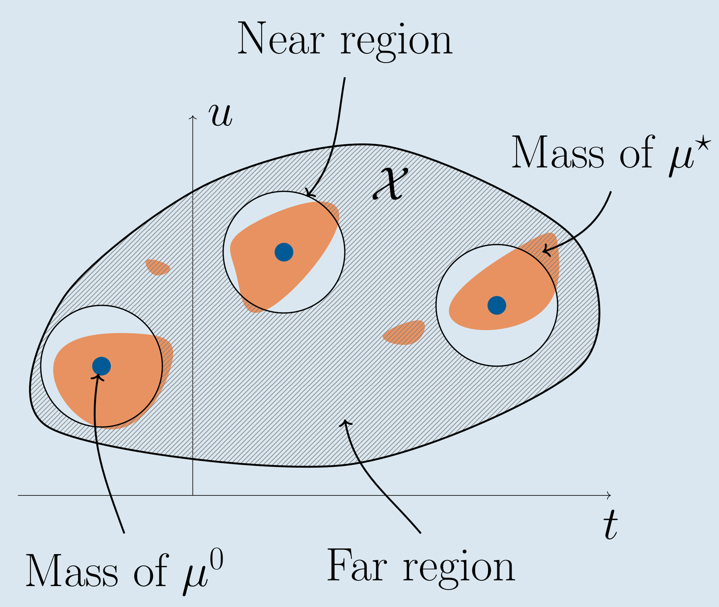
$$\|\mu\|_{\mathrm{TV}} = \sum_{j=1}^{p} |a_j| .$$

The BLASSO: Regularization parameter $\kappa > 0$.

$$\operatorname*{arg\,min}_{\mu \in \mathcal{M}(\mathcal{X})^+} J(\mu) \quad \text{with} \quad J(\mu) \coloneqq \frac{1}{2}\|\Psi\mu - y\|_{\mathbb{L}}^2 + \kappa \|\mu\|_{\mathrm{TV}}$$

Convex problem!

Statistical guarantees



Assumptions:

- Diagonal covariance matrices;
- Component separation;
- Suitable choices of κ and y.

Result: The expected discrepancy in mass assigned to regions of the parameter space between μ_ω^0 and a BLASSO solution $\mu^\star$ scales as $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$ (n: sample size).

We estimate μ_ω^0 rather than μ^0 (technical).

Algorithm: CPGD

Goal: Approximate numerically the solution to the BLASSO.

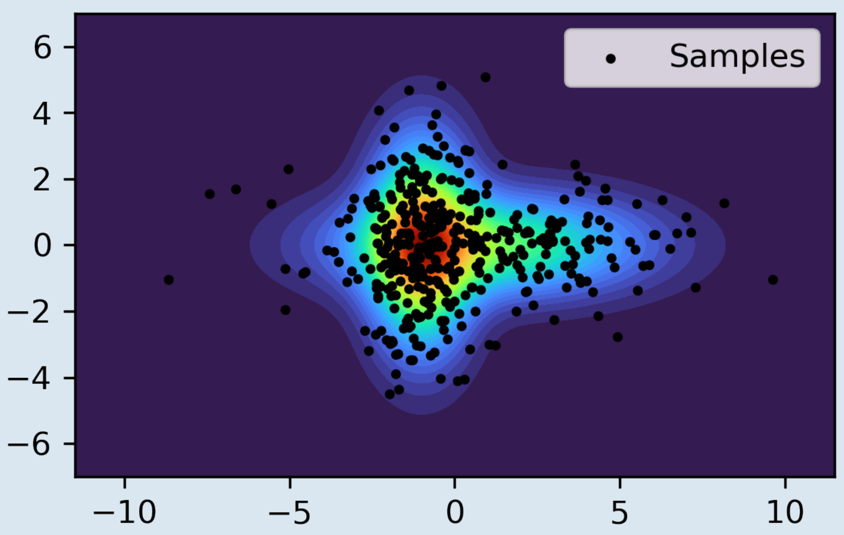
Limitation: Must discretize $\to$ J is no longer convex on the space of discrete measures.

Algorithm: Relies on *Conic retraction* and *Natural gradient descent.*
Notations:
- J'_μ: Fréchet derivative of J at point μ.
- $\nabla^{\mathfrak{g}}$: Riemannian gradient w.r.t. Fisher-Rao geometry.
  $\nabla^{\mathfrak{g}} J'_\mu(t,U) = \mathfrak{g}_{(t,U)}^{-1} \nabla J'_\mu(t,U)$ with $\mathfrak{g}$ the metric tensor.

For diagonal covariance matrices

Require: Step sizes $\alpha, \beta \geq 0$, number of iterates K, initialization $\mu_\omega^1 = \sum_{j=1}^p \omega_j^1 \delta_{(t_j^1, U_j^1)}$
  for $k = 1, \ldots, K$ **do**
    for $j = 1, \ldots, p$ **do**
      $\omega_j^{k+1} \leftarrow \omega_j^k e^{-\alpha J'_{\mu_\omega^k}(t_j^k, U_j^k)}$ ▷ Weight update
      $(t_j^{k+1}, U_j^{k+1}) \leftarrow (t_j^k, U_j^k) - \beta \nabla^{\mathfrak{g}} J'_{\mu_\omega^k}(t_j^k, U_j^k)$ ▷ Location-scale update
    end for
    $\mu_\omega^{k+1} \leftarrow \sum_{j=1}^p \omega_j^{k+1} \delta_{(t_j^{k+1}, U_j^{k+1})}$
  end for

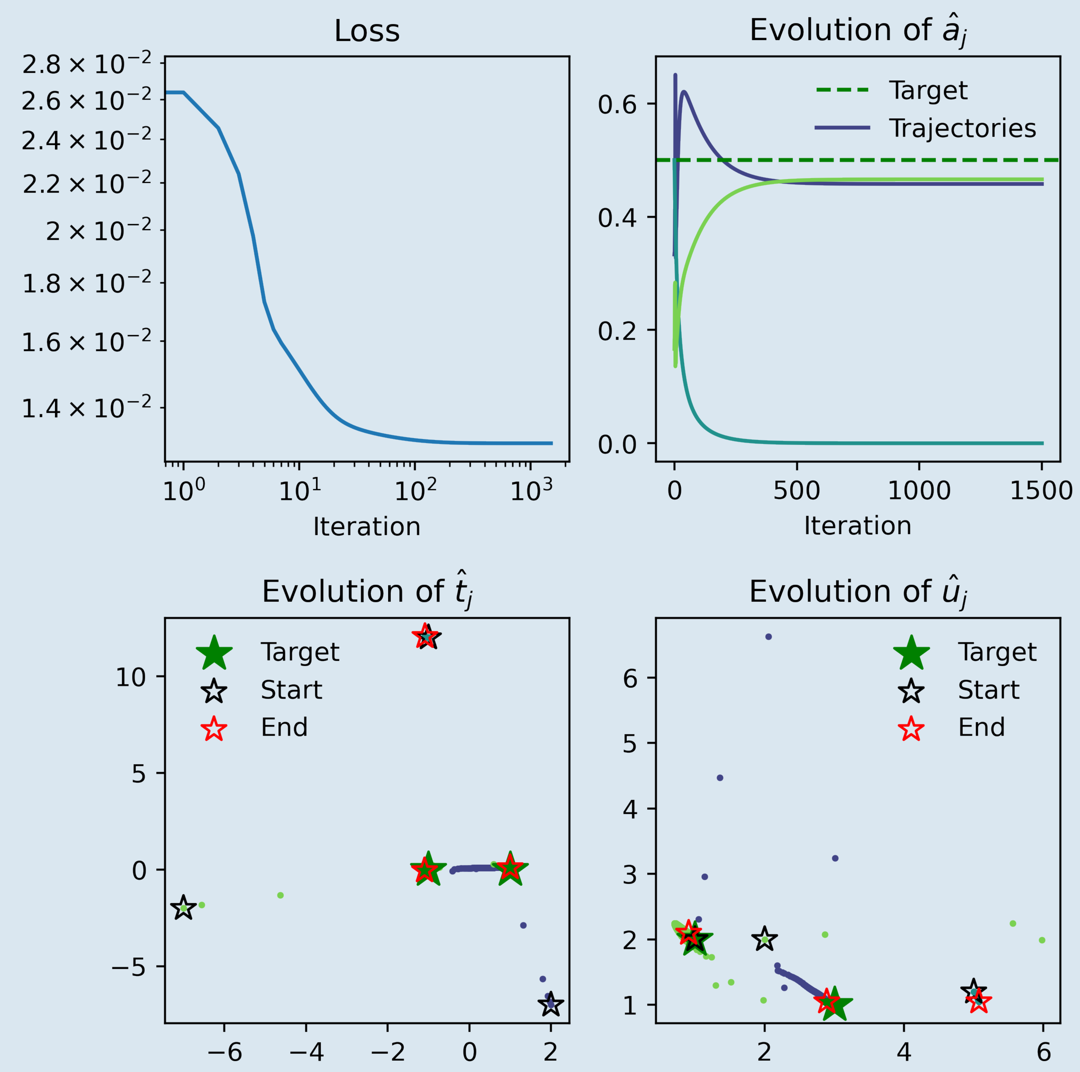
Numerical experiment



2D Target Gaussian mixture

Setting: Gaussian mixture in 2 dimensions, diagonal covariance matrices.
2 target components.
We have access to $n = 400$ samples of the target mixture.

Algo: Initialization with 3 particles.



References

- Poon et al. (2023), "The geometry of off-the-grid compressed sensing".
- De Castro et al. (2021), "SuperMix: Sparse Regularization for Mixtures".
- Chizat (2022), "Sparse optimization on measures with over-parameterized gradient descent".
- Nielsen (2020), "An Elementary Introduction to Information Geometry".
- **Associated paper:** Giard et al. (2025), *Gaussian Mixture Model with unknown diagonal covariances via continuous sparse regularization*. Submitted. arXiv:2509.12889