Prediction and testing of mixtures of features issued from a continuous dictionary

Clément Hardy
joint work with C. Butucea (CREST, ENSAE, IP Paris), J.-F. Delmas (ENPC), A. Dutfoy (EDF)

Luxembourg, August 31, 2023

Industrial motivation: spectroscopy

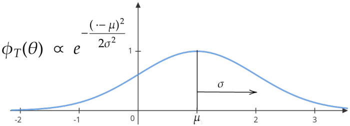
Wave numbers (cm-1)	Peak assignment
3690-3400-3364-3200-3014	-OH
2952-2920-2850	$\nu - CH_2, CH_3$ Aliphatic
1731	$\nu - C = O$
1647	$\nu - C = C$ de $HC = CH_2$
1540	$\nu - C = C$ de R-CR=CH-R, δ CH2 Aliphatic
1419	δCH_2, δ-CH Aliphatic
1160-1082	ν Si-O (SiO_2)
1009-909	ν Si-O (Si-OH)
825	C-Cl
664	CH Aromatic

Location of absorption spikes of chemical components for polychloroprene samples ([Tchalla, 2017]).

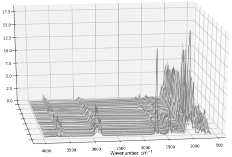
Continuous dictionary

$$\theta = (\mu, \sigma) \in \Theta = \mathbb{R} \times \mathbb{R}_+^*$$

$$\phi_T(\theta) \propto e^{-\frac{(\cdot - \mu)^2}{2\sigma^2}}$$



Infrared signals



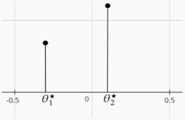
The model

A noisy signal $y = (y(t),\ t \in \mathbb{R})$ is observed on the grid $t_1, \cdots, t_T$:

$$\mathbf{y} = \underbrace{\sum_{k=1}^{s} \beta_k^\star \, \phi_T(\theta_k^\star)}_{\text{mixture of spikes}} + \underbrace{w}_{\text{noise}} .$$

Goal: recover from y the parameters $\beta^\star$ and $\vartheta^\star = (\theta_1^\star, \cdots, \theta_s^\star)$.

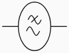
Motivation : low-pass filter



Sources ponctuelles

$$\mu^\star = \sum_{k=1}^{s} \beta_k^\star \, \delta_{\theta_k^\star}$$

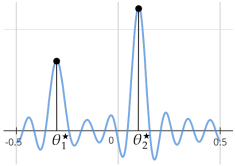
$\ast$



Filtre passe-bas (Dirichlet)

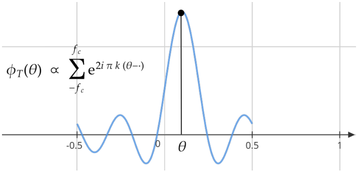
$$t \mapsto \sum_{-f_c}^{f_c} e^{2i\pi k t} = \frac{\sin(\pi\,(2f_c+1)\,t)}{\sin(\pi\,t)}$$

Filtered signal



Continuous dictionary

$\theta \in \Theta = \mathbb{R}/\mathbb{Z}$ and $T = 2f_c + 1$



The model

A noisy signal $y = (y(t),\ t \in \mathbb{R}/\mathbb{Z})$ is observed:

$$\mathbf{y} = \underbrace{\sum_{k=1}^{s} \beta_k^\star \, \phi_T(\theta_k^\star)}_{\text{mixture of spikes}} + \underbrace{w}_{\text{noise}} .$$

Goal: recover from y the parameters $\beta^\star$ and $\vartheta^\star = (\theta_1^\star, \cdots, \theta_s^\star)$.

Plan of the talk

- **I The model**
- **II Estimation and Prediction**
- **III Tests**

I The model

The model

We observe a random element y of a Hilbert space H_T (e.g.: $\mathbb{R}^T$, $L^2(\lambda_T)$,...) with scalar product $\langle \cdot, \cdot \rangle_T$ (and norm $\|\cdot\|_T$).

Model

$$\mathbf{y} = \underbrace{\sum_{k=1}^{s} \beta_k^\star \, \phi_T(\theta_k^\star)}_{\text{signal}} \quad + \quad \underbrace{w_T}_{noise}.$$

Notations

- T increases with the amount of information of the observation (number of grid points, 1/noise level...).
- $\theta_k^\star \in \Theta \subset \mathbb{R}$ and $\beta_k^\star \in \mathbb{R}$, for all k.
- Continuous dictionary $(\phi_T(\theta), \theta \in \Theta)$ of elements of H_T of norm 1. The map ϕ_T is continuous on Θ.
- w_T Gaussian process.

The model: Gaussian noise (I)

Assumptions on the noise (H1)

For all $f \in H_T$, the random variable $\langle f, w_T \rangle_T$ is centered Gaussian with:

$$\mathrm{Var}\left(\langle f, w_T \rangle_T\right) \leq \Delta_T \left\| f \right\|_T^2.$$

The model: Gaussian noise (I)

Assumptions on the noise (H1)

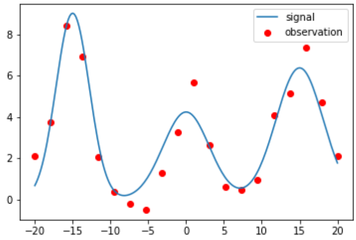
For all $f \in H_T$, the random variable $\langle f, w_T \rangle_T$ is centered Gaussian with:

$$\mathrm{Var}\left(\langle f, w_T \rangle_T\right) \leq \Delta_T \left\| f \right\|_T^2.$$

Ex: spectroscopy

- **Regular grid:** $t_1 < \cdots < t_T$ on $\mathbb{R}$ with step-size $\Delta_T = \frac{t_T - t_1}{T}$.
- **Observations:** $y(t_i) = \mathit{signal}(t_i) + w_T(t_i)$, $1 \leq i \leq T$.
- **Noise:** $w_T(t_i)$ i.i.d $\sim \mathcal{N}(0, 1)$.

$H_T = L^2(\lambda_T)$ where $\lambda_T(\mathrm{d}t) = \Delta_T \sum_{j=1}^{T} \delta_{t_j}(\mathrm{d}t)$.



Δ_T = step-size.

The model: Gaussian noise (II)

Assumptions on the noise (H1)

For all $f \in H_T$, the random variable $\langle f, w_T \rangle_T$ is centered Gaussian with:

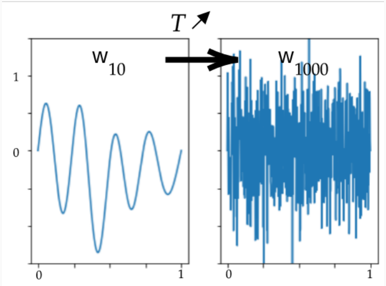
$$\mathrm{Var}\left(\langle f, w_T \rangle_T\right) \leq \Delta_T \, \|f\|_T^2.$$

Ex: low-pass filter

- **Observations:** $(y(t),\ t \in \mathbb{R}/\mathbb{Z})$ s.t $y \in L^2(\mathrm{Leb})$.
- **Truncated white noise:** $w_T = \frac{1}{\sqrt{T}} \sum_{k=1}^{T} G_k \, \psi_k$,

  -G_k i.i.d $\sim \mathcal{N}(0,1)$,
  -$(\psi_k, k \in \mathbb{N})$ o.n.b. of $L^2(\mathrm{Leb})$.

Thus $\|w_T\|_{L^2(\mathrm{Leb})}$ is of order 1.



$\Delta_T = 1/T$.

II Estimation and Prediction

Estimators

$$(\hat{\beta}, \hat{\vartheta}) \in \underset{\beta \in \mathbb{R}^K, \vartheta \in \Theta_T^K}{\operatorname{argmin}} \quad \frac{1}{2} \|y - \beta\, \Phi_T(\vartheta)\|_T^2 + \kappa \|\beta\|_{\ell_1}. \qquad (\mathcal{P}_1(\kappa))$$

- K is an upper bound for s.
- Θ_T is a compact interval.
- $\kappa > 0$ tuning parameter.

$\Phi_T(\vartheta) \in H_T^K$ is defined by:

$$\Phi_T(\vartheta) = (\phi_T(\theta_1), \ldots, \phi_T(\theta_K))^\top.$$

Estimators

Estimators

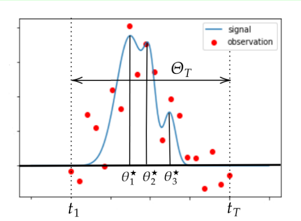
$$(\hat{\beta}, \hat{\vartheta}) \in \underset{\beta \in \mathbb{R}^K, \vartheta \in \Theta_T^K}{\operatorname{argmin}} \ \frac{1}{2} \|y - \beta \, \Phi_T(\vartheta)\|_T^2 + \kappa \|\beta\|_{\ell_1}. \qquad (\mathcal{P}_1(\kappa))$$

- K is an upper bound for s.
- Θ_T is a compact interval.
- $\kappa > 0$ tuning parameter.

$\Phi_T(\vartheta) \in H_T^K$ is defined by:

$$\Phi_T(\vartheta) = (\phi_T(\theta_1), \ldots, \phi_T(\theta_K))^\top.$$

Ex: spectroscopy



Prediction risk

Estimators

Let $(\hat{\beta}, \hat{\vartheta}) \in \mathbb{R}^K \times \Theta^K$ be measurable functions of y solutions of $(\mathcal{P}_1(\kappa))$ "approximating" $(\beta^\star = (\beta_1^\star, \cdots, \beta_s^\star), \vartheta^\star = (\theta_1^\star, \cdots, \theta_s^\star))$. We define:

Prediction risk

Estimators

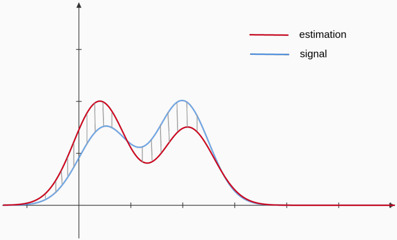
Let $(\hat{\beta}, \hat{\vartheta}) \in \mathbb{R}^K \times \Theta^K$ be measurable functions of y solutions of $(\mathcal{P}_1(\kappa))$ "approximating" $(\beta^\star = (\beta_1^\star, \cdots, \beta_s^\star), \vartheta^\star = (\theta_1^\star, \cdots, \theta_s^\star))$. We define:

Prediction risk

$$\left\| \beta^\star \Phi_T(\vartheta^\star) - \hat{\beta}\, \Phi_T(\hat{\vartheta}) \right\|_T,$$

where $\Phi_T(\hat{\vartheta}) \in H_T^K$ is defined by:

$$\Phi_T(\hat{\vartheta}) = (\phi_T(\hat{\theta}_1), \ldots, \phi_T(\hat{\theta}_K))^\top,$$



Estimators

Let $(\hat{\beta}, \hat{\vartheta}) \in \mathbb{R}^K \times \Theta^K$ be measurable functions of y solutions of $(\mathcal{P}_1(\kappa))$ "approximating" $(\beta^\star = (\beta_1^\star, \cdots, \beta_s^\star), \vartheta^\star = (\theta_1^\star, \cdots, \theta_s^\star))$. We define:

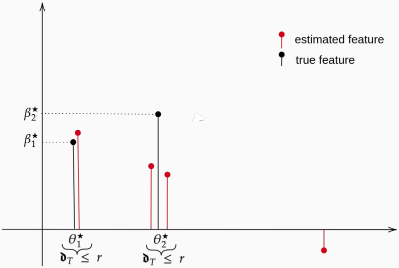
Estimation risks (I)

$$\sum_{k=1}^{s} \left| \beta_k^\star - \sum_{\ell \in S_k(r)} \hat{\beta}_\ell \right| \quad \text{and} \quad \sum_{\ell \in S(r)^c} |\hat{\beta}_\ell|,$$

where the set $S(r)$ is given by:

$$S(r) = \bigcup_{1 \le k \le s} S_k(r),$$

with
$S_k(r) = \left\{ \ell,\ \hat{\beta}_\ell \neq 0 \text{ and } \mathfrak{d}_T(\hat{\theta}_\ell, \theta_k^\star) \le r \right\}.$



Estimation: estimation risks (II)

Estimators

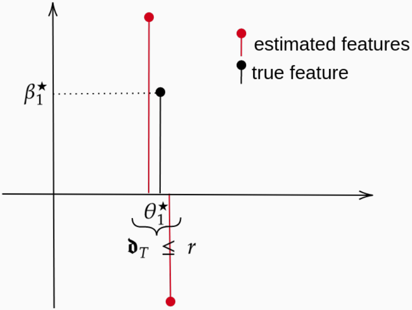
Let $(\hat{\beta}, \hat{\vartheta}) \in \mathbb{R}^K \times \Theta^K$ be measurable functions of y solutions of $(\mathcal{P}_1(\kappa))$ "approximating" $(\beta^\star = (\beta_1^\star, \cdots, \beta_s^\star), \vartheta^\star = (\theta_1^\star, \cdots, \theta_s^\star))$. We define:

Estimation risks (II)

$$\sum_{k=1}^{s} \left| |\beta_k^\star| - \sum_{\ell \in S_k(r)} |\hat{\beta}_\ell| \right|,$$

with

$S_k(r) = \left\{ \ell,\ \hat{\beta}_\ell \neq 0 \text{ and } \mathfrak{d}_T(\hat{\theta}_\ell, \theta_k^\star) \leq r \right\}.$



Estimation: the Beurling Lasso (I)

The Beurling Lasso [De Castro & Gamboa, 2012]

$$\min_{\mu \,\in\, \mathcal{M}(\Theta_T)} \quad \frac{1}{2}\,\|y - \langle \phi_T, \mu\rangle\|_T^2 + \kappa\|\mu\|_{TV}. \qquad (\mathcal{P}_2(\kappa))$$

- $\mathcal{M}(\Theta_T)$ the set of measures on Θ_T.
- $\langle \phi_T, \mu\rangle = \int \phi_T(\theta)\,\mu(\mathrm{d}\theta)$.
- $\|\cdot\|_{TV}$ the total variation of a norm.
- $\kappa > 0$ tuning parameter.

Remark

Let $\mu = \sum_{k=1}^{K} \beta_k\,\delta_{\theta_k}$ (atomic measure), then:

$$\tfrac{1}{2}\,\|y - \langle \phi_T, \mu\rangle\|_T^2 + \kappa\|\mu\|_{TV} = \tfrac{1}{2}\|y - \beta\,\boldsymbol{\Phi}_T(\vartheta)\|_T^2 + \kappa\|\beta\|_{\ell_1}.$$

Estimation: the Beurling Lasso (II)

The Beurling Lasso [De Castro & Gamboa, 2012]

$$\min_{\mu \,\in\, \mathcal{M}(\Theta_T)} \quad \frac{1}{2} \left\| y - \langle \phi_T, \mu \rangle \right\|_T^2 + \kappa \|\mu\|_{TV}. \qquad (\mathcal{P}_2(\kappa))$$

Estimation: the Beurling Lasso (II)

The Beurling Lasso [De Castro & Gamboa, 2012]

$$\min_{\mu \in \mathcal{M}(\Theta_T)} \quad \frac{1}{2} \|y - \langle \phi_T, \mu \rangle\|_T^2 + \kappa \|\mu\|_{TV}. \qquad (\mathcal{P}_2(\kappa))$$

$\mathcal{P}_1(\kappa)$ **and** $\mathcal{P}_2(\kappa)$

- Existence of a solution to $\mathcal{P}_2(\kappa)$ [Bredies & Pikkarainen, 2013].
- If $\mathcal{P}_2(\kappa)$ admits a solution $\hat{\mu} = \sum_{k=1}^{K} \hat{\beta}_k \delta_{\hat{\theta}_k}$ then $(\hat{\beta}, \hat{\vartheta})$ is a solution to $\mathcal{P}_1(\kappa)$.
- If H_T has finite dimension K, then there exists a solution to $\mathcal{P}_2(\kappa)$ with K atoms at most, [Boyer et al, 2019].

Estimation: the Beurling Lasso (II)

The Beurling Lasso [De Castro & Gamboa, 2012]

$$\min_{\mu \in \mathcal{M}(\Theta_T)} \frac{1}{2} \|y - \langle \phi_T, \mu \rangle\|_T^2 + \kappa \|\mu\|_{TV}. \qquad (\mathcal{P}_2(\kappa))$$

$\mathcal{P}_1(\kappa)$ and $\mathcal{P}_2(\kappa)$

- Existence of a solution to $\mathcal{P}_2(\kappa)$ [Bredies & Pikkarainen, 2013].
- If $\mathcal{P}_2(\kappa)$ admits a solution $\hat{\mu} = \sum_{k=1}^{K} \hat{\beta}_k \delta_{\hat{\theta}_k}$ then $(\hat{\beta}, \hat{\vartheta})$ is a solution to $\mathcal{P}_1(\kappa)$.
- If H_T has finite dimension K, then there exists a solution to $\mathcal{P}_2(\kappa)$ with K atoms at most, [Boyer et al, 2019].

Ex : low-pass filter

The observation space is $H_T = L^2(\mathrm{Leb})$.

Estimation: the Beurling Lasso (II)

The Beurling Lasso [De Castro & Gamboa, 2012]

$$\min_{\mu \in \mathcal{M}(\Theta_T)} \quad \frac{1}{2} \|y - \langle \phi_T, \mu \rangle\|_T^2 + \kappa \|\mu\|_{TV}. \qquad (\mathcal{P}_2(\kappa))$$

$\mathcal{P}_1(\kappa)$ and $\mathcal{P}_2(\kappa)$

- Existence of a solution to $\mathcal{P}_2(\kappa)$ [Bredies & Pikkarainen, 2013].
- If $\mathcal{P}_2(\kappa)$ admits a solution $\hat{\mu} = \sum_{k=1}^{K} \hat{\beta}_k \delta_{\hat{\theta}_k}$ then $(\hat{\beta}, \hat{\vartheta})$ is a solution to $\mathcal{P}_1(\kappa)$.
- If H_T has finite dimension K, then there exists a solution to $\mathcal{P}_2(\kappa)$ with K atoms at most, [Boyer et al, 2019].

Numerical implementation

- Modified Frank-Wolfe algorithm [Boyd, Schiebinger & Recht, 2017], [Denoyelle, Duval, Peyré & Soubies 2020], [Globabae & Poon, 2022].

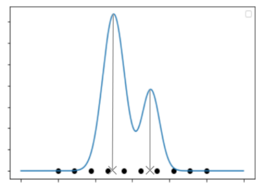
Estimation: why gridless methods

Estimation on a grid on the space of parameters

- Discrete grid of K points $\vartheta^{\mathcal{G}} = (\theta_1^{\mathcal{G}}, \cdots, \theta_K^{\mathcal{G}})$.
- Solve the Lasso problems:

$$\hat{\beta} \in \operatorname*{argmin}_{\beta \in \mathbb{R}^K} \quad \frac{1}{2}\|y - \beta\Phi_{\mathcal{G}}\|_T^2 + \kappa\|\beta\|_{\ell_1},$$

where $\Phi_{\mathcal{G}} = (\phi_T(\theta_1^{\mathcal{G}}), \cdots, \phi_T(\theta_K^{\mathcal{G}}))^\top$.



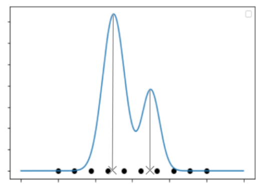
Estimation: why gridless methods

Estimation on a grid on the space of parameters

- Discrete grid of K points $\vartheta^{\mathcal{G}} = (\theta_1^{\mathcal{G}}, \cdots, \theta_K^{\mathcal{G}})$.
- Solve the Lasso problems:

$$\hat{\beta} \in \underset{\beta \in \mathbb{R}^K}{\operatorname{argmin}} \quad \frac{1}{2} \|y - \beta \Phi_{\mathcal{G}}\|_T^2 + \kappa \|\beta\|_{\ell_1},$$

where $\Phi_{\mathcal{G}} = (\phi_T(\theta_1^{\mathcal{G}}), \cdots, \phi_T(\theta_K^{\mathcal{G}}))^\top$.



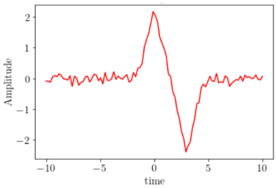
Inconvenients related to the refinement of the grid on Θ

- Strong correlations between the lines $\Phi_{\mathcal{G}} \implies$ numerical problems.
- The size of the grid grows exponentially with d when $\Theta \subset \mathbb{R}^d$.
- In the location model: clusters of spikes in the neighbourhood of the true spikes [Duval & Peyré, 2017].

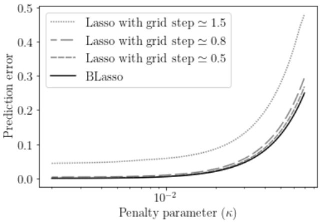
Estimation: numerical aspects (Blasso V Lasso on a grid)

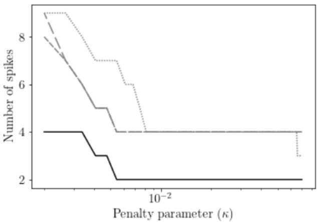
Numerical experiment

- Signal in $H_T = \mathbb{R}^T$ with $T = 100$, mixture of two Gaussian-shaped spikes with $\theta_1^\star = 0$ and $\theta_2^\star = 3$ and amplitudes in [-10,10] uniformly distributed, corrupted by i.i.d. centered Gaussian r. v. with $\sigma = 0.1$.



Results (GitHub: ClementHardy/PySFW)





Estimation: gridless methods

Bibliography

- **BLasso**: [De Castro and Gamboa, 2012], [Bredies & Pikkarainen, 2013].
- **Super-resolution** and **compressed sensing**: [Candès and Fernandez-Granda, 2013, 2014] (BLasso), [Bhaskar, Tang & Recht, 2013] (*atomic norm denoising*)...
- **Prediction and estimation** (dictionary composed of complex-valued Fourier basis): [Tang, Bhaskar & Recht 2014], [Boyer, De Castro & Salmon 2017].
- **Recover the support of a mesure and robustness of BLasso**: [Duval & Peyré, 2015] (spike deconvolution with weak noise $\|w_T\|_T \ll 1$) .
- **General geometric framework for BLasso**: [Poon, Keriven & Peyré, 2021].
- **Mixture (density) model**: [De Castro, Gadat, Marteau & Maugis, 2020].

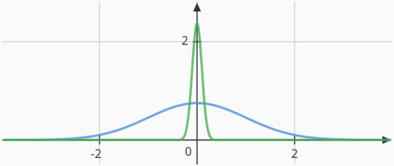
Estimation: assumptions (I)

Smoothness of the dictionary functions (H2)

- $\varphi_T : \Theta \to H_T$ is $\mathcal{C}^3$.
- $\|\varphi_T(\theta)\|_T > 0$ on Θ.
- $\phi_T(\theta) = \varphi_T(\theta)/\|\varphi_T(\theta)\|_T$.
- $\|\partial_\theta \phi_T(\theta)\|_T^2 > 0$ on Θ.

Ex: spectroscopy

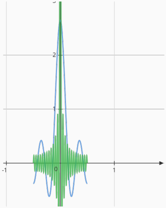
$$\varphi_T(\theta) = \mathrm{e}^{-(\theta-\cdot)^2/2\sigma_T^2}, \quad \theta \in \Theta = \mathbb{R}.$$



Ex: low-pass filter

$$\varphi_T(\theta) = \sum_{k=-f_c}^{f_c} \mathrm{e}^{2i\pi k(\theta-\cdot)} = \frac{\sin(\pi(\theta-\cdot)/\sigma_T)}{\sin(\pi(\theta-\cdot))},$$

$$\theta \in \Theta = \mathbb{R}/\mathbb{Z},\ \sigma_T = \frac{1}{T} = \frac{1}{2f_c+1}, \quad T \in 2\mathbb{N}^* + 1.$$



Estimation: assumptions (II)

Kernel and approximating kernel

We define a kernel on Θ^2 to measure the correlation between components of the dictionary:

$$\mathcal{K}_T(\theta, \theta') = \langle \phi_T(\theta), \phi_T(\theta') \rangle_T,$$

and an approximating symmetric kernel $\mathcal{K}^{\text{prox}}$ sur Θ_∞^2.

Estimation: assumptions (II)

Kernel and approximating kernel

We define a kernel on Θ^2 to measure the correlation between components of the dictionary:

$$\mathcal{K}_T(\theta, \theta') = \langle \phi_T(\theta), \phi_T(\theta') \rangle_T,$$

and an approximating symmetric kernel $\mathcal{K}^{\text{prox}}$ sur Θ_∞^2.

Ex: spectroscopy

If $\Delta_T \to 0$ and $\sigma_T = cst$, $\mathcal{K}^{\text{prox}} = \lim_{T \to +\infty} \mathcal{K}_T$.

Estimation: assumptions (II)

Kernel and approximating kernel

We define a kernel on Θ^2 to measure the correlation between components of the dictionary:

$$\mathcal{K}_T(\theta, \theta') = \langle \phi_T(\theta), \phi_T(\theta') \rangle_T,$$

and an approximating symmetric kernel $\mathcal{K}^{\text{prox}}$ sur Θ_∞^2.

Ex: spectroscopy

If $\Delta_T \to 0$ and $\sigma_T = cst$, $\qquad \mathcal{K}^{\text{prox}} = \lim_{T \to +\infty} \mathcal{K}_T$.

Assumptions on the approximating (H3)

The kernel $\mathcal{K}^{\text{prox}}$ is $\mathcal{C}^{3,3}$ with bounded derivatives + other smoothness assumptions. It is locally concave on the diagonal and strictly less than 1 outside the diagonal.

Estimation: assumptions (III)

Fisher-Rao metric on the space of parameters

$$\mathfrak{d}_{\mathcal{K}}(\theta,\theta') = \inf_{\gamma} \int_0^1 |\dot{\gamma}_s| \sqrt{\partial_{x,y}\mathcal{K}(\gamma_s,\gamma_s)}\,\mathrm{d}s$$

inf. on the set of smooth paths $\gamma : [0,1] \to \Theta$ such that $\gamma_0 = \theta$ and $\gamma_1 = \theta'$.

$\to$ **invariance** $\mathfrak{d}_{\mathcal{K}_\varphi}(\theta,\theta') = \mathfrak{d}_{\mathcal{K}_{\varphi\circ h}}(h^{-1}(\theta), h^{-1}(\theta'))$.

Estimation: assumptions (III)

Fisher-Rao metric on the space of parameters

$$\mathfrak{d}_{\mathcal{K}}(\theta, \theta') = \inf_{\gamma} \int_0^1 |\dot{\gamma}_s| \sqrt{\partial_{x,y}\mathcal{K}(\gamma_s, \gamma_s)}\, \mathrm{d}s$$

inf. on the set of smooth paths $\gamma : [0,1] \to \Theta$ such that $\gamma_0 = \theta$ and $\gamma_1 = \theta'$.

$\to$ **invariance** $\mathfrak{d}_{\mathcal{K}_\varphi}(\theta, \theta') = \mathfrak{d}_{\mathcal{K}_{\varphi \circ h}}(h^{-1}(\theta), h^{-1}(\theta'))$.

Univariate examples

- **Translated spikes model (spectroscopy / low-pass filter):**

$$\mathfrak{d}_{\mathcal{K}_T}(\theta, \theta') \sim |\theta - \theta'| \text{ (Euclidean distance).}$$

- **Scale model:** $H = L^2(\mathrm{Leb})$ and $\varphi(\theta) = e^{-\cdot\theta}$ with $\Theta = \mathbb{R}_+^*$ and

$$\mathfrak{d}_{\mathcal{K}_T}(\theta, \theta') \propto |\log(\theta/\theta')|.$$

  We have $\mathfrak{d}_{\mathcal{K}_T}(\theta, \theta + \varepsilon) \underset{\theta \to 0}{\to} +\infty$ ($\neq$ Euclidean distance).

Estimation: assumptions (IV)

Proximity of $\mathcal{K}_T$ and $\mathcal{K}^{\text{prox}}$

- **Proximity between kernels:**

$$\mathcal{V}_T = \max_{i,j \in \{0,\cdots,3\}} \sup_{\Theta_T^2} |\mathcal{K}_T^{[i,j]} - \mathcal{K}^{\text{prox}[i,j]}|.$$

- **equivalent metrics** : $\mathfrak{d}_{\mathcal{K}_T}$ and $\mathfrak{d}_{\mathcal{K}^{\text{prox}}}$: $\quad \mathfrak{d}_{\mathcal{K}^{\text{prox}}} / \rho_T \leq \mathfrak{d}_{\mathcal{K}_T} \leq \rho_T \, \mathfrak{d}_{\mathcal{K}^{\text{prox}}}$

Estimation: assumptions (IV)

Proximity of $\mathcal{K}_T$ and $\mathcal{K}^{\mathbf{prox}}$

- **Proximity between kernels:**

$$\mathcal{V}_T = \max_{i,j \in \{0,\cdots,3\}} \sup_{\Theta_T^2} |\mathcal{K}_T^{[i,j]} - \mathcal{K}^{\text{prox}[i,j]}|.$$

- **equivalent metrics** : $\mathfrak{d}_{\mathcal{K}_T}$ and $\mathfrak{d}_{\mathcal{K}^{\text{prox}}}$: $\quad \mathfrak{d}_{\mathcal{K}^{\text{prox}}} / \rho_T \leq \mathfrak{d}_{\mathcal{K}_T} \leq \rho_T \, \mathfrak{d}_{\mathcal{K}^{\text{prox}}}$

Proximity assumptions between $\mathcal{K}_T$ and $\mathcal{K}^{\mathbf{prox}}$ (H4)

$$s\, \mathcal{V}_T \leq C \quad \text{and} \quad \rho_T \leq \rho.$$

Estimation: assumptions (IV)

Proximity of $\mathcal{K}_T$ and $\mathcal{K}^{\text{prox}}$

- **Proximity between kernels:**

$$\mathcal{V}_T = \max_{i,j\in\{0,\cdots,3\}} \sup_{\Theta_T^2} |\mathcal{K}_T^{[i,j]} - \mathcal{K}^{\text{prox}[i,j]}|.$$

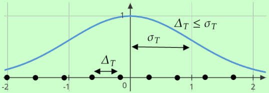
- **equivalent metrics** : $\mathfrak{d}_{\mathcal{K}_T}$ and $\mathfrak{d}_{\mathcal{K}^{\text{prox}}}$: $\quad \mathfrak{d}_{\mathcal{K}^{\text{prox}}} / \rho_T \leq \mathfrak{d}_{\mathcal{K}_T} \leq \rho_T\, \mathfrak{d}_{\mathcal{K}^{\text{prox}}}$

Proximity assumptions between $\mathcal{K}_T$ and $\mathcal{K}^{\text{prox}}$ (H4)

$$s\, \mathcal{V}_T \leq C \quad \text{and} \quad \rho_T \leq \rho.$$

Ex: spectroscopy

T : number of grid points.
Δ_T : grid step-size.
σ_T : scale parameter.



$$\frac{\Delta_T}{\sigma_T} \to 0 \text{ and } \frac{\Delta_T\, T}{\sigma_T} \to +\infty \implies \mathcal{V}_T \to 0 \text{ and } \rho_T \to 1.$$

Estimation: bounds on prediction and estimation errors

Theorem 1

We observe $y \in H_T$ with unknown parameters $\beta^\star \in \mathbb{R}^s$ and $\vartheta^\star = (\theta_1^\star, \cdots, \theta_s^\star) \in \Theta_T^s$ avec $s \leq K$ such that assumptions H1-H4 are checked and, for all $\ell \neq k$,

$$\mathfrak{d}_{\mathcal{K}_T}(\theta_\ell^\star, \theta_k^\star) \gtrsim \delta(s) \ \textbf{(séparation)}.$$

Then, the estimators $\hat{\beta}$ and $\hat{\vartheta}$ defined by $\mathcal{P}_1(\kappa)$ with

$$\kappa \geq \mathcal{C}_1 \sqrt{\Delta_T \log \tau} \quad \text{and} \quad \tau > 1,$$

with **the following bounds on the prediction and estimation risks:**

$$\left\| \beta^\star \, \Phi_T(\vartheta^\star) - \hat{\beta} \, \Phi_T(\hat{\vartheta}) \right\|_T \leq \mathcal{C}_0 \sqrt{s} \, \kappa,$$

$$\sum_{k=1}^{s} \left| |\beta_k^\star| - \sum_{\ell \in S_k(r)} |\hat{\beta}_\ell| \right| \quad + \quad \sum_{k=1}^{s} \left| \beta_k^\star - \sum_{\ell \in S_k(r)} \hat{\beta}_\ell \right| \quad + \quad \left\| \hat{\beta}_{S(r)^c} \right\|_{\ell_1} \leq \mathcal{C}_0 \, \kappa \, s.$$

with probability larger than: $1 - \mathcal{C}_2 \left(\frac{|\Theta_T|_{\mathfrak{d}_T}}{\tau \sqrt{\log \tau}} \vee \frac{1}{\tau} \right)$.

Remark: the bounds do not depend on K!

Estimation: separation between parameters

Separation condition

$$\mathfrak{d}_{\mathcal{K}_T}(\theta_\ell^\star, \theta_k^\star) \gtrsim \delta(s),\ \ell \neq k.$$

Estimation: separation between parameters

Separation condition

$$\mathfrak{d}_{\mathcal{K}_T}(\theta_\ell^\star, \theta_k^\star) \gtrsim \delta(s),\ \ell \neq k.$$

Definition of the separation $\delta(s)$

$$\delta(s) = \inf \Big\{ \delta > 0 : \max_{1 \leq \ell \leq s} \sum_{k=1, k\neq \ell}^{s} |\mathcal{K}^{\mathrm{prox}[i,j]}(\theta_\ell, \theta_k)| \leq u \quad \text{for all}$$

$$(i,j) \in \{0,1\} \times \{0,1,2\} \text{ and } (\theta_1, \cdots, \theta_s) \in \Theta^s \quad \text{s.t.} \quad \mathfrak{d}_{\mathcal{K}^{\mathrm{prox}}}(\theta_\ell, \theta_k) > \delta, \ell \neq k \Big\}.$$

Estimation: separation between parameters

Separation condition

$$\mathfrak{d}_{\mathcal{K}_T}(\theta_\ell^\star, \theta_k^\star) \gtrsim \delta(s),\ \ell \neq k.$$

Definition of the separation $\delta(s)$

$$\delta(s) = \inf \Big\{ \delta > 0 : \max_{1 \leq \ell \leq s} \sum_{k=1, k\neq \ell}^{s} |\mathcal{K}^{\mathrm{prox}[i,j]}(\theta_\ell, \theta_k)| \leq u \quad \text{for all}$$

$$(i,j) \in \{0,1\} \times \{0,1,2\} \text{ and } (\theta_1, \cdots, \theta_s) \in \Theta^s \quad \text{s.t.} \quad \mathfrak{d}_{\mathcal{K}^{\mathrm{prox}}}(\theta_\ell, \theta_k) > \delta, \ell \neq k \Big\}.$$

Ex: spectroscopy

$$\mathfrak{d}_{\mathcal{K}_T}(\theta, \theta') \sim \frac{|\theta - \theta'|}{\sigma_T} \quad \text{and} \quad \delta(s) \lesssim 1/u, \text{ for some } u > 0.$$

Estimation: bounds on the prediction risk

Ex: spectroscopy

- **Grid:** $t_1 < \cdots < t_T$ regular on $\mathbb{R}$ with step-size Δ_T.
- **Noise:** $w_T(t_j)$ i.i.d $\sim \mathcal{N}(0,1)$.
- **Dictionary:** $(\varphi_T(\theta) = \mathrm{e}^{-(\theta-\cdot)^2/2\sigma_T^2}, \theta \in \Theta = \mathbb{R})$.
- **Separation:** $|\theta_\ell^\star - \theta_k^\star| \gtrsim \sigma_T$ for $\ell \neq k$.

We have for $\sigma_T \gtrsim \Delta_T$:

$$\frac{1}{\sqrt{T}} \left\| \hat{\beta}\Phi_T(\hat{\vartheta}) - \beta^\star \Phi_T(\vartheta^\star) \right\|_{\ell_2} \lesssim \sqrt{\frac{s \log(T)}{T}},$$

with probability larger than $1 - C\left(\frac{t_T - t_1}{\sigma_T\, T\, \sqrt{\log(T)}} \vee \frac{1}{T}\right)$.

Estimation: bounds on the prediction risk

Ex: spectroscopy

- **Grid:** $t_1 < \cdots < t_T$ regular on $\mathbb{R}$ with step-size Δ_T.
- **Noise:** $w_T(t_j)$ i.i.d $\sim \mathcal{N}(0,1)$.
- **Dictionary:** $(\varphi_T(\theta) = \mathrm{e}^{-(\theta-\cdot)^2/2\sigma_T^2}, \theta \in \Theta = \mathbb{R})$.
- **Separation:** $|\theta^\star_\ell - \theta^\star_k| \gtrsim \sigma_T$ for $\ell \neq k$.

We have for $\sigma_T \gtrsim \Delta_T$:

$$\frac{1}{\sqrt{T}}\left\|\hat{\beta}\Phi_T(\hat{\vartheta}) - \beta^\star\Phi_T(\vartheta^\star)\right\|_{\ell_2} \lesssim \sqrt{\frac{s\log(T)}{T}},$$

with probability larger than $1 - C\left(\frac{t_T - t_1}{\sigma_T\, T\,\sqrt{\log(T)}} \vee \frac{1}{T}\right)$.

$\rightarrow$ The upper bound is of the same order as that for the Lasso estimator in the linear regression model.

Estimation: bounds on the prediction risk

Ex: low-pass filter

- **Truncated white noise:** $w_T = \frac{1}{\sqrt{T}} \sum_{k=1}^{T} G_k \, \psi_k$.
- **Dictionary:** $(\varphi_T(\theta) = \frac{\sin(T\pi(\theta - \cdot))}{\sin(\pi(\theta - \cdot))}, \, \theta \in \Theta = \mathbb{R}/\mathbb{Z})$.
- **Separation:** $|\theta_\ell^\star - \theta_k^\star| \gtrsim \delta(s)/T$ pour $\ell \neq k$.

We have for $T \gtrsim s$:

$$\left\| \hat{\beta}\Phi_T(\hat{\vartheta}) - \beta^\star \Phi_T(\vartheta^\star) \right\|_{L^2(\mathrm{Leb})} \lesssim \sqrt{\frac{s \log(T)}{T}},$$

with probability larger than $1 - \frac{C}{T\sqrt{\log(T)}}$.

III Tests

Tests: $\beta^\star \Phi_T(\vartheta^\star) = \beta^0 \Phi_T(\vartheta^0)$

Goodness-of-fit:

Let $(\beta^0, \vartheta^0) \in (\mathbb{R}^*)^{s^0} \times \Theta_T^{s^0}$ (**known**) and $(\beta^\star, \vartheta^\star) \in (\mathbb{R}^*)^{s} \times \Theta_T^{s}$ (**unknown**).

$$\begin{cases} H_0 & : \quad \beta^\star \Phi_T(\vartheta^\star) = \beta^0 \Phi_T(\vartheta^0), \\ H_1(\rho) & : \quad \left\| \beta^\star \Phi_T(\vartheta^\star) - \beta^0 \Phi_T(\vartheta^0) \right\|_T \geq \rho. \end{cases}$$

Tests: $\beta^\star \Phi_T(\vartheta^\star) = \beta^0 \Phi_T(\vartheta^0)$

Goodness-of-fit:

Let $(\beta^0, \vartheta^0) \in (\mathbb{R}^*)^{s^0} \times \Theta_T^{s^0}$ (**known**) and $(\beta^\star, \vartheta^\star) \in (\mathbb{R}^*)^{s} \times \Theta_T^{s}$ (**unknown**).

$$\begin{cases} H_0 & : \quad \beta^\star \Phi_T(\vartheta^\star) = \beta^0 \Phi_T(\vartheta^0), \\ H_1(\rho) & : \quad \left\| \beta^\star \Phi_T(\vartheta^\star) - \beta^0 \Phi_T(\vartheta^0) \right\|_T \geq \rho. \end{cases}$$

Testing risk and minimal separation

A test Ψ is a measurable function y taking values in $\{0, 1\}$:
$\Psi = 0$ **accept** H_0 and $\Psi = 1$ **reject** H_0.

- **The maximal testing risk**:

$$R_\rho(\Psi) = \underbrace{\sup_{(\beta^\star, \vartheta^\star) \in H_0} \mathbb{E}_{(\beta^\star, \vartheta^\star)}[\Psi]}_{\text{1st type error prob.}} + \underbrace{\sup_{(\beta^\star, \vartheta^\star) \in H_1(\rho)} \mathbb{E}_{(\beta^\star, \vartheta^\star)}[1 - \Psi]}_{\text{2nd type error prob.}}.$$

- **The minimax separation distance for testing at** $\alpha \in (0, 1)$:

$$\rho^\star(\alpha) = \inf\{\rho > 0 : \inf_{\Psi} R_\rho(\Psi) \leq \alpha\}.$$

Tests: upper bound on the minimax separation

Framework

- $H_T = L^2(\lambda_T)$ with $\lambda_T \to \mathrm{Leb}$ ($\mathbb{R}$ ou $\mathbb{R}/\mathbb{Z}$).
- Dictionary: $(\varphi_T(\theta) = h(\theta - \cdot, \sigma_T),\, \theta \in \Theta)$.

Ex: spectroscopy / low-pass filter.

Tests: upper bound on the minimax separation

Framework

- $H_T = L^2(\lambda_T)$ with $\lambda_T \to \mathrm{Leb}$ ($\mathbb{R}$ ou $\mathbb{R}/\mathbb{Z}$).
- Dictionary: $(\varphi_T(\theta) = h(\theta - \cdot, \sigma_T), \theta \in \Theta)$.

Ex: spectroscopy / low-pass filter.

Proposition 1

Under the assumptions H1-H4 (for $s \in \mathbb{N}$ and $s^0 \in \mathbb{N}$) assume also that :

$$|\theta^\star_\ell - \theta^\star_k| \gtrsim \sigma_T \, \delta(s) \quad \ell \neq k \text{ and that } |\theta^0_\ell - \theta^0_k| \gtrsim \sigma_T \, \delta(s^0) \quad \ell \neq k.$$

We have for $|\Theta_T|/\sigma_T \geq 1$ and $\alpha \in (0,1)$:

$$\rho^\star(\alpha) \lesssim \left(\frac{\mathrm{Var}\left(\|w_T\|^2_{L^2(\lambda_T)} \right)}{\alpha} \right)^{1/4} \wedge \sqrt{(s \vee s^0 \vee 1)\Delta_T \log\left(\frac{c\,|\Theta_T|}{\alpha\,\sigma_T} \right)}.$$

Tests: upper bound on the minimax separation

Corollary 1

Discrete process on a regular grid of T points on $\mathbb{R}/\mathbb{Z}$ with $w_T(t_j)$ i.i.d $\sim \mathcal{N}(0,1)$ for $1 \leq j \leq T$.

$$\begin{cases} H_0 & : \quad s = 0, \\ H_1(\rho) & : \quad \|\beta^\star\|_{\ell_2} \geq \rho. \end{cases}$$

The bound becomes:

$$\rho^\star(\alpha) \lesssim \min\left(\frac{1}{(\alpha T)^{\frac{1}{4}}}, \sqrt{\frac{s}{T}\log\left(\frac{c}{\alpha\,\sigma_T}\right)}\right).$$

Tests: upper bound on the minimax separation

Corollary 1

Discrete process on a regular grid of T points on $\mathbb{R}/\mathbb{Z}$ with $w_T(t_j)$ i.i.d $\sim \mathcal{N}(0,1)$ for $1 \leq j \leq T$.

$$\begin{cases} H_0 & : \quad s = 0, \\ H_1(\rho) & : \quad \|\beta^\star\|_{\ell_2} \geq \rho. \end{cases}$$

The bound becomes:

$$\rho^\star(\alpha) \lesssim \min\left(\frac{1}{(\alpha T)^{\frac{1}{4}}}, \sqrt{\frac{s}{T}\log\left(\frac{c}{\alpha\,\sigma_T}\right)}\right).$$

Remarks

- In the sparse linear regression model, a third regime appears with the rate $\frac{K^{1/4}}{\sqrt{T}}$; Ingster, Verzelen, Tsybakov (2010), Nickl and van de Geer, (2013), showed that the minimax testing rate is $\frac{1}{T^{1/4}} \wedge \sqrt{\frac{s\log(K)}{T}} \wedge \frac{K^{1/4}}{\sqrt{T}}$.

Conclusion and perspectives

Estimation and tests

Under least separation conditions between the true non linear parameters:

- Prediction risks of the same order as for the Lasso-type estimators when $\vartheta^\star$ is known.
- Testing separation rate is of the same order as for signal detection when $\vartheta^\star$ is given.
- (Simultaneous reconstruction (analogous to group-Lasso) when many signals share a common structure).

Perspectives

- $\Theta \subseteq \mathbb{R}^d$
- Improve on the non-linear parameter separation conditions in general
- Extend the testing problems