Off-the-grid learning of mixtures

Cristina Butucea (ENSAE), Jean-François Delmas (ENPC),
Anne Dutfoy (EDF R&D), Clément Hardy (EDF R&D, ENPC)

cristina.butucea@ensae.fr

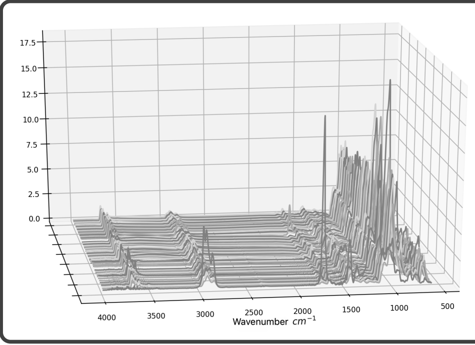
arxiv:2207.00171

Symposium on Inverse Problems, September 19-21, 2022

Infrared 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

Figure: Table of the location of peaks and their corresponding bonds for the polychloroprene samples ([Tchalla, 2017]).



$$y(t) = \sum_{k=1}^{s} \beta_k^\star \, \phi(\theta_k^\star, t) + w_T(t), \; (\phi(\theta, \cdot), \theta \in \Theta) \text{ continuous dictionary.}$$

We observe y a random element of the Hilbert space $(H_T, <\cdot,\cdot>_T)$, for $T \in \mathbb{N}^*$.

Continuous dictionary $\{\varphi_T(\theta), \theta \in \Theta\}$ of non-degenerate elements of H_T and the normalized functions

$$\phi_T(\theta) = \frac{\varphi_T(\theta)}{\|\varphi_T(\theta)\|_T}.$$

If H_T is a space of functions, we denote $\varphi_T(\theta) = \varphi_T(\theta, \cdot)$.

We observe y a random element of the Hilbert space $(H_T, < \cdot, \cdot >_T)$, for $T \in \mathbb{N}^*$.

Continuous dictionary $\{\varphi_T(\theta),\, \theta \in \Theta\}$ of non-degenerate elements of H_T and the normalized functions

$$\phi_T(\theta) = \frac{\varphi_T(\theta)}{\|\varphi_T(\theta)\|_T}.$$

If H_T is a space of functions, we denote $\varphi_T(\theta) = \varphi_T(\theta, \cdot)$.

We assume

$$y = \sum_{k=1}^{K} \beta_k^* \cdot \phi_T(\theta_k^*) + w_T,$$

where

- w_T is a centered Gaussian element of H_T,
- β^* in $\mathbb{R}^K$, $s-$sparse,
- $\{\theta_k^*\}_{k=1}^K$ included in Θ.

Model

$$y = \beta^* \cdot \Phi_T(\theta^*) + w_T, \quad \beta^* \in \mathbb{R}^K,$$

where β^* - row vector and $\Phi_T = (\phi_T(\theta_1^*), ..., \phi_T(\theta_K^*))^\top$.

Model

$$y = \beta^* \cdot \Phi_T(\theta^*) + w_T, \quad \beta^* \in \mathbb{R}^K,$$

where β^* - row vector and $\Phi_T = (\phi_T(\theta_1^*), ..., \phi_T(\theta_K^*))^\top$.

a) Discrete model Let $t_1 < ... < t_T$ in [0,1] be the design points, and $G_1, ..., G_T$ i.i.d. $N(0, \sigma^2)$, s.t.

$$y(t_j) = \beta^\star \Phi_T(\theta^\star, t_j) + G_j, \quad j = 1, ..., T.$$

We let $H_T = \mathbb{L}_2(\lambda_T)$, where $\lambda_T(dt) = \frac{1}{T}\sum_{j=1}^T \delta_{t_j}(dt)$. The noise process can be written:

$$w_T(t) = \sum_{j=1}^T G_j \cdot I(t = t_j).$$

Model

$$y = \beta^* \cdot \Phi_T(\theta^*) + w_T, \quad \beta^* \in \mathbb{R}^K,$$

where β^* - row vector and $\Phi_T = (\phi_T(\theta_1^*), ..., \phi_T(\theta_K^*))^\top$.

a) Discrete model Let $t_1 < ... < t_T$ in [0,1] be the design points, and $G_1, ..., G_T$ i.i.d. $N(0, \sigma^2)$, s.t.

$$y(t_j) = \beta^\star \Phi_T(\theta^\star, t_j) + G_j, \quad j = 1, ..., T.$$

We let $H_T = \mathbb{L}_2(\lambda_T)$, where $\lambda_T(dt) = \frac{1}{T}\sum_{j=1}^T \delta_{t_j}(dt)$. The noise process can be written:

$$w_T(t) = \sum_{j=1}^T G_j \cdot I(t = t_j).$$

Then, for any f in H_T,

$$Var(< f, w_T >_T) = Var\left(\frac{1}{T}\sum_{j=1}^T f(t_j)G_j\right) = \frac{\sigma^2}{T}\|f\|_T^2.$$

b) Continuous model with truncated or coloured noise: Let

$$w_T = \sum_{k:p_k>0} \sqrt{\xi_k} G_k \psi_k, \quad \{G_k\}_k \text{ i.i.d, } N(0, \sigma^2),$$

b) Continuous model with truncated or coloured noise: Let

$$w_T = \sum_{k: p_k > 0} \sqrt{\xi_k} G_k \psi_k, \quad \{G_k\}_k \text{ i.i.d, } N(0, \sigma^2),$$

where $\{\psi_k,\, k \in \mathbb{N}\}$ o.n.b. of continuous functions of $(\mathbb{L}_2[0,1], Leb)$;
and we choose $\{p_k\}_{k \in \mathbb{N}}$ and $\{\xi_k\}_{k \in \mathbb{N}}$ sequences of positive real numbers such that

$$\sum_k p_k \xi_k < \infty.$$

Examples

b) Continuous model with truncated or coloured noise: Let

$$w_T = \sum_{k:p_k>0} \sqrt{\xi_k} G_k \psi_k, \quad \{G_k\}_k \text{ i.i.d, } N(0,\sigma^2),$$

where $\{\psi_k,\, k \in \mathbb{N}\}$ o.n.b. of continuous functions of $(\mathbb{L}_2[0,1], Leb)$;
and we choose $\{p_k\}_{k\in\mathbb{N}}$ and $\{\xi_k\}_{k\in\mathbb{N}}$ sequences of positive real numbers such that

$$\sum_k p_k \xi_k < \infty.$$

We define the weighted Hilbert space:

$$H_T = \overline{< \{\psi_k, \quad k : p_k > 0\} >},$$

with $< f, g >_T = \sum_k p_k \cdot < f, \psi_k > \cdot < g, \psi_k >$. Typically, $p_k = \frac{1}{T} I(1 \le k \le T)$.

b) Continuous model with truncated or coloured noise: Let

$$w_T = \sum_{k:p_k>0} \sqrt{\xi_k} G_k \psi_k, \quad \{G_k\}_k \text{ i.i.d, } N(0,\sigma^2),$$

where $\{\psi_k,\, k \in \mathbb{N}\}$ o.n.b. of continuous functions of $(\mathbb{L}_2[0,1], Leb)$;
and we choose $\{p_k\}_{k\in\mathbb{N}}$ and $\{\xi_k\}_{k\in\mathbb{N}}$ sequences of positive real numbers such that

$$\sum_k p_k \xi_k < \infty.$$

We define the weighted Hilbert space:

$$H_T = \overline{< \{\psi_k, \quad k : p_k > 0\} >},$$

with $< f, g >_T = \sum_k p_k \cdot < f, \psi_k > \cdot < g, \psi_k >$. Typically, $p_k = \frac{1}{T} I(1 \leq k \leq T)$.

Then, for all f in H_T,

$$Var(< f, w_T >_T) = Var(\sum_k p_k \cdot < f, \psi_k > \cdot \sqrt{\xi_k} G_k)$$

$$= \sum_k p_k^2 < f, \psi_k >^2 \xi_k \sigma^2 \leq \sigma^2 \sup_k (p_k \xi_k) \cdot \|f\|_T^2$$

They can be stated and applied to:
-learning mixtures, compressed sensing, two-layer neural networks, low-rank tensor product of matrices, super-resolution in signal processing.

They can be stated and applied to:
-learning mixtures, compressed sensing, two-layer neural networks, low-rank tensor product of matrices, super-resolution in signal processing.

Beurling-Lasso (BLasso) de Castro and Gamboa, 2012; Bredies and Pikkarainen, 2013;
- convex optimization problem over a set of Radon measures $\mathcal{M}(\Theta)$ on the space Θ:

$$\mathcal{P}(\kappa): \quad \arg\min_{\mu \in \mathcal{M}(\Theta)} \frac{1}{2}\|y - \Phi\mu\|_T^2 + \kappa|\mu|(\Theta),$$

where $\Phi : \mathcal{M}(\Theta) \to H_T$ is the acquisition operator and $|\mu|$ denotes the total variation of the measure μ.

They can be stated and applied to:
-learning mixtures, compressed sensing, two-layer neural networks, low-rank tensor product of matrices, super-resolution in signal processing.

Beurling-Lasso (BLasso) de Castro and Gamboa, 2012; Bredies and Pikkarainen, 2013;
- convex optimization problem over a set of Radon measures $\mathcal{M}(\Theta)$ on the space Θ:

$$\mathcal{P}(\kappa): \quad \arg\min_{\mu\in\mathcal{M}(\Theta)} \frac{1}{2}\|y-\Phi\mu\|_T^2+\kappa|\mu|(\Theta),$$

where $\Phi : \mathcal{M}(\Theta) \to H_T$ is the acquisition operator and $|\mu|$ denotes the total variation of the measure μ.

Remark that $\Phi\mu = \int \phi(w,\cdot)d\mu(w)$ is equal to $\sum_k \beta_k^* \phi(\theta_k^*,\cdot)$ for

$$d\mu(w) = \sum_{k\in S^\star} \beta_k^* \delta_{\theta_k^*}(dw).$$

Note that $|\mu|(\Theta) = \sum_{k\in S^\star} |\beta_k^\star|$.

The dual problem

$$\mathcal{D}(\kappa): \quad \arg\max_{p:\|\Phi^\star p\|_\infty \leq 1} < y, p >_T - \frac{\kappa}{2}\|p\|_T^2$$

The dual problem

$$\mathcal{D}(\kappa): \quad \arg\max_{p:\|\Phi^\star p\|_\infty \leq 1} < y, p >_T - \frac{\kappa}{2}\|p\|_T^2$$

The measure μ_κ solution to the problem $\mathcal{P}(\kappa)$ and p_κ the unique solution of $\mathcal{D}(\kappa)$ are related through:

$$\begin{cases} \Phi^\star p_\kappa & \in \partial|\mu_\kappa| \\ -p_\kappa & = \frac{1}{\kappa}(\Phi\mu_\kappa - y) \end{cases}$$

where the subdifferential $\partial|\mu|$ is the set of continuous functions g, vanishing at infinity, bounded by 1: $\|g\|_\infty \leq 1$, such that $\int_\Theta g d\mu = |\mu|$.

The dual problem

$$\mathcal{D}(\kappa): \quad \arg \max_{p:\|\Phi^\star p\|_\infty \leq 1} < y, p >_T - \frac{\kappa}{2}\|p\|_T^2$$

The measure μ_κ solution to the problem $\mathcal{P}(\kappa)$ and p_κ the unique solution of $\mathcal{D}(\kappa)$ are related through:

$$\begin{cases} \Phi^\star p_\kappa & \in \partial|\mu_\kappa| \\ -p_\kappa & = \frac{1}{\kappa}(\Phi\mu_\kappa - y) \end{cases}$$

where the subdifferential $\partial|\mu|$ is the set of continuous functions g, vanishing at infinity, bounded by 1: $\|g\|_\infty \leq 1$, such that $\int_\Theta g d\mu = |\mu|$.

Definition: $\eta_\kappa := \Phi^\star p_\kappa$ is a dual certificate of μ_κ.

The dual problem

$$\mathcal{D}(\kappa): \quad \arg \max_{p:\|\Phi^\star p\|_\infty \leq 1} < y, p >_T - \frac{\kappa}{2}\|p\|_T^2$$

The measure μ_κ solution to the problem $\mathcal{P}(\kappa)$ and p_κ the unique solution of $\mathcal{D}(\kappa)$ are related through:

$$\begin{cases} \Phi^\star p_\kappa & \in \partial|\mu_\kappa| \\ -p_\kappa & = \frac{1}{\kappa}(\Phi\mu_\kappa - y) \end{cases}$$

where the subdifferential $\partial|\mu|$ is the set of continuous functions g, vanishing at infinity, bounded by 1: $\|g\|_\infty \leq 1$, such that $\int_\Theta g d\mu = |\mu|$.

Definition: $\eta_\kappa := \Phi^\star p_\kappa$ is a dual certificate of μ_κ.

Remark: -the solution to the problem $\mathcal{P}(\kappa)$ is not necessarily a discrete measure; if $N := dim(Im(\Phi))$ is finite then a solution which is a discrete measure with at most N atoms can be found.
Therefore, we proceed with a slightly different optimization problem so that we recover a discrete mixture as solution.

Let

$$(\hat{\beta}, \hat{\theta}) := \arg \min_{\beta \in \mathbb{R}^K, \theta \in (\Theta_T)^K} \frac{1}{2} \|y - \beta \Phi_T(\theta)\|_T^2 + \kappa \|\beta\|_1$$

where $\Theta_T \subset \Theta$ is a compact set.

1.3 Estimator

Let

$$(\hat{\beta}, \hat{\theta}) := \arg \min_{\beta \in \mathbb{R}^K, \theta \in (\Theta_T)^K} \frac{1}{2} \|y - \beta \Phi_T(\theta)\|_T^2 + \kappa \|\beta\|_1$$

where $\Theta_T \subset \Theta$ is a compact set.

The algorithms used to solve numerically (also the BLasso):

1. Sliding Franck-Wolfe algorithm (Denoyel et al. 2019)
2. conic particle gradient descent (Chizat, 2021)

Let

$$(\hat{\beta}, \hat{\theta}) := \arg \min_{\beta \in \mathbb{R}^K, \theta \in (\Theta_T)^K} \frac{1}{2} \|y - \beta \Phi_T(\theta)\|_T^2 + \kappa \|\beta\|_1$$

where $\Theta_T \subset \Theta$ is a compact set.

The algorithms used to solve numerically (also the BLasso):

1. Sliding Franck-Wolfe algorithm (Denoyel et al. 2019)
2. conic particle gradient descent (Chizat, 2021)

We will give high-probability bounds for the prediction risk

$$\|\hat{\beta}\Phi(\hat{\theta}) - \beta^* \Phi(\theta^*)\|_T^2$$

and some estimation results.

Let

$$(\hat\beta, \hat\theta) := \arg \min_{\beta \in \mathbb{R}^K, \theta \in (\Theta_T)^K} \frac{1}{2} \|y - \beta \Phi_T(\theta)\|_T^2 + \kappa \|\beta\|_1$$

where $\Theta_T \subset \Theta$ is a compact set.

The algorithms used to solve numerically (also the BLasso):

1. Sliding Franck-Wolfe algorithm (Denoyel et al. 2019)
2. conic particle gradient descent (Chizat, 2021)

We will give high-probability bounds for the prediction risk

$$\|\hat\beta \Phi(\hat\theta) - \beta^* \Phi(\theta^*)\|_T^2$$

and some estimation results.

Bibliography:
-For known θ^*, linear regression model! Bühlmann and van de Geer 2011, Giraud 2015.

Bibliography

Self-modeling **non-linear regression**: Golub, Pereyra, 1973; Kneip, Gasser, 1988 (consistency results for finite dimensional model);
BLasso : de Castro and Gamboa, 2012;
Super-resolution and **compressed sensing**: Candès and Fernandez-Granda, 2013, 2014; Tang, 2015; ...
Overcomplete dictionary techniques, sparse coding: Donoho, Elad, Temlyakov 2006; Tang, Baskhar, Recht, 2013; ...

Bibliography

Self-modeling **non-linear regression**: Golub, Pereyra, 1973; Kneip, Gasser, 1988 (consistency results for finite dimensional model);
BLasso : de Castro and Gamboa, 2012;
Super-resolution and **compressed sensing**: Candès and Fernandez-Granda, 2013, 2014; Tang, 2015; ...
Overcomplete dictionary techniques, sparse coding: Donoho, Elad, Temlyakov 2006; Tang, Baskhar, Recht, 2013; ...

Off-the-grid methods for the regression model -
-Fourier basis features: Tang, Baskhar, Recht 2015; Boyer, de Castro, Salmon, 2017;
-Location families: "Fixed-grid + Lasso" produces clusters of spikes around true location parameters - Duval, Peyré, 2017;
Off-the-grid method produces perturbations in the location and amplitude estimation - Duval, Peyré, 2015;

for the density model - de Castro, Gadat, Marteau, Maugis-Rabusseau, 2021 - solve the BLasso for a different risk measure, find rates for the prediction risk under minimal separation conditions.

Bibliography

Self-modeling **non-linear regression**: Golub, Pereyra, 1973; Kneip, Gasser, 1988 (consistency results for finite dimensional model);
BLasso : de Castro and Gamboa, 2012;
Super-resolution and **compressed sensing**: Candès and Fernandez-Granda, 2013, 2014; Tang, 2015; ...
Overcomplete dictionary techniques, sparse coding: Donoho, Elad, Temlyakov 2006; Tang, Baskhar, Recht, 2013; ...

Off-the-grid methods for the regression model -
-Fourier basis features: Tang, Baskhar, Recht 2015; Boyer, de Castro, Salmon, 2017;
-Location families: "Fixed-grid + Lasso" produces clusters of spikes around true location parameters - Duval, Peyré, 2017;
Off-the-grid method produces perturbations in the location and amplitude estimation - Duval, Peyré, 2015;

for the density model - de Castro, Gadat, Marteau, Maugis-Rabusseau, 2021 - solve the BLasso for a different risk measure, find rates for the prediction risk under minimal separation conditions.

-Non translation invariant models: Poon, Keriven, Peyré, 2021 describe the natural geometric framework of the BLasso, show that the resulting measure recovers the true one in Wasserstein metric.

Smoothness of the dictionary: Assume $\varphi_T : \Theta \to H_T$ is of class $\mathcal{C}^3$ and that $\|\varphi_T(\theta)\|_T > 0$ on Θ. Moreover, we assume that

$$g_T(\theta) := \|\partial_\theta \phi_T(\theta)\|_T^2 > 0, \text{ on } \Theta.$$

Smoothness of the dictionary: Assume $\varphi_T : \Theta \to H_T$ is of class $\mathcal{C}^3$ and that $\|\varphi_T(\theta)\|_T > 0$ on Θ. Moreover, we assume that

$$g_T(\theta) := \|\partial_\theta \phi_T(\theta)\|_T^2 > 0, \text{ on } \Theta.$$

Examples:
a) Locations families, i.e.

$$\varphi_T(\theta, t) = v\left(\frac{t-\theta}{\sigma_0}\right)$$

for some known spread parameter σ_0:
- Gaussian family: $v(t) = \exp(-\frac{1}{2}t^2)$
- Cauchy family: $v(t) = (1+t^2)^{-1}$
-$sinc$-kernel: $v(t) = \frac{\sin(\pi t)}{\pi t}$
but not the Laplace kernel $v(t) = \exp(-\frac{1}{2}|t|)$.

Smoothness of the dictionary: Assume $\varphi_T : \Theta \to H_T$ is of class $\mathcal{C}^3$ and that $\|\varphi_T(\theta)\|_T > 0$ on Θ. Moreover, we assume that

$$g_T(\theta) := \|\partial_\theta \phi_T(\theta)\|_T^2 > 0, \text{ on } \Theta.$$

Examples:

a) Locations families, i.e.

$$\varphi_T(\theta, t) = v\left(\frac{t-\theta}{\sigma_0}\right)$$

for some known spread parameter σ_0:
- Gaussian family: $v(t) = \exp(-\frac{1}{2}t^2)$
- Cauchy family: $v(t) = (1+t^2)^{-1}$
-$sinc$-kernel: $v(t) = \frac{\sin(\pi t)}{\pi t}$
but not the Laplace kernel $v(t) = \exp(-\frac{1}{2}|t|)$.

b) Scaling families, i.e.

$$\varphi_T(\theta, t) = v(\theta \cdot t).$$

-Laplace transform for $v(t) = exp(-t)$.

We define the kernel $\mathcal{K}_T$ on Θ^2 by:

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

2.2 Kernel and Riemannian metric

We define the kernel $\mathcal{K}_T$ on Θ^2 by:

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

We have

$$g_T(\theta) = \partial^2_{xy} \mathcal{K}_T(\theta, \theta),$$

defining an intrinsic **Riemannian metric** on Θ^2:

$$\mathfrak{d}_T(\theta, \theta') = |G_T(\theta) - G_T(\theta')|,$$

where G_T is a primitive of $\sqrt{g_T}$.

In particular, we use Taylor expansion in θ wrt the metric $\mathfrak{d}_T$ and covariant derivatives.

Approximating limit kernel

The kernel has the properties

$$\mathcal{K}_T(\theta,\theta) = 1, \quad \mathcal{K}_T^{[1,0]}(\theta,\theta) = 0, \quad \mathcal{K}_T^{[2,0]}(\theta,\theta) = -1,$$

$$\mathcal{K}_T^{[2,1]}(\theta,\theta) = 0 \quad \text{and} \quad \sup_{\Theta^2} |\mathcal{K}_T^{[0,0]}| \le 1.$$

Denote by $h_T(\theta) = \mathcal{K}_T^{[3,3]}(\theta,\theta)$.

Approximating limit kernel

The kernel has the properties

$$\mathcal{K}_T(\theta,\theta) = 1, \quad \mathcal{K}_T^{[1,0]}(\theta,\theta) = 0, \quad \mathcal{K}_T^{[2,0]}(\theta,\theta) = -1,$$

$$\mathcal{K}_T^{[2,1]}(\theta,\theta) = 0 \quad \text{and} \quad \sup_{\Theta^2} |\mathcal{K}_T^{[0,0]}| \leq 1.$$

Denote by $h_T(\theta) = \mathcal{K}_T^{[3,3]}(\theta,\theta)$.

We assume there exists an approximating **limit kernel** $\mathcal{K}_\infty$ on Θ_∞ which are free of T, satisfying smoothness conditions and boundedness conditions:

$$\inf_{\Theta_\infty} g_\infty > 0, \quad \sup_{\Theta_\infty} h_\infty < +\infty, \text{ and } \sup_{\Theta_\infty^2} |\mathcal{K}_\infty^{[i,j]}| < +\infty \quad \text{for all } i,j \in \{0,1,2\}.$$

Approximating limit kernel

The kernel has the properties

$$\mathcal{K}_T(\theta,\theta) = 1, \quad \mathcal{K}_T^{[1,0]}(\theta,\theta) = 0, \quad \mathcal{K}_T^{[2,0]}(\theta,\theta) = -1,$$

$$\mathcal{K}_T^{[2,1]}(\theta,\theta) = 0 \quad \text{and} \quad \sup_{\Theta^2} |\mathcal{K}_T^{[0,0]}| \le 1.$$

Denote by $h_T(\theta) = \mathcal{K}_T^{[3,3]}(\theta,\theta)$.

We assume there exists an approximating **limit kernel** $\mathcal{K}_\infty$ on Θ_∞ which are free of T, satisfying smoothness conditions and boundedness conditions:

$$\inf_{\Theta_\infty} g_\infty > 0, \quad \sup_{\Theta_\infty} h_\infty < +\infty, \text{ and } \sup_{\Theta_\infty^2} |\mathcal{K}_\infty^{[i,j]}| < +\infty \quad \text{for all } i,j \in \{0,1,2\}.$$

Proximity to the limit kernel. There exist a constant $L > 0$:

$$\max\{\max_{i,j\in\{0,1,2\}} \sup_{\Theta_T^2} |\mathcal{K}_T^{[i,j]} - \mathcal{K}_\infty^{[i,j]}|, \quad \sup_{\Theta_T} |h_T - h_\infty|\} \le L.$$

Example: mixture of Gaussian features, discrete regression model

We observe y on a regular grid on $\Theta_T = [-a_T, a_T]$ with step

$$\Delta_T = \frac{2a_T}{T}.$$

The Gaussian features have spread σ_0.

Example: mixture of Gaussian features, discrete regression model

We observe y on a regular grid on $\Theta_T = [-a_T, a_T]$ with step

$$\Delta_T = \frac{2a_T}{T}.$$

The Gaussian features have spread σ_0.

The limit space: if $a_T \to \infty$ and $\Delta_T \to 0$, then $\lambda_\infty = Leb$ on $\Theta_\infty = \mathbb{R}$.
We calculate

$$\phi_\infty(\theta) = \frac{1}{\pi^{\frac{1}{4}}\sqrt{\sigma_0}}\varphi(\theta), \quad \mathcal{K}_\infty(\theta, \theta') = v\left(\frac{\theta - \theta'}{\sqrt{2}\,\sigma_0}\right) \quad \text{and} \quad g_\infty(\theta) = \frac{1}{2\sigma_0^2}$$

and

$$\mathfrak{d}_\infty(\theta, \theta') = \frac{|\theta - \theta'|}{\sqrt{2}\,\sigma_0}.$$

Let $\mathcal{Q}^\star$ be a set of s elements in Θ_T. Suppose that

$$\mathfrak{d}_T(\theta, \theta') > 2r \text{ for all } \theta, \theta' \in \mathcal{Q}^\star,$$

and that there exist positive constants C_N, C'_N, C_F, C_B, with $C_F < 1$, depending on r and $\mathcal{K}_\infty$ such that

for any application $v : \mathcal{Q}^\star \to \{-1, 1\}$ there exists an element $p \in H_T$ satisfying:

1. For all $\theta^\star \in \mathcal{Q}^\star$ and $\theta \in \mathcal{B}_T(\theta^\star, r)$, we have $|\langle \phi_T(\theta), p \rangle_T| \leq 1 - C_N\, \mathfrak{d}_T(\theta^\star, \theta)^2$.
2. For all $\theta^\star \in \mathcal{Q}^\star$ and $\theta \in \mathcal{B}_T(\theta^\star, r)$, we have $|\langle \phi_T(\theta), p \rangle_T - v(\theta^\star)| \leq C'_N\, \mathfrak{d}_T(\theta^\star, \theta)^2$.
3. For all θ in Θ_T and $\theta \notin \bigcup_{\theta^\star \in \mathcal{Q}^\star} \mathcal{B}_T(\theta^\star, r)$ (far region), we have

$$|\langle \phi_T(\theta), p \rangle_T| \leq 1 - C_F.$$

4. We have $\|p\|_T \leq C_B \sqrt{s}$.

Existence of Interpolating Certificate

Let $\mathcal{Q}^\star$ be a set of s elements in Θ_T. Suppose that

$$\mathfrak{d}_T(\theta, \theta') > 2r \text{ for all } \theta, \theta' \in \mathcal{Q}^\star,$$

and that there exist positive constants C_N, C'_N, C_F, C_B, with $C_F < 1$, depending on r and $\mathcal{K}_\infty$ such that

for any application $v : \mathcal{Q}^\star \to \{-1, 1\}$ there exists an element $p \in H_T$ satisfying:

1. For all $\theta^\star \in \mathcal{Q}^\star$ and $\theta \in \mathcal{B}_T(\theta^\star, r)$, we have $|\langle \phi_T(\theta), p \rangle_T| \leq 1 - C_N\, \mathfrak{d}_T(\theta^\star, \theta)^2$.
2. For all $\theta^\star \in \mathcal{Q}^\star$ and $\theta \in \mathcal{B}_T(\theta^\star, r)$, we have $|\langle \phi_T(\theta), p \rangle_T - v(\theta^\star)| \leq C'_N\, \mathfrak{d}_T(\theta^\star, \theta)^2$.
3. For all θ in Θ_T and $\theta \notin \bigcup_{\theta^\star \in \mathcal{Q}^\star} \mathcal{B}_T(\theta^\star, r)$ (far region), we have

$$|\langle \phi_T(\theta), p \rangle_T| \leq 1 - C_F.$$

4. We have $\|p\|_T \leq C_B \sqrt{s}$.

The interpolating certificate is

$$\eta : \theta \mapsto \langle \phi_T(\theta), p \rangle_T.$$

Assume that

$$\mathfrak{d}_T(\theta, \theta') > 2r \text{ for all } \theta, \theta' \in \mathcal{Q}^\star$$

and that there exist positive constants c_N, c_F, c_B depending on r and $\mathcal{K}_\infty$, such that for any application $v : \mathcal{Q}^\star \to \{-1, 1\}$ there exists an element $q \in H_T$ satisfying:

1. For all $\theta^\star \in \mathcal{Q}^\star$ and $\theta \in \mathcal{B}_T(\theta^\star, r)$, we have:
$$|\langle \phi_T(\theta), q \rangle_T - v(\theta^\star)\ \mathrm{sign}(\theta - \theta^\star)\mathfrak{d}_T(\theta, \theta^\star)| \leq c_N\, \mathfrak{d}_T(\theta^\star, \theta)^2.$$
2. For all θ in Θ_T and $\theta \notin \bigcup_{\theta^\star \in \mathcal{Q}^\star} \mathcal{B}_T(\theta^\star, r)$ (far region), we have $|\langle \phi_T(\theta), q \rangle_T| \leq c_F$.
3. $||q||_T \leq c_B \sqrt{s}$.

Existence of Interpolating Derivative Certificate

Assume that

$$\mathfrak{d}_T(\theta,\theta') > 2r \text{ for all } \theta,\theta' \in \mathcal{Q}^\star$$

and that there exist positive constants c_N, c_F, c_B depending on r and $\mathcal{K}_\infty$, such that for any application $v : \mathcal{Q}^\star \to \{-1,1\}$ there exists an element $q \in H_T$ satisfying:

1. For all $\theta^\star \in \mathcal{Q}^\star$ and $\theta \in \mathcal{B}_T(\theta^\star, r)$, we have:

$$|\langle \phi_T(\theta), q\rangle_T - v(\theta^\star)\ \mathrm{sign}(\theta-\theta^\star)\mathfrak{d}_T(\theta,\theta^\star)| \le c_N\, \mathfrak{d}_T(\theta^\star,\theta)^2.$$

2. For all θ in Θ_T and $\theta \notin \bigcup_{\theta^\star\in\mathcal{Q}^\star} \mathcal{B}_T(\theta^\star, r)$ (far region), we have $|\langle \phi_T(\theta), q\rangle_T| \le c_F$.

3. $||q||_T \le c_B\sqrt{s}$.

The interpolating derivative certificate is

$$\theta \mapsto \langle \phi_T(\theta), q\rangle_T.$$

Assume we observe the random element y of H_T under the regression model $\beta^\star$ and $\vartheta^\star = (\theta_1^\star, \cdots, \theta_K^\star)$ a vector with entries in Θ_T, a compact interval of $\mathbb{R}$, such that:

1. **Admissible noise:** For any f in H_T, for a noise level $\sigma > 0$ and a decay rate for the noise variance $\Delta_T > 0$:
$$Var(< f, w_T >_T) \leq \sigma^2 \Delta_T \|f\|_T^2.$$
2. **Regularity of the dictionary φ_T:** The dictionary function φ_T satisfies the smoothness conditions and g_T the positivity conditions .
3. **Regularity of the limit kernel:** The kernel $\mathcal{K}_\infty$ and the functions g_∞ and h_∞, defined on an interval $\Theta_\infty \subset \Theta$ satisfy the smoothness conditions .
4. **Proximity to the limit kernel:** The kernel $\mathcal{K}_T$ is sufficiently close to the limit kernel $\mathcal{K}_\infty$.
5. **Existence of certificates:** The set of unknown parameters $\mathcal{Q}^\star = \{\theta_k^\star,\, k \in S^\star\}$, with $S^\star = \mathrm{Supp}(\beta^\star)$, satisfies Assumptions for existence of certificates with the same $r > 0$.

Then, there exist finite positive constants $\mathcal{C}_0$, $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ depending on the kernel $\mathcal{K}_\infty$ defined on Θ_∞ and on r such that for any $\tau > 0$ and a tuning parameter:

$$\kappa \geq \mathcal{C}_1 \sigma \sqrt{\Delta_T \log \tau},$$

we have the prediction error bound :

$$\left\| \hat{\beta} \Phi_T(\hat{\vartheta}) - \beta^\star \Phi_T(\vartheta^\star) \right\|_T^2 \leq \mathcal{C}_0 \, s \, \kappa^2,$$

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

Then, there exist finite positive constants $\mathcal{C}_0$, $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ depending on the kernel $\mathcal{K}_\infty$ defined on Θ_∞ and on r such that for any $\tau > 0$ and a tuning parameter:

$$\kappa \geq \mathcal{C}_1 \sigma \sqrt{\Delta_T \log \tau},$$

we have the prediction error bound :

$$\left\| \hat{\beta} \Phi_T(\hat{\vartheta}) - \beta^\star \Phi_T(\vartheta^\star) \right\|_T^2 \leq \mathcal{C}_0 \, s \, \kappa^2,$$

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

Moreover, with the same probability, the difference of the ℓ_1-norms of $\hat{\beta}$ and $\beta^\star$ is bounded by:

$$\left| \| \, \hat{\beta} \|_{\ell_1} - \| \, \beta^\star \|_{\ell_1} \right| \leq \mathcal{C}_3 \, \kappa \, s.$$

There can be no clusters of large values $\hat{\beta}_\ell$ in the neighborhood of one β_k^* which can compensate to estimate β_k^*:

$$\sum_{k \in S^\star} \left| |\beta_k^\star| - \sum_{\ell \in \tilde{S}_k(r)} |\hat{\beta}_\ell| \right| \leq \mathcal{C}_3 \, \kappa \, s, \quad \sum_{k \in S^\star} \left| \beta_k^\star - \sum_{\ell \in \tilde{S}_k(r)} \hat{\beta}_\ell \right| \leq \mathcal{C}_4 \, \kappa \, s$$

There can be no clusters of large values $\hat{\beta}_\ell$ in the neighborhood of one β_k^* which can compensate to estimate β_k^*:

$$\sum_{k\in S^\star}\left||\beta_k^\star| - \sum_{\ell\in\tilde{S}_k(r)}|\hat{\beta}_\ell|\right| \leq \mathcal{C}_3\,\kappa\,s, \quad \sum_{k\in S^\star}\left|\beta_k^\star - \sum_{\ell\in\tilde{S}_k(r)}\hat{\beta}_\ell\right| \leq \mathcal{C}_4\,\kappa\,s$$

and the estimator $\hat{\beta}_\ell$ drops to 0 when $\hat{\theta}_\ell$ is outside the $r-$neighbourhood of the true set of non-linear parameters:

$$\left\|\hat{\beta}_{\tilde{S}(r)^c}\right\|_{\ell_1} \leq \mathcal{C}_5\,\kappa\,s,$$

with the same probability.

Results - Prediction and estimation 2/2

There can be no clusters of large values $\hat{\beta}_\ell$ in the neighborhood of one β_k^* which can compensate to estimate β_k^*:

$$\sum_{k \in S^\star} \left| |\beta_k^\star| - \sum_{\ell \in \tilde{S}_k(r)} |\hat{\beta}_\ell| \right| \leq \mathcal{C}_3\, \kappa\, s, \quad \sum_{k \in S^\star} \left| \beta_k^\star - \sum_{\ell \in \tilde{S}_k(r)} \hat{\beta}_\ell \right| \leq \mathcal{C}_4\, \kappa\, s$$

and the estimator $\hat{\beta}_\ell$ drops to 0 when $\hat{\theta}_\ell$ is outside the $r-$neighbourhood of the true set of non-linear parameters:

$$\left\| \hat{\beta}_{\tilde{S}(r)^c} \right\|_{\ell_1} \leq \mathcal{C}_5\, \kappa\, s,$$

with the same probability.

Quality of estimation of the non-linear parameters:

$$\sum_{k \in S^\star} \sum_{\ell \in \tilde{S}_k(r)} \left| \hat{\beta}_\ell \right| \mathfrak{d}_T(\hat{\theta}_\ell, \theta_k^\star)^2 \leq \mathcal{C}_6\, \kappa\, s,$$

with the same probability.

We consider a very general framework including discrete and continuous models with Gaussian, possibly correlated, noise and various dictionaries of smooth functions (deconvolution, Laplace transform, ...)

Discussion

We consider a very general framework including discrete and continuous models with Gaussian, possibly correlated, noise and various dictionaries of smooth functions (deconvolution, Laplace transform, ...)

The upper bound on the prediction risk is
- nearly the same as for the linear regression in the discrete model, whp,
- free of K
- involves controls of tails of sup of linear functionals of a Gaussian process (Azaïs and Wschebor, 2009)

Discussion

We consider a very general framework including discrete and continuous models with Gaussian, possibly correlated, noise and various dictionaries of smooth functions (deconvolution, Laplace transform, ...)

The upper bound on the prediction risk is
- nearly the same as for the linear regression in the discrete model, whp,
- free of K
- involves controls of tails of sup of linear functionals of a Gaussian process (Azaïs and Wschebor, 2009)

We assumed existence of certificates! Next we construct such certificates under separation conditions of the non-linear parameters (of order s in theory, can be reduced to constant for location models)!

For the location models (deconvolution):
-the spread of the features can help to reduce the euclidean distance between the location parameters!

Starting the proof:

By definition:

$$\frac{1}{2}\|y - \hat{\beta}\Phi_T(\hat{\theta})\|_T^2 + \kappa\|\hat{\beta}\|_1 \leq \frac{1}{2}\|y - \beta^\star\Phi_T(\theta^\star)\|_T^2 + \kappa\|\beta^\star\|_1$$

gives

$$\frac{1}{2}\|\beta^\star\Phi_T(\theta^\star) - \hat{\beta}\Phi_T(\hat{\theta})\|_T^2 \leq \langle\hat{\beta}\Phi_T(\hat{\theta}) - \beta^\star\Phi_T(\theta^\star), w_T\rangle_T + \kappa(\|\beta^\star\|_1 - \|\hat{\beta}\|_1).$$

Starting the proof:

By definition:

$$\frac{1}{2}\|y-\hat{\beta}\Phi_T(\hat{\theta})\|_T^2+\kappa\|\hat{\beta}\|_1 \leq \frac{1}{2}\|y-\beta^\star\Phi_T(\theta^\star)\|_T^2+\kappa\|\beta^\star\|_1$$

gives

$$\frac{1}{2}\|\beta^\star\Phi_T(\theta^\star)-\hat{\beta}\Phi_T(\hat{\theta})\|_T^2 \leq \langle\hat{\beta}\Phi_T(\hat{\theta})-\beta^\star\Phi_T(\theta^\star),w_T\rangle_T+\kappa(\|\beta^\star\|_1-\|\hat{\beta}\|_1).$$

Decompose

$$\hat{\beta}\Phi_T(\hat{\theta})-\beta^\star\Phi_T(\theta^\star)=\sum_{k\in S^\star}\left(\sum_{\ell\in\tilde{S}_k(r)}\hat{\beta}_\ell\Phi_T(\hat{\theta}_\ell)-\beta_k^\star\Phi_T(\theta_k^\star)\right)+\sum_{\ell\in\tilde{S}^c(r)}\hat{\beta}_\ell\Phi_T(\hat{\theta}_\ell)$$

and use Taylor expansion for $\Phi_T(\hat{\theta}_\ell)$ at $\theta_k^\star$.

Starting the proof:

By definition:

$$\frac{1}{2}\|y-\hat\beta\Phi_T(\hat\theta)\|_T^2+\kappa\|\hat\beta\|_1\le\frac{1}{2}\|y-\beta^\star\Phi_T(\theta^\star)\|_T^2+\kappa\|\beta^\star\|_1$$

gives

$$\frac{1}{2}\|\beta^\star\Phi_T(\theta^\star)-\hat\beta\Phi_T(\hat\theta)\|_T^2\le\langle\hat\beta\Phi_T(\hat\theta)-\beta^\star\Phi_T(\theta^\star),w_T\rangle_T+\kappa(\|\beta^\star\|_1-\|\hat\beta\|_1).$$

Decompose

$$\hat\beta\Phi_T(\hat\theta)-\beta^\star\Phi_T(\theta^\star)=\sum_{k\in S^\star}\left(\sum_{\ell\in\tilde S_k(r)}\hat\beta_\ell\Phi_T(\hat\theta_\ell)-\beta_k^\star\Phi_T(\theta_k^\star)\right)+\sum_{\ell\in\tilde S^c(r)}\hat\beta_\ell\Phi_T(\hat\theta_\ell)$$

and use Taylor expansion for $\Phi_T(\hat\theta_\ell)$ at $\theta_k^\star$. The first term of the expansion writes

$$\sum_{k\in S^\star}\left(\sum_{\ell\in\tilde S_k(r)}\hat\beta_\ell-\beta_k^\star\right)\langle\Phi_T(\theta_k^\star),w_T\rangle_T\le\sum_{k\in S^\star}\left|\sum_{\ell\in\tilde S_k(r)}\hat\beta_\ell-\beta_k^\star\right|\cdot\sup_\theta\langle\Phi_T(\theta),w_T\rangle_T,$$

then use a certificate and probabilistic bounds, etc.

For α, ξ in $\mathbb{R}^s$, we construct the family

$$p_{\alpha,\xi} = \sum_{k=1}^{s} \alpha_k \phi_T(\theta_k^\star) + \sum_{k=1}^{s} \xi_k \, \phi_T^{[1]}(\theta_k^\star)$$

and certificates will be obtained by finding α, ξ to check the constraints. We get:

$$\eta_{\alpha,\xi}(\theta) := \langle \phi_T(\theta), p_{\alpha,\xi} \rangle_T = \sum_{k=1}^{s} \alpha_k \, \mathcal{K}_T(\theta, \theta_k^\star) + \sum_{k=1}^{s} \xi_k \, \mathcal{K}_T^{[0,1]}(\theta, \theta_k^\star).$$

3.2 Sufficient conditions for constructing the certificates

For α, ξ in $\mathbb{R}^s$, we construct the family

$$p_{\alpha,\xi} = \sum_{k=1}^{s} \alpha_k \phi_T(\theta_k^\star) + \sum_{k=1}^{s} \xi_k \, \phi_T^{[1]}(\theta_k^\star)$$

and certificates will be obtained by finding α, ξ to check the constraints. We get:

$$\eta_{\alpha,\xi}(\theta) := \langle \phi_T(\theta), p_{\alpha,\xi} \rangle_T = \sum_{k=1}^{s} \alpha_k \, \mathcal{K}_T(\theta, \theta_k^\star) + \sum_{k=1}^{s} \xi_k \, \mathcal{K}_T^{[0,1]}(\theta, \theta_k^\star).$$

Local curvature of the kernels around the diagonal is controled:

$$\varepsilon_T(r) = 1 - \sup\left\{ |\mathcal{K}_T(\theta, \theta')|; \quad \theta, \theta' \in \Theta_T \text{ such that } \mathfrak{d}_T(\theta', \theta) \geq r \right\},$$
$$\nu_T(r) = -\sup\left\{ \mathcal{K}_T^{[0,2]}(\theta, \theta'); \quad \theta, \theta' \in \Theta_T \text{ such that } \mathfrak{d}_T(\theta', \theta) \leq r \right\}.$$

In the example 'mixture of Gaussian features':

$$\varepsilon_\infty(r) = 1 - \exp(-\frac{1}{2} r^2), \quad \nu_\infty(r) = (1 - r^2) \exp(-\frac{1}{2} r^2).$$

Separation conditions

We define the set $\Theta^s_{T,\delta} \subset \Theta^s_T$ of vector of parameters of dimension $s \in \mathbb{N}^*$ and separation $\delta > 0$ as:

$$\Theta^s_{T,\delta} = \Big\{ (\theta_1, \cdots, \theta_s) \in \Theta^s_T : \ \mathfrak{d}_T(\theta_\ell, \theta_k) > \delta \text{ for all distinct } k, \ell \in \{1, \dots, s\} \Big\}.$$

and, for $u > 0$, a measure of the **decoherence of the features**:

$$\delta_T(u,s) = \inf \Big\{ \delta > 0 : \ \max_{1 \le \ell \le s} \sum_{k=1, k \neq \ell}^{s} |\mathcal{K}_T^{[i,j]}(\theta_\ell, \theta_k)| \le u \\ \text{for all } (i,j) \in \{0,1\} \times \{0,1,2\} \text{ and } (\theta_1, \cdots, \theta_s) \in \Theta^s_{T,\delta} \Big\}.$$

Existence of the interpolating certificate

Let $T \in \mathbb{N}$, $s \in \mathbb{N}^*$ and $r > 0$. We assume that:

1. **Regularity of the dictionary** φ_T;
2. **Regularity of the limit kernel** $\mathcal{K}_\infty$ and we have $r \in \left(0, 1/\sqrt{2L_{2,0}}\right)$, and also $\varepsilon_\infty(r/\rho) > 0$ and $\nu_\infty(\rho r) > 0$.
3. **Decoherence of the features:** There exists $u_\infty \in \left(0, H_\infty^{(2)}(r,\rho)\right)$ such that:

$$\delta_\infty(u_\infty, s) < +\infty.$$

4. **Closeness of the metrics** $\mathfrak{d}_T$ **and** $\mathfrak{d}_\infty$ controled by some ρ_T
5. **Proximity of the kernels** $\mathcal{K}_T$ **and** $\mathcal{K}_\infty$.

Existence of the interpolating certificate

Let $T \in \mathbb{N}$, $s \in \mathbb{N}^*$ and $r > 0$. We assume that:

1. **Regularity of the dictionary** φ_T;
2. **Regularity of the limit kernel** $\mathcal{K}_\infty$ and we have $r \in \left(0, 1/\sqrt{2L_{2,0}}\right)$, and also $\varepsilon_\infty(r/\rho) > 0$ and $\nu_\infty(\rho r) > 0$.
3. **Decoherence of the features:** There exists $u_\infty \in \left(0, H_\infty^{(2)}(r,\rho)\right)$ such that:

$$\delta_\infty(u_\infty, s) < +\infty.$$

4. **Closeness of the metrics $\mathfrak{d}_T$ and $\mathfrak{d}_\infty$** controled by some ρ_T
5. **Proximity of the kernels $\mathcal{K}_T$ and $\mathcal{K}_\infty$.**

Then, with the positive constants C_N, C'_N, $C_B = 2$ and $C_F \leq 1$ there exist an interpolating certificate (with the same r) for any subset $\mathcal{Q}^\star = \{\theta_i^\star, 1 \leq i \leq s\}$ such that for all $\theta \neq \theta' \in \mathcal{Q}^\star$:

$$\mathfrak{d}_T(\theta, \theta') > 2 \max(r, \rho_T\, \delta_\infty(u_\infty, s)).$$

Existence of the interpolating derivative certificate

Let $T \in \mathbb{N}$ and $s \in \mathbb{N}^*$. We assume that:

1. **Regularity of the dictionary φ_T**;
2. **Regularity of the limit kernel $\mathcal{K}_\infty$:**
3. **Decoherence of the features:** There exists $u'_\infty \in (0, 1/6)$, such that:

$$\delta_\infty(u'_\infty, s) < +\infty.$$

4. **Proximity of the kernels $\mathcal{K}_T$ and $\mathcal{K}_\infty$**

Let $T \in \mathbb{N}$ and $s \in \mathbb{N}^*$. We assume that:

1. **Regularity of the dictionary φ_T;**
2. **Regularity of the limit kernel $\mathcal{K}_\infty$:**
3. **Decoherence of the features:** There exists $u'_\infty \in (0, 1/6)$, such that:

$$\delta_\infty(u'_\infty, s) < +\infty.$$

4. **Proximity of the kernels $\mathcal{K}_T$ and $\mathcal{K}_\infty$**

Then, with the positive constants c_N, $c_B = 2$ and c_F
there exists an interpolating derivative certificate for any $r > 0$ and any subset $\mathcal{Q}^\star = \{\theta_i^\star, 1 \leq i \leq s\}$ such that for all $\theta \neq \theta' \in \mathcal{Q}^\star$:

$$\mathfrak{d}_T(\theta, \theta') > 2 \max(r, \rho_T \, \delta_\infty(u'_\infty, s)).$$

Next

Location families (spike deconvolution) - more explicit separation bounds free of s and decreasing when the spread of the feature decreases!

Next

Location families (spike deconvolution) - more explicit separation bounds free of s and decreasing when the spread of the feature decreases!

Group-BLasso: given a collection of signals $y_1, ..., y_n$, off-the-grid prediction by penalizing with a global matrix norm:

$$\sum_{k \in S^\star} \|\beta_{k,\cdot}\|_2 \text{ or } \sum_{k \in S^\star} \|\beta_k(\cdot)\|_p,\ p \in [1, 2]$$

Inference on the signals: clustering, outliers, etc.

Next

Location families (spike deconvolution) - more explicit separation bounds free of s and decreasing when the spread of the feature decreases!

Group-BLasso: given a collection of signals $y_1, ..., y_n$, off-the-grid prediction by penalizing with a global matrix norm:

$$\sum_{k \in S^\star} \|\beta_{k,\cdot}\|_2 \text{ or } \sum_{k \in S^\star} \|\beta_k(\cdot)\|_p,\ p \in [1,2]$$

Inference on the signals: clustering, outliers, etc.

Testing the goodness-of-fit of such a signal,
or that a new signal contains only components in the prescribed list!