



Estimation et tests pour des mélanges de composantes hilbertiennes issues d'un dictionnaire continu

Clément Hardy

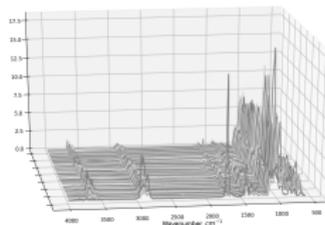
Soutenance de thèse, 16 février 2023

Motivation industrielle : spectroscopie

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	$\delta CH_2, \delta$ -CH Aliphatic
1160-1082	ν Si-O (SiO_2)
1009-909	ν Si-O (Si-OH)
825	C-Cl
664	CH Aromatic

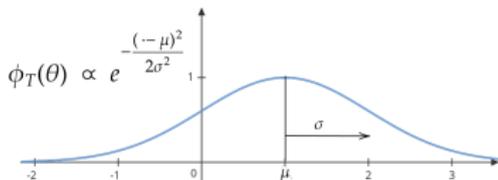
Positions des pics d'absorption et composés chimiques associés pour des échantillons de polychloroprène ([Tchalla, 2017]).

Spectres infrarouges



Dictionnaire continu

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



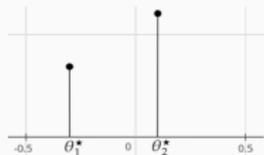
Le modèle

Un spectre $y = (y(t), t \in \mathbb{R})$ observé sur une grille t_1, \dots, t_T tel que:

$$y = \underbrace{\sum_{k=1}^s \beta_k^* \phi_T(\theta_k^*)}_{\text{mélange de pics}} + \underbrace{w}_{\text{bruit}}$$

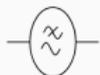
Objectif: retrouver à partir de y les paramètres β^* et $\vartheta^* = (\theta_1^*, \dots, \theta_s^*)$.

Motivation : filtre passe-bas



Sources ponctuelles

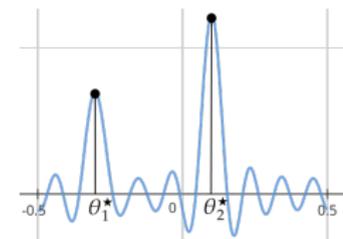
$$\mu^* = \sum_{k=1}^s \beta_k^* \delta_{\theta_k^*}$$



* Filtre passe-bas (Dirichlet)

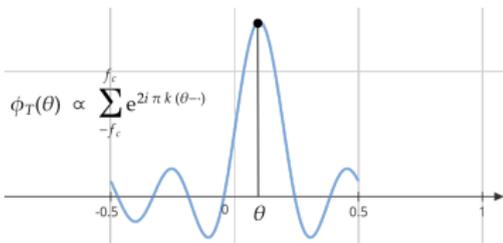
$$t \mapsto \sum_{-f_c}^{f_c} e^{2i\pi kt} = \frac{\sin(\pi(2f_c + 1)t)}{\sin(\pi t)}$$

Signal filtré



Dictionnaire continu

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



$$\phi_T(\theta) \propto \sum_{-f_c}^{f_c} e^{2i\pi k(\theta - \cdot)}$$

Le modèle

Une observation $y = (y(t), t \in \mathbb{R}/\mathbb{Z})$ telle que:

$$y = \underbrace{\sum_{k=1}^s \beta_k^* \phi_T(\theta_k^*)}_{\text{mélange de pics}} + \underbrace{w}_{\text{bruit}}$$

Objectif: retrouver à partir de y les paramètres β^* et $\vartheta^* = (\theta_1^*, \dots, \theta_s^*)$.

- **I Le modèle**
- **II Estimation**
- **III Tests**
- **IV Aspects numériques**

Le modèle

Le modèle

On observe un élément aléatoire y d'un espace de Hilbert H_T (ex: \mathbb{R}^T , $L^2(\lambda_T), \dots$) muni d'un produit scalaire $\langle \cdot, \cdot \rangle_T$ (et d'une norme $\|\cdot\|_T$).

Modèle

$$y = \underbrace{\sum_{k=1}^s \beta_k^* \phi_T(\theta_k^*)}_{\text{signal}} + \underbrace{w_T}_{\text{bruit}}.$$

Notations

- T croît avec la qualité des observations (nombre d'observations, niveau du bruit...).
- $\Theta \subset \mathbb{R}$ est l'espace des paramètres.
- $(\phi_T(\theta), \theta \in \Theta)$ sont des éléments normalisés de H_T formant un dictionnaire continu. L'application ϕ_T est continue sur Θ .
- w_T est un processus gaussien.

Le modèle : le bruit (I)

Hypothèse sur le bruit (H1)

Pour tout $f \in H_T$, la variable aléatoire $\langle f, w_T \rangle_T$ est une variable aléatoire gaussienne centrée satisfaisant:

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

Le modèle : le bruit (I)

Hypothèse sur le bruit (H1)

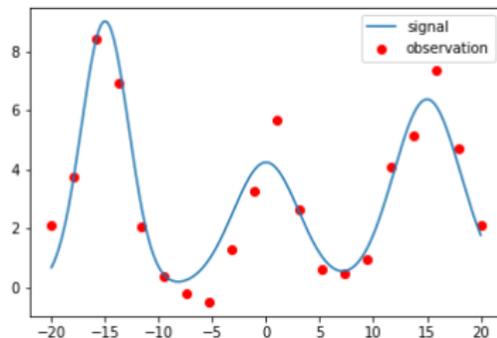
Pour tout $f \in H_T$, la variable aléatoire $\langle f, w_T \rangle_T$ est une variable aléatoire gaussienne centrée satisfaisant:

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

Ex: spectroscopie

- **Grille régulière:** $t_1 < \dots < t_T$ sur \mathbb{R} de pas de discrétisation $\Delta_T = \frac{t_T - t_1}{T}$.
- **Observations:**
 $y(t_i) = \text{signal} + w_T(t_i), 1 \leq i \leq T$.
- **Bruit:** $w_T(t_i)$ i.i.d $\sim \mathcal{N}(0, 1)$.

$$H_T = L^2(\lambda_T) \text{ où } \lambda_T(dt) = \Delta_T \sum_{j=1}^T \delta_{t_j}(dt).$$



$\Delta_T =$ pas de discrétisation .

Le modèle : le bruit (II)

Hypothèse sur le bruit (H1)

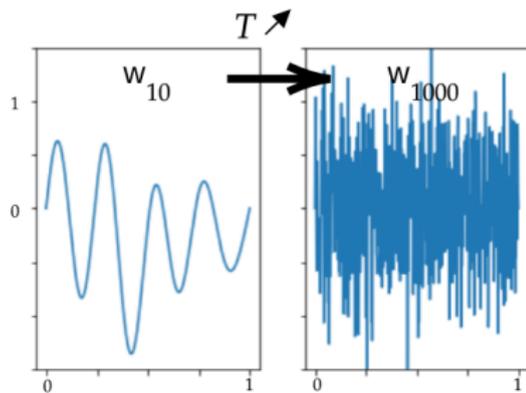
Pour tout $f \in H_T$, la variable aléatoire $\langle f, w_T \rangle_T$ est une variable aléatoire gaussienne centrée satisfaisant:

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

Ex: filtre passe-bas

- **Observations:** $(y(t), t \in \mathbb{R}/\mathbb{Z})$ t.q $y \in L^2(\text{Leb})$.
- **Bruit blanc tronqué:**
 $w_T = \frac{1}{\sqrt{T}} \sum_{k=1}^T G_k \psi_k$
- G_k i.i.d $\sim \mathcal{N}(0, 1)$,
- $(\psi_k, k \in \mathbb{N})$ b.o.n de $L^2(\text{Leb})$.

On a alors $\|w_T\|_{L^2(\text{Leb})}$ d'ordre 1.



$$\Delta_T = 1/T.$$

Estimation

Estimateurs

On construit un couple $(\hat{\beta}, \hat{\vartheta}) \in \mathbb{R}^K \times \Theta^K$ de fonctions mesurables de y “approchant” les paramètres du modèle $(\beta^* = (\beta_1^*, \dots, \beta_s^*), \vartheta^* = (\theta_1^*, \dots, \theta_s^*))$.

Estimation : risque de prédiction

Estimateurs

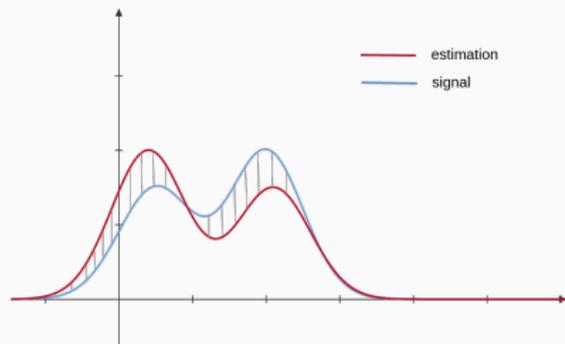
On construit un couple $(\hat{\beta}, \hat{\vartheta}) \in \mathbb{R}^K \times \Theta^K$ de fonctions mesurables de y "approchant" les paramètres du modèle $(\beta^* = (\beta_1^*, \dots, \beta_s^*), \vartheta^* = (\theta_1^*, \dots, \theta_s^*))$.

Risque de prédiction

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

où $\Phi_T(\hat{\vartheta}) \in H_T^K$ est défini par:

$$\Phi_T(\hat{\vartheta}) = (\phi_T(\hat{\theta}_1), \dots, \phi_T(\hat{\theta}_K))^T.$$



Estimation : risques d'estimation (I)

Estimateurs

On construit un couple $(\hat{\beta}, \hat{\vartheta}) \in \mathbb{R}^K \times \Theta^K$ de fonctions mesurables de y "approchant" les paramètres du modèle $(\beta^* = (\beta_1^*, \dots, \beta_s^*), \vartheta^* = (\theta_1^*, \dots, \theta_s^*))$.

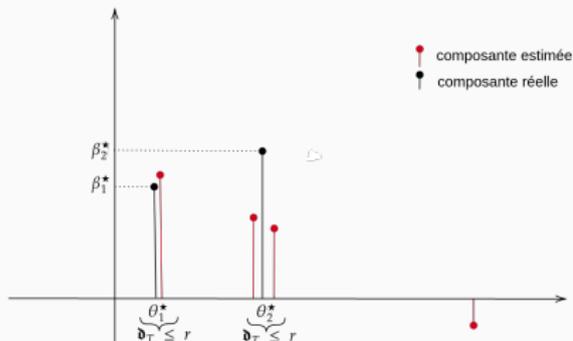
Risques d'estimation (I)

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

où l'ensemble d'indices $S(r)$ est donné par:

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

$$\text{avec } S_k(r) = \left\{ \ell, \hat{\beta}_\ell \neq 0 \text{ et } \vartheta_T(\hat{\theta}_\ell, \theta_k^*) \leq r \right\}.$$



Estimation : risques d'estimation (II)

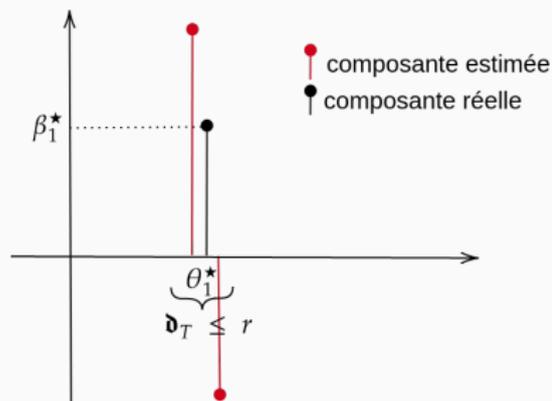
Estimateurs

On construit un couple $(\hat{\beta}, \hat{\vartheta}) \in \mathbb{R}^K \times \Theta^K$ de fonctions mesurables de y "approchant" les paramètres du modèle $(\beta^* = (\beta_1^*, \dots, \beta_s^*), \vartheta^* = (\theta_1^*, \dots, \theta_s^*))$.

Risques d'estimation (II)

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

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



Estimateurs

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

- K est une borne sur la parcimonie s .
- Θ_T est un intervalle compact.
- $\kappa > 0$ paramètre de pénalisation à régler.

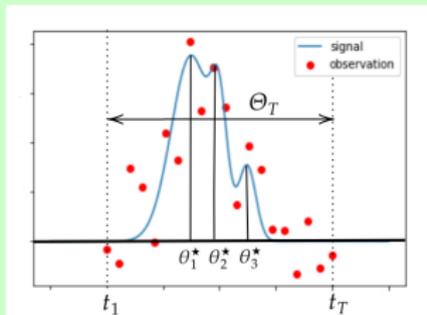
Estimation : définition des estimateurs

Estimateurs

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

- K est une borne sur la parcimonie s .
- Θ_T est un intervalle compact.
- $\kappa > 0$ paramètre de pénalisation à régler.

Ex: spectroscopie



Estimation : le Beurling Lasso (I)

Le Beurling Lasso [De Castro & Gamboa, 2012]

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

- $\mathcal{M}(\Theta_T)$ est l'ensemble des mesures de Radon sur Θ_T .
- $\langle \phi_T, \mu \rangle = \int \phi_T(\theta) \mu(d\theta)$.
- $\|\cdot\|_{TV}$ est la norme en variation totale sur les mesures.
- $\kappa > 0$ est un paramètre de pénalisation à régler.

Remarque

Posons $\mu = \sum_{k=1}^K \beta_k \delta_{\theta_k}$ (mesure atomique), on a alors:

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

Estimation : le Beurling Lasso (II)

Le Beurling Lasso [De Castro & Gamboa, 2012]

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

Estimation : le Beurling Lasso (II)

Le Beurling Lasso [De Castro & Gamboa, 2012]

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

Lien entre $\mathcal{P}_1(\kappa)$ et $\mathcal{P}_2(\kappa)$

- Existence d'une solution à $\mathcal{P}_2(\kappa)$ [Bredies & Pikkarainen, 2013].
- Si $\mathcal{P}_2(\kappa)$ admet une solution $\hat{\mu} = \sum_{k=1}^K \hat{\beta}_k \delta_{\hat{\theta}_k}$ alors $(\hat{\beta}, \hat{\vartheta})$ est solution de $\mathcal{P}_1(\kappa)$.
- Lorsque H_T est de dimension finie égale à K , il existe une solution à $\mathcal{P}_2(\kappa)$ composée d'au plus K atomes [Boyer et al, 2019].

Estimation : le Beurling Lasso (II)

Le Beurling Lasso [De Castro & Gamboa, 2012]

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

Lien entre $\mathcal{P}_1(\kappa)$ et $\mathcal{P}_2(\kappa)$

- Existence d'une solution à $\mathcal{P}_2(\kappa)$ [Bredies & Pikkarainen, 2013].
- Si $\mathcal{P}_2(\kappa)$ admet une solution $\hat{\mu} = \sum_{k=1}^K \hat{\beta}_k \delta_{\hat{\theta}_k}$ alors $(\hat{\beta}, \hat{\vartheta})$ est solution de $\mathcal{P}_1(\kappa)$.
- Lorsque H_T est de dimension finie égale à K , il existe une solution à $\mathcal{P}_2(\kappa)$ composée d'au plus K atomes [Boyer et al, 2019].

Ex : filtre passe-bas

L'espace d'observation est $H_T = L^2(\text{Leb})$.

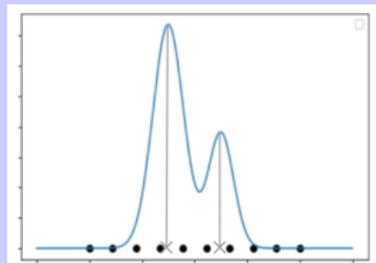
Estimation : de l'intérêt des méthodes sans grille

Estimation utilisant une grille sur l'espace des paramètres

- Discrétiser l'espace des paramètres sur une grille de K points $\vartheta^{\mathcal{G}} = (\theta_1^{\mathcal{G}}, \dots, \theta_K^{\mathcal{G}})$.
- Résoudre le problème Lasso:

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

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



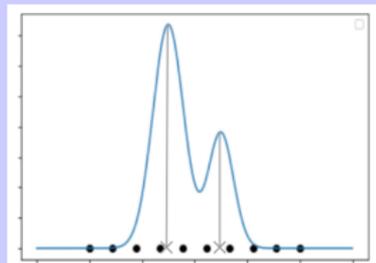
Estimation : de l'intérêt des méthodes sans grille

Estimation utilisant une grille sur l'espace des paramètres

- Discrétiser l'espace des paramètres sur une grille de K points $\vartheta^{\mathcal{G}} = (\theta_1^{\mathcal{G}}, \dots, \theta_K^{\mathcal{G}})$.
- Résoudre le problème Lasso:

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

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



Inconvénients liés au raffinement de la grille sur Θ

- Corrélation forte entre les lignes de $\Phi_{\mathcal{G}} \implies$ pb. num.
- Explosion du nombre de points nécessaires dans la grille lorsque $\Theta \subset \mathbb{R}^d$.
- Dans le modèle de pics translatés: formation de clusters de pics au voisinage des positions des pics à estimer [Duval & Peyré, 2017].

Bibliographie

- **BLasso**: [De Castro and Gamboa, 2012], [Bredies & Pikkarainen, 2013].
- **Super-résolution et acquisition comprimée**: [Candès and Fernandez-Granda, 2013, 2014] (BLasso), [Bhaskar, Tang & Recht, 2013] (*atomic norm denoising*)...
- **Risques de prédiction et d'estimation** (dictionnaire composé d'exponentielles complexes): [Tang, Bhaskar & Recht 2014], [Boyer, De Castro & Salmon 2017].
- **Recouvrement du support d'une mesure et robustesse du BLasso**: [Duval & Peyré, 2015] (modèle de déconvolution de pics avec faible bruit $\|w_T\|_T \ll 1$).
- **Cadre géométrique général du BLasso**: [Poon, Keriven & Peyré, 2021].
- **Modèle de mélange de densités**: [De Castro, Gadat, Marteau & Maugis, 2020].

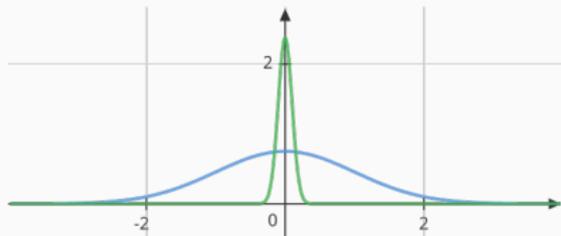
Estimation : hypothèses (I)

Hypothèses sur la régularité du dictionnaire (H2)

- $\phi_T : \Theta \rightarrow H_T$ est \mathcal{C}^3 et $\|\phi_T(\theta)\|_T = 1$ sur Θ .
- $\|\partial_\theta \phi_T(\theta)\|_T^2 > 0$ sur Θ .

Ex: spectroscopie

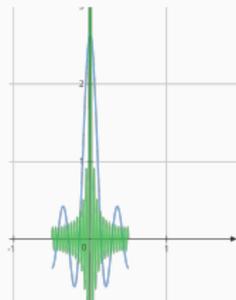
$$\phi_T(\theta) \propto e^{-(\theta - \cdot)^2 / 2\sigma_T^2}, \quad \theta \in \Theta = \mathbb{R}.$$



Ex: filtre passe-bas

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

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



Estimation : hypothèses (II)

Noyau et noyau approximant

On définit un noyau sur Θ^2 pour mesurer la corrélation entre les composantes du dictionnaire:

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

ainsi qu'un noyau approximant symétrique $\mathcal{K}^{\text{prox}}$ sur Θ_∞^2 .

Estimation : hypothèses (II)

Noyau et noyau approximant

On définit un noyau sur Θ^2 pour mesurer la corrélation entre les composantes du dictionnaire:

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

ainsi qu'un noyau approximant symétrique $\mathcal{K}^{\text{prox}}$ sur Θ_∞^2 .

Ex: spectroscopie

$$\Delta_T \rightarrow 0 \text{ et } \sigma_T = \text{cst}, \quad \mathcal{K}^{\text{prox}} = \lim_{T \rightarrow +\infty} \mathcal{K}_T.$$

Estimation : hypothèses (II)

Noyau et noyau approximant

On définit un noyau sur Θ^2 pour mesurer la corrélation entre les composantes du dictionnaire:

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

ainsi qu'un noyau approximant symétrique $\mathcal{K}^{\text{prox}}$ sur Θ_∞^2 .

Ex: spectroscopie

$$\Delta_T \rightarrow 0 \text{ et } \sigma_T = \text{cst}, \quad \mathcal{K}^{\text{prox}} = \lim_{T \rightarrow +\infty} \mathcal{K}_T.$$

Hypothèses sur la régularité du noyau approximant (H3)

- Le noyau $\mathcal{K}^{\text{prox}}$ est $\mathcal{C}^{3,3}$ avec des dérivées bornées + autres régularités.
- Le noyau $\mathcal{K}^{\text{prox}}$ est **localement concave sur la diagonale** et strictement inférieur à 1 en dehors.

Métrique de Fisher sur l'espace des paramètres

$$d_{\mathcal{K}}(\theta, \theta') = \inf_{\gamma} \int_0^1 |\dot{\gamma}_s| \sqrt{\partial_{x,y} \mathcal{K}(\gamma_s, \gamma_s)} ds$$

inf. sur l'ensemble des chemins réguliers $\gamma : [0, 1] \rightarrow \Theta$ tels que $\gamma_0 = \theta$ et $\gamma_1 = \theta'$.

→ **invariance** $d_{\mathcal{K}_\phi}(\theta, \theta') = d_{\mathcal{K}_{\phi \circ h}}(h^{-1}(\theta), h^{-1}(\theta'))$.

Estimation : hypothèses (III)

Métrique de Fisher sur l'espace des paramètres

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

inf. sur l'ensemble des chemins réguliers $\gamma : [0, 1] \rightarrow \Theta$ tels que $\gamma_0 = \theta$ et $\gamma_1 = \theta'$.

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

Exemples unidimensionnels

- Modèles de pics translétés (spectroscopie / filtre passe-bas):

$$\mathfrak{d}_{\mathcal{K}_T}(\theta, \theta') \sim \frac{|\theta - \theta'|}{\sigma_T} \text{ (métrique euclidienne).}$$

- Modèles d'échelle: $H = L^2(\text{Leb})$ et $\phi(\theta) \propto e^{-\theta}$ avec $\Theta = \mathbb{R}_+^*$ et

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

On a $\mathfrak{d}_{\mathcal{K}_T}(\theta, \theta + \varepsilon) \xrightarrow{\varepsilon \rightarrow 0} +\infty$ (\neq métrique euclidienne).

Estimation : hypothèses (IV)

Proximité de \mathcal{K}_T et $\mathcal{K}^{\text{prox}}$

- Proximité entre les noyaux:

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

- Equivalence de $\mathfrak{d}_{\mathcal{K}_T}$ et $\mathfrak{d}_{\mathcal{K}^{\text{prox}}}$: $\mathfrak{d}_{\mathcal{K}^{\text{prox}}}/\rho_T \leq \mathfrak{d}_{\mathcal{K}_T} \leq \rho_T \mathfrak{d}_{\mathcal{K}^{\text{prox}}}$.

Estimation : hypothèses (IV)

Proximité de \mathcal{K}_T et $\mathcal{K}^{\text{prox}}$

- Proximité entre les noyaux:

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

- Equivalence de $\partial_{\mathcal{K}_T}$ et $\partial_{\mathcal{K}^{\text{prox}}}$: $\partial_{\mathcal{K}^{\text{prox}}} / \rho_T \leq \partial_{\mathcal{K}_T} \leq \rho_T \partial_{\mathcal{K}^{\text{prox}}}$.

Hypothèses de proximité entre \mathcal{K}_T et $\mathcal{K}^{\text{prox}}$ (H4)

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

Estimation : hypothèses (IV)

Proximité de \mathcal{K}_T et $\mathcal{K}^{\text{prox}}$

- Proximité entre les noyaux:

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

- Equivalence de $\partial_{\mathcal{K}_T}$ et $\partial_{\mathcal{K}^{\text{prox}}}$: $\partial_{\mathcal{K}^{\text{prox}}} / \rho_T \leq \partial_{\mathcal{K}_T} \leq \rho_T \partial_{\mathcal{K}^{\text{prox}}}$.

Hypothèses de proximité entre \mathcal{K}_T et $\mathcal{K}^{\text{prox}}$ (H4)

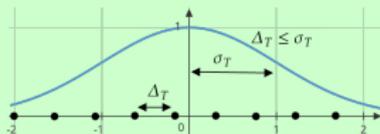
$$s \mathcal{V}_T \leq C \quad \text{et} \quad \rho_T \leq \rho.$$

Ex: spectroscopie

T : nombre de points de discrétisation.

Δ_T : pas de discrétisation.

σ_T : paramètre d'échelle.



$$\frac{\Delta_T}{\sigma_T} \rightarrow 0 \quad \text{et} \quad \frac{\Delta_T T}{\sigma_T} \rightarrow +\infty \implies \mathcal{V}_T \rightarrow 0 \quad \text{et} \quad \rho_T \rightarrow 1.$$

Théorème 1

On observe $y \in H_T$ de paramètres inconnus $\beta^* \in \mathbb{R}^s$ et $\vartheta^* = (\theta_1^*, \dots, \theta_s^*) \in \Theta_T^s$ avec $s \leq K$ tel que les hypothèses H1-H4 sont vérifiées et pour tout $\ell \neq k$,

$$d_{\mathcal{K}_T}(\theta_\ell^*, \theta_k^*) \gtrsim \delta(s) \text{ (séparation)}.$$

Alors, on a pour les estimateurs $\hat{\beta}$ et $\hat{\vartheta}$ définis par $\mathcal{P}_1(\kappa)$ avec

$$\kappa \geq C_1 \sqrt{\Delta_T \log \tau} \quad \text{et} \quad \tau > 1,$$

les bornes suivantes sur les risques de prédiction et d'estimation:

$$\left\| \beta^* \Phi_T(\vartheta^*) - \hat{\beta} \Phi_T(\hat{\vartheta}) \right\|_T \leq C_0 \sqrt{s} \kappa,$$

$$\sum_{k=1}^s \left| |\beta_k^*| - \sum_{\ell \in S_k(r)} |\hat{\beta}_\ell| \right| + \sum_{k=1}^s \left| \beta_k^* - \sum_{\ell \in S_k(r)} \hat{\beta}_\ell \right| + \left\| \hat{\beta}_{S(r)^c} \right\|_{\ell_1} \leq C_0 \kappa s.$$

avec probabilité au moins: $1 - C_2 \left(\frac{|\Theta_T| d_T}{\tau \sqrt{\log \tau}} \vee \frac{1}{\tau} \right)$.

Remarque: les bornes ne dépendent pas de K !

Condition de séparation

$$d_{\mathcal{K}_T}(\theta_\ell^*, \theta_k^*) \gtrsim \delta(s), \ell \neq k.$$

Estimation : séparation entre les paramètres

Condition de séparation

$$\partial_{\mathcal{K}_T}(\theta_\ell^*, \theta_k^*) \gtrsim \delta(s), \ell \neq k.$$

Définition de la séparation $\delta(s)$

$$\delta(s) = \inf \left\{ \delta > 0 : \max_{1 \leq \ell \leq s} \sum_{k=1, k \neq \ell}^s |\mathcal{K}^{\text{prox}[i,j]}(\theta_\ell, \theta_k)| \leq u \quad \text{pour tout} \right. \\ \left. (i, j) \in \{0, 1\} \times \{0, 1, 2\} \text{ et } (\theta_1, \dots, \theta_s) \in \Theta^s \text{ t.q. } \partial_{\mathcal{K}^{\text{prox}}}(\theta_\ell, \theta_k) > \delta, \ell \neq k \right\}.$$

Estimation : séparation entre les paramètres

Condition de séparation

$$\partial_{\mathcal{K}_T}(\theta_\ell^*, \theta_k^*) \gtrsim \delta(s), \ell \neq k.$$

Définition de la séparation $\delta(s)$

$$\delta(s) = \inf \left\{ \delta > 0 : \max_{1 \leq \ell \leq s} \sum_{k=1, k \neq \ell}^s |\mathcal{K}^{\text{prox}[i,j]}(\theta_\ell, \theta_k)| \leq u \text{ pour tout} \right. \\ \left. (i,j) \in \{0,1\} \times \{0,1,2\} \text{ et } (\theta_1, \dots, \theta_s) \in \Theta^s \text{ t.q. } \partial_{\mathcal{K}^{\text{prox}}}(\theta_\ell, \theta_k) > \delta, \ell \neq k \right\}.$$

Ex: spectroscopie

$$\partial_{\mathcal{K}_T}(\theta, \theta') \sim \frac{|\theta - \theta'|}{\sigma_T} \text{ et } \delta(s) \lesssim 1/u.$$

Estimation : borne sur le risque de prédiction (spectroscopie)

Ex: spectroscopie

- **Grille:** $t_1 < \dots < t_T$ régulière sur \mathbb{R} de pas Δ_T .
- **Bruit:** $w_T(t_j)$ i.i.d $\sim \mathcal{N}(0, 1)$.
- **Dictionnaire:** $(\phi_T(\theta) \propto e^{-(\theta - \cdot)^2 / 2\sigma_T^2}, \theta \in \Theta = \mathbb{R})$.
- **Séparation:** $|\theta_\ell^* - \theta_k^*| \gtrsim \sigma_T$ pour $\ell \neq k$.

On a pour $T \gtrsim s(t_T - t_1)\sigma_T^{-1}$:

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

avec proba. au moins $1 - C \left(\frac{t_T - t_1}{\sigma_T T \sqrt{\log(T)}} \vee \frac{1}{T} \right)$.

Estimation : borne sur le risque de prédiction (spectroscopie)

Ex: spectroscopie

- **Grille:** $t_1 < \dots < t_T$ régulière sur \mathbb{R} de pas Δ_T .
- **Bruit:** $w_T(t_j)$ i.i.d $\sim \mathcal{N}(0, 1)$.
- **Dictionnaire:** $(\phi_T(\theta) \propto e^{-(\theta - \cdot)^2 / 2\sigma_T^2}, \theta \in \Theta = \mathbb{R})$.
- **Séparation:** $|\theta_\ell^* - \theta_k^*| \gtrsim \sigma_T$ pour $\ell \neq k$.

On a pour $T \gtrsim s(t_T - t_1)\sigma_T^{-1}$:

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

avec proba. au moins $1 - C \left(\frac{t_T - t_1}{\sigma_T T \sqrt{\log(T)}} \vee \frac{1}{T} \right)$.

→ La borne sur le risque de prédiction obtenue est de l'ordre de celle obtenue pour l'estimateur Lasso dans le cadre du modèle de régression linéaire de grande dimension.

Ex: filtre passe-bas

- **Bruit blanc tronqué:** $w_T = \frac{1}{\sqrt{T}} \sum_{k=1}^T G_k \psi_k$.
- **Dictionnaire:** $(\phi_T(\theta) \propto \frac{\sin(T\pi(\theta-\cdot))}{\sin(\pi(\theta-\cdot))}, \theta \in \Theta = \mathbb{R}/\mathbb{Z})$.
- **Séparation:** $|\theta_\ell^* - \theta_k^*| \gtrsim \delta(s)/T$ pour $\ell \neq k$.

On a pour $T \gtrsim s$:

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

avec proba. au moins $1 - \frac{C}{T\sqrt{\log(T)}}$.

Modèle (plusieurs signaux)

On observe n signaux $(Y(i) \in H_T, 1 \leq i \leq n)$ dont l'union des composantes est de cardinal s :

$$Y(i) = \sum_{k=1}^s B_k^*(i) \phi_T(\theta_k^*) + W_T(i) \quad \text{pour } 1 \leq i \leq n,$$

où les coefficients linéaires sont obtenus par l'application $B^* : i \mapsto B^*(i) = (B_1^*(i), \dots, B_s^*(i))$.

$$W_T(i) \text{ i.i.d } \sim w_T.$$

Estimation : structure commune entre plusieurs signaux

Modèle (plusieurs signaux)

On observe n signaux ($Y(i) \in H_T$, $1 \leq i \leq n$) dont l'union des composantes est de cardinal s :

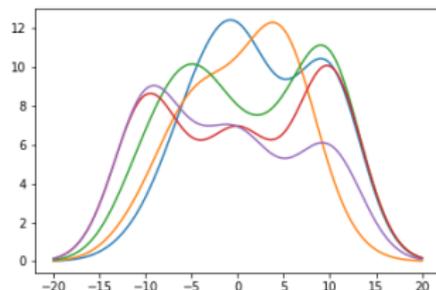
$$Y(i) = \sum_{k=1}^s B_k^*(i) \phi_T(\theta_k^*) + W_T(i) \quad \text{pour } 1 \leq i \leq n,$$

où les coefficients linéaires sont obtenus par l'application $B^* : i \mapsto B^*(i) = (B_1^*(i), \dots, B_s^*(i))$.

$$W_T(i) \text{ i.i.d } \sim w_T.$$

Question

Reconstruction simultanée v.s individuelle: peut-on tirer profit de la structure commune des signaux pour accélérer leur reconstruction à partir des observations ?



Estimateurs

$$(\hat{B}, \hat{\vartheta}) \in \underset{B \in \mathbb{R}^{n \times K}, \vartheta \in \Theta_T^K}{\operatorname{argmin}} \frac{1}{2n} \sum_{i=1}^n \|Y(i) - B(i)\Phi_T(\vartheta)\|_T^2 + \kappa \sum_{k=1}^K \sqrt{\sum_{i=1}^n B_k(i)^2}$$

Bibliographie

- Group-Lasso pour les modèles linéaires [Yuan & Lin, 2006], résultats d'optimalité (au sens minimax) [Lounici, Pontil, van de Geer & Tsybakov, 2011].
- Group-BLasso [Golbabaev & Poon, 2022].

Estimation : reconstruction simultanée v.s individuelle

Théorème 2

Sous les hypothèses H1-H4 et avec pour tout $\ell \neq k$,

$$\vartheta_T(\theta_\ell^*, \theta_k^*) \gtrsim \delta(s) \text{ (séparation)}.$$

Alors, pour $\tau > 1$ et un certain choix de κ , on a la **borne suivante sur le risque de prédiction**:

$$\frac{1}{n} \sum_{i=1}^n \left\| B^*(i) \Phi_T(\vartheta^*) - \hat{B}(i) \Phi_T(\hat{\vartheta}) \right\|_T^2 \lesssim s \Delta_T \left(1 + \frac{\log(\tau)}{n} \right),$$

avec probabilité au moins: $1 - \mathcal{C}_2 \left(\frac{1}{\tau^2 \log(\tau)} + \frac{|\Theta_T| \vartheta_T e^{-n/3}}{\tau \sqrt{\log(\tau)}} \right)$.

Estimation : reconstruction simultanée v.s individuelle

Théorème 2

Sous les hypothèses H1-H4 et avec pour tout $\ell \neq k$,

$$\vartheta_T(\theta_\ell^*, \theta_k^*) \gtrsim \delta(s) \text{ (séparation)}.$$

Alors, pour $\tau > 1$ et un certain choix de κ , on a la **borne suivante sur le risque de prédiction**:

$$\frac{1}{n} \sum_{i=1}^n \left\| B^*(i) \Phi_T(\vartheta^*) - \hat{B}(i) \Phi_T(\hat{\vartheta}) \right\|_T^2 \lesssim s \Delta_T \left(1 + \frac{\log(\tau)}{n} \right),$$

avec probabilité au moins: $1 - C_2 \left(\frac{1}{\tau^2 \log(\tau)} + \frac{|\Theta_T| \vartheta_T e^{-n/3}}{\tau \sqrt{\log(\tau)}} \right)$.

Reconstruction	simultanée	individuelle
Borne	$s \Delta_T \left(1 + \frac{\log(\tau)}{n} \right)$	$s \Delta_T \log(\tau)$
Proba.	$1 - C \left(\frac{1}{\tau^2 \log(\tau)} + \frac{ \Theta_T \vartheta_T e^{-n/3}}{\tau \sqrt{\log(\tau)}} \right)$	$1 - C' \left(\frac{n}{\tau} + \frac{ \Theta_T \vartheta_T n}{\tau \sqrt{\log(\tau)}} \right)$

Estimation : reconstruction simultanée v.s individuelle

Théorème 2

Sous les hypothèses H1-H4 et avec pour tout $\ell \neq k$,

$$\partial_T(\theta_\ell^*, \theta_k^*) \gtrsim \delta(s) \text{ (séparation)}.$$

Alors, pour $\tau > 1$ et un certain choix de κ , on a la **borne suivante sur le risque de prédiction**:

$$\frac{1}{n} \sum_{i=1}^n \left\| B^*(i) \Phi_T(\vartheta^*) - \hat{B}(i) \Phi_T(\hat{\vartheta}) \right\|_T^2 \lesssim s \Delta_T \left(1 + \frac{\log(\tau)}{n} \right),$$

avec probabilité au moins: $1 - C_2 \left(\frac{1}{\tau^2 \log(\tau)} + \frac{|\Theta_T| \partial_T e^{-n/3}}{\tau \sqrt{\log(\tau)}} \right)$.

Reconstruction	simultanée	individuelle
Borne	$s \Delta_T \left(1 + \frac{\log(\tau)}{n} \right)$	$s \Delta_T \log(\tau)$
Proba.	$1 - C \left(\frac{1}{\tau^2 \log(\tau)} + \frac{ \Theta_T \partial_T e^{-n/3}}{\tau \sqrt{\log(\tau)}} \right)$	$1 - C' \left(\frac{n}{\tau} + \frac{ \Theta_T \partial_T n}{\tau \sqrt{\log(\tau)}} \right)$

→ La borne sur le risque de prédiction obtenue est de l'ordre de celle obtenue pour l'estimateur Group-Lasso dans le cadre du modèle de régression linéaire de grande dimension.

Tests

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

Test d'adéquation à un modèle:

Soient $(\beta^0, \vartheta^0) \in (\mathbb{R}^*)^{s^0} \times \Theta_T^{s^0}$ (**connus**) et $(\beta^*, \vartheta^*) \in (\mathbb{R}^*)^s \times \Theta_T^s$ (**inconnus**).

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

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

Test d'adéquation à un modèle:

Soient $(\beta^0, \vartheta^0) \in (\mathbb{R}^*)^{s^0} \times \Theta_T^{s^0}$ (**connus**) et $(\beta^*, \vartheta^*) \in (\mathbb{R}^*)^s \times \Theta_T^s$ (**inconnus**).

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

Risque et séparation minimale

Un test Ψ est une fonction mesurable de y à valeurs dans $\{0, 1\}$:

$\Psi = 0$ **acceptation** de H_0 et $\Psi = 1$ **rejet** de H_0 .

- **Le risque de test maximal:**

$$R_\rho(\Psi) = \underbrace{\sup_{(\beta^*, \vartheta^*) \in H_0} \mathbb{E}_{(\beta^*, \vartheta^*)}[\Psi]}_{\text{Rejeter } H_0 \text{ à tort}} + \underbrace{\sup_{(\beta^*, \vartheta^*) \in H_1(\rho)} \mathbb{E}_{(\beta^*, \vartheta^*)}[1 - \Psi]}_{\text{Accepter } H_0 \text{ à tort}}.$$

- **La séparation minimax du problème de test pour $\alpha \in (0, 1)$:**

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

Cadre

- $H_T = L^2(\lambda_T)$ avec $\lambda_T \rightarrow \text{Leb}(\mathbb{R} \text{ ou } \mathbb{R}/\mathbb{Z})$.
- Dictionnaire : $(\phi_T(\theta) \propto h(\theta - \cdot, \sigma_T), \theta \in \Theta)$.

Ex: spectroscopie / filtre passe-bas .

Cadre

- $H_T = L^2(\lambda_T)$ avec $\lambda_T \rightarrow \text{Leb}(\mathbb{R} \text{ ou } \mathbb{R}/\mathbb{Z})$.
- Dictionnaire : $(\phi_T(\theta) \propto h(\theta - \cdot, \sigma_T), \theta \in \Theta)$.

Ex: spectroscopie / filtre passe-bas .

Proposition 1

Sous les hypothèses H1-H4 (pour $s \in \mathbb{N}$ et $s^0 \in \mathbb{N}$) et si:

$$|\theta_\ell^* - \theta_k^*| \gtrsim \sigma_T \delta(s) \quad \ell \neq k \text{ et } |\theta_\ell^0 - \theta_k^0| \gtrsim \sigma_T \delta(s^0) \quad \ell \neq k.$$

On a pour $|\Theta_T|/\sigma_T \geq 1$ et $\alpha \in (0, 1)$:

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

Corollaire 1

Processus discret sur une grille régulière de T points sur le tore \mathbb{R}/\mathbb{Z} avec $w_T(t_j)$ i.i.d $\sim \mathcal{N}(0, 1)$ pour $1 \leq j \leq T$.

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

La borne devient:

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

Corollaire 1

Processus discret sur une grille régulière de T points sur le tore \mathbb{R}/\mathbb{Z} avec $w_T(t_j)$ i.i.d $\sim \mathcal{N}(0, 1)$ pour $1 \leq j \leq T$.

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

La borne devient:

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

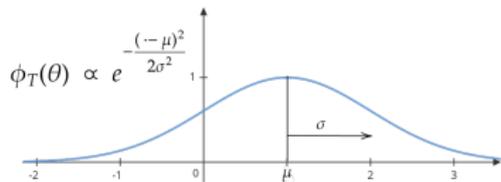
→ On peut en déduire une borne sup. sur la séparation minimax asymptotique (*i.e* $s, T \rightarrow \infty$) \approx borne inf. sur la séparation minimax asymptotique pour le modèle de régression linéaire en grande dimension [Ingster, Tsybakov & Verzelen, 2010].

Aspects numériques

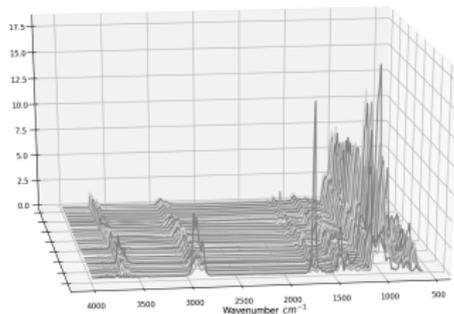
Aspects numériques : application à la spectroscopie IR

Dictionnaire continu

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



Spectres infrarouges



Bibliographie

Blaso \rightarrow problème convexe sur un espace de dimension infinie.

- Modifications de l'algorithme de Frank-Wolfe [Boyd, Schiebinger & Recht, 2017], [Denoyelle, Duval, Peyré & Soubies 2020], [Globabae & Poon, 2022].

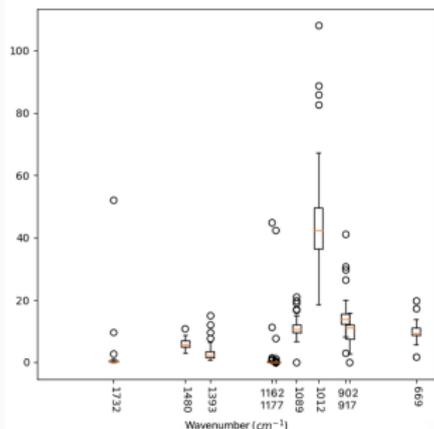
Résolution via sliding Frank-Wolfe

$$(\hat{B}, \hat{\vartheta}) \in \underset{B \in \mathbb{R}^{n \times K}, \vartheta \in \Theta_T^K}{\operatorname{argmin}} \frac{1}{2n} \sum_{i=1}^n \|Y(i) - B(i)\Phi_T(\vartheta)\|_T^2 + \kappa \sum_{k=1}^K \sqrt{\sum_{i=1}^n B_k(i)^2}$$

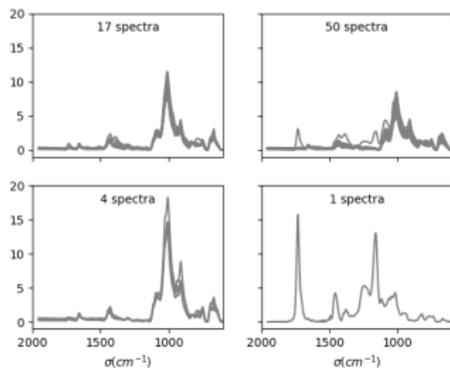
Aspects numériques : application à la spectroscopie IR

Résolution via sliding Frank-Wolfe

$$(\hat{B}, \hat{\vartheta}) \in \underset{B \in \mathbb{R}^{n \times K}, \vartheta \in \Theta_T^K}{\operatorname{argmin}} \frac{1}{2n} \sum_{i=1}^n \|Y(i) - B(i)\Phi_T(\vartheta)\|_T^2 + \kappa \sum_{k=1}^K \sqrt{\sum_{i=1}^n B_k(i)^2}$$



Dispersion des amplitudes associées aux 10 principaux pics d'absorption déterminés pour les spectres de polychloroprène vieilli en milieu marin.



Regroupement des spectres de polychloroprène vieilli en milieu marin par niveaux d'usage.

Estimation et tests

Sous des conditions de séparation entre les paramètres non linéaires à estimer:

- Risques de prédiction de l'ordre de ceux connus lorsque ϑ^* est donné.
- Reconstruction simultanée $>$ reconstruction individuelle lorsque les signaux partagent une structure commune.
- Intensité minimale pour la détection d'un signal de l'ordre de celle nécessaire pour le modèle de régression linéaire en grande dimension.

Perspectives

- $\Theta \subseteq \mathbb{R}^d$
- Tester $H_0 : \mathcal{Q}^* = \{\theta_1^*, \dots, \theta_s^*\} \subseteq \mathcal{Q}^0$ sans contraintes sur les signes des entrées de β^* .
- Améliorer les conditions de séparation.