Régression sur variétés pénalisée topologiquement

Olympio Hacquard [1]

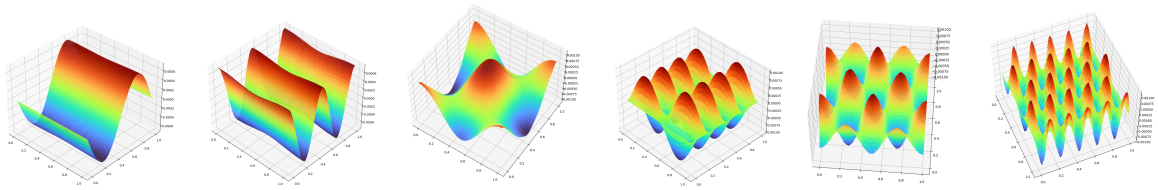
[1] *Laboratoire mathématique d'Orsay, 307 rue Michel Magat, 91400 Orsay, olympio.hacquard@universite-paris-saclay.fr*

Résumé. Nous nous intéressons à un problème de régression où les données sont supportées sur une variété compacte. Afin de tirer profit de la nature géométrique sous-jacente du problème, on effectue la régression sur la base constituée des fonctions propres de l'opérateur de Laplace-Beltrami sur la variété, régularisée par une pénalité topologique. Les pénalités proposées sont construites sur la topologie des sous-ensemble de niveaux soit de ces fonctions propres, soit de la fonction à reconstruire. Nous allons voir que cette approche permet d'offrir une bonne performance en terme de reconstruction à la fois sur des exemples simulés et réels. De plus, nous proposons des résultats théoriques sur l'estimateur de la fonction de régression garantissant à la fois une reconstruction fidèle et une certaine régularité topologique, permettant ainsi de valider notre méthode dans le cas où la fonction à reconstruire est suffisamment lisse d'un point de vue topologique.

Mots-clés. Régression sur variété, Apprentissage statistique, Laplacien de graphe, Persistance topologique.

Abstract. We study a regression problem on a compact manifold. In order to take advantage of the underlying geometry and topology of the data, the regression task is performed on the basis of the first several eigenfunctions of the Laplace-Beltrami operator of the manifold, that are regularized with topological penalties. The proposed penalties are based on the topology of the sub-level sets of either the eigenfunctions or the estimated function. The overall approach is shown to yield promising and competitive performance on various applications to both synthetic and real data sets. We also provide theoretical guarantees on the regression function estimates, on both its prediction error and its smoothness (in a topological sense). Taken together, these results support the relevance of our approach in the case where the targeted function is "topologically smooth".

Keywords. Manifold regression, Statistical learning, Graph Laplacian, Topological persistence.



(a) EF 1 (b) EF 2 (c) EF 10 (d) EF 27 (e) EF 54 (f) EF 75

Figure 1: Quelques fonctions propres du Laplacien du graphe aux 8 plus proches voisins pour 10000 points sur le tore.

1 Modèle statistique

On observe un $n-$échantillon $(X_i, Y_i)_{i=1}^n$ où les X_i vivent sur une sous-variété de $\mathbb{R}^D$ compacte sans bord $\mathcal{M}$ et où les Y_i sont des étiquettes réelles. On suppose que les données sont simulées de sorte que pour tout i,

$$Y_i = f^\star(X_i) + \varepsilon_i$$

où ε est un bruit sous-gaussien indépendant sur chaque entrée et l'on cherche à estimer la fonction $f^\star$. Pour ce faire, on introduit la base $(\Phi_i)_{i=1}^p$ des fonctions propres de l'opérateur de Laplace-Beltrami (voir Zelditch (2003)) ou éventuellement une approximation via les vecteurs propres du Laplacien discret d'un graphe construit sur les données. On peut voir quelques uns de ces vecteurs propres en Figure 1.

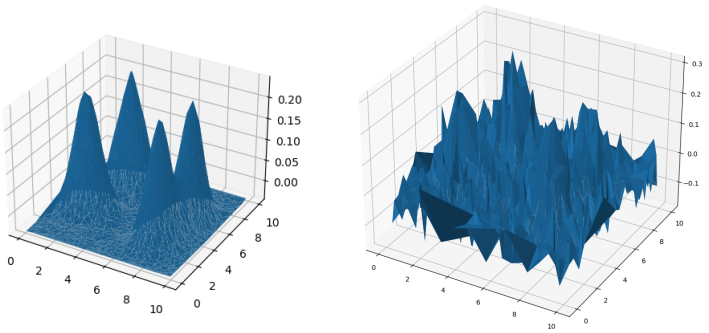
Une fois le nombre p de fonctions propres choisi, le problème revient à trouver $\theta \in \mathbb{R}^p$ tel que $\sum_{i=1}^p \theta_i \Phi_i$ est une bonne approximation de f^*. Ainsi, on introduit la matrice de design $\mathbf{X} \in \mathbb{R}^{n \times p}$, où $\mathbf{X}_{ij} = \Phi_j(X_i)$, et on cherche un minimiseur de

$$\mathcal{L}(\theta) = \|Y - \mathbf{X}\theta\|_2^2 + \mu\Omega(\theta),$$

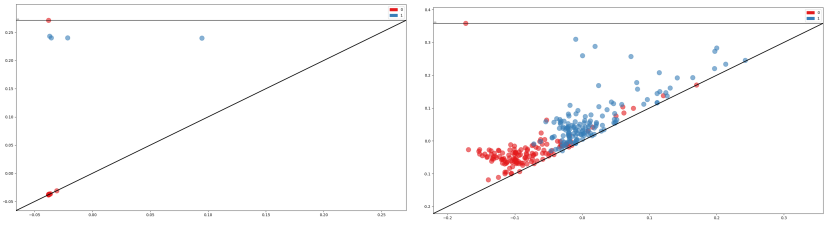
où Ω est un terme de pénalité visant à promouvoir une bonne généralisation (voir Massart (2007)).

2 Pénalités topologiques

Inspiré de travaux récents de Chen & al. (2019) et Carriere & al. (2020), on cherche à introduire un terme de régularisation topologique. Les pénalités proposées sont basées sur la notion de persistance topologique d'une fonction(voir Boissonnat, Chazal et Yvinec (2018)). Lorsque les sous-ensembles de niveau d'une fonction varient de $-\infty$ à $+\infty$, leur topologie change, et plus précisément des composantes homologiques naissent et meurent. La persistance χ est définie comme la somme des temps de vie de chaque composante topologique.



(a) Fonction originale (b) Fonction bruitée



(c) Fonction originale (d) Fonction bruitée

Figure 2: Influence du bruit sur les diagrammes de persistance

Une représentation usuelle est celle des diagrammes de persistance qui est un multi-ensemble de points où pour chacun est représenté en abscisse le temps où la composante topologique est née et en ordonnée le temps où celle-ci est morte. Lorsque l'on bruite une fonction, de nombreux points sont ajoutés près de la diagonale, et le paradigme classique est de considérer que les points éloignées de la diagonale correspondent à de vraies composantes topologiques de la fonction tandis que ceux près de la diagonale correspondent à du bruit (voir figure 2).

La première pénalitée considérée est :

$$\Omega_1(\theta) = \sum_{i=1}^{p} |\theta_i| \chi(\Phi_i).$$

Cette pénalité est convexe en θ et pénalise chaque fonction propre individuellement, utilisant les propriétés de sélection du LASSO afin d'éliminer les fonctions propres qui oscillent trop et ont donc une persistance trop élevée.

La seconde pénalité est :

$$\Omega_2 = \chi\left(\sum_{i=1}^{p} \theta_i \Phi_i\right).$$

Celle-ci est non-convexe et a pour objectif de pénaliser directement la géométrie de la fonction de régression que l'on cherche à estimer, dans le but de la lisser et de réduire le bruit.

3 Illustration expérimentale

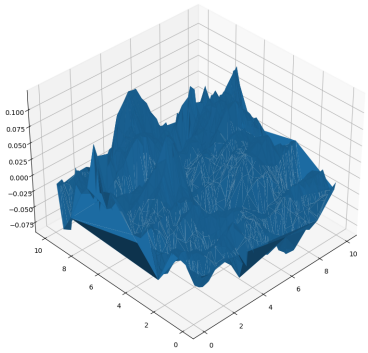
Les deux pénalités ont été essayées sur de nombreuses données à la fois réelles et synthétiques et ont été comparées à des pénalités plus usuelles (Lasso ou variation totale), ainsi qu'à des méthodes à noyau standard. L'approche qui a conduit à la meilleure reconstruction consiste à minimiser la fonction de perte $\mathcal{L}$ avec la pénalité Ω_1 qui va servir de modèle de sélection de variable et ne va garder que quelques fonctions propres de basse persistance, puis à minimiser à nouveau la fonction de perte sur ce sous-ensemble des fonctions propres avec la pénalité Ω_2. On a ainsi une reconstruction souvent meilleure que les méthodes usuelles, notamment lorsque les données sont fortement bruitées ou que l'on cherche à apprendre le modèle sur un faible nombre de données. On peut voir en figure 3 le résultat de cette méthode appliquée à la reconstruction de la somme de quatre Gaussiennes et comparée à un Lasso sur les vecteurs propres du Laplacien de graphe. On voit que la fonction reconstruite avec une pénalité topologique est très lisse et que le nombre de pics est fidèlement reconstruits. Cela se traduit sur les diagrammes par un faible nombre de points de persistance faible et quatre points très persistants correspondant à chaque mode de chaque Gaussienne. Par opposition, la fonction reconstruite par un Lasso est encore assez bruitée et il y a de nombreux points de faible persistance dans le diagramme. De plus, seulement deux ou trois modes peuvent être identifiés.

4 Discussion théorique

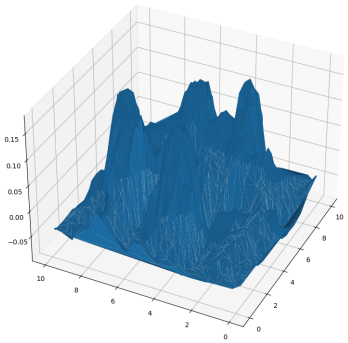
L'étude théorique de la pénalité Ω_1 est très simple puisqu'il s'agit d'un Lasso pondéré. Toutes les propriétés connues du Lasso (voir Buhlmann & Van de Geer (2011) pour un traitement exhaustif) découlent alors. Dans le cas de la pénalité non-convexe Ω_2, on a le résultat suivant :

Théorème 1 *Supposons* $f^\star = \sum_{j=1}^{p} \theta_j^\star \Phi_j$. *Alors, le minimum* $\hat{\theta}$ *pour la pénalité* Ω_2 *vérifie, avec une probabilité supérieure à* $1 - 3e^{-x} - \exp\left(\frac{-0.1n}{C(\mathcal{M})p} + \ln(p)\right)$ *:*

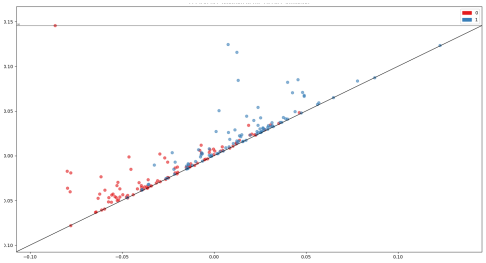
$$\|\theta^\star - \hat{\theta}\|^2 \leq 32\left[\frac{p\sigma^2}{n} + \frac{C(\mathcal{M})p\sqrt{2x}}{n^{3/2}}\right](1 + C_0\sqrt{x})^2 + 8C(\mathcal{M})^2 p(2\nu(f^\star) + \zeta)^2\mu^2.$$



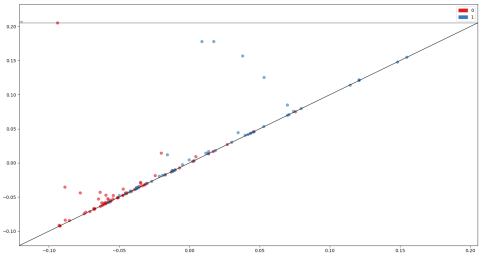
(a) Function estimated by a Lasso



(b) Function estimated by the topological penalty Ω_2



(c) Persistence diagram of the Lasso estimate



(d) Persistence diagram of the Ω_2 penalty estimate

Figure 3: Reconstruction of the sum of four Gaussians.

où C_0 est une constante universelle, $C_{\mathcal{M}}$ une constante dépendant uniquement de $\mathcal{M}$ et de sa métrique, et $\nu(f^\star)$ est le nombre de points dans le diagramme de persistance de $f^\star$. De plus, sous les mêmes hypothèses :

$$\chi(\hat{f}) \leq \chi(f^\star) + \frac{32}{\mu}\left[\frac{p\sigma^2}{n} + \frac{C(\mathcal{M})p\sqrt{2x}}{n^{3/2}}\right](1 + C_0\sqrt{x})^2.$$

Ce résultat est vérifié avec grande probabilité si le nombre d'échantillons est au moins d'ordre $p\ln(p)$. On a alors une vitesse de convergence d'ordre p/n ce qui est standard sans hypothèse supplémentaire de parcimonie. On voit également apparaître un terme de biais permettant de calibrer le paramètre μ. La deuxième partie du théorème assure la consistance topologique de la fonction reconstruite $\hat{f}$.

Bibliographie

Pascal Massart (2007) Concentration inequalities and model selection.
Mathieu Carriere, Frederic Chazal, Marc Glisse, Yuichi Ike, and Hariprasad Kannan (2020). A note on stochastic subgradient descent for persistence-based functionals: convergence and practical aspects *Arxiv preprint*.
Chao Chen, Xiuyan Ni, Qinxun Bai, and Yusu Wang. A topological regularizer for classifiers via persistent homology (2019). *22nd International Conference on Artificial Intelligence and Statistics* pp.2573–2582.
Jean-Daniel Boissonnat, Frederic Chazal, and Mariette Yvine (2018). Geometric and topological inference. *Cambridge University Press*
Peter Buhlmann and Sara Van De Geer (2011). Statistics for high-dimensional data: methods, theory and applications. *Springer Science & Business Media*.
Steve Zelditch (2017). Eigenfunctions of the Laplacian on a Riemannian manifold, *volume 125 of CBMS Regional Conference Series in Mathematics. American Mathematical Soc.*