Filtrage adaptatif de signaux définis sur des graphes de grande taille

Elie Chedemail, Basile de Loynes, Fabien Navarro, Baptiste Olivier

Filtrage adaptatif de signaux définis sur des graphes de grande taille

Elie Chedemail [1,2] & Basile de Loynes[2] & Fabien Navarro[2] & Baptiste Olivier[1]

[1] *Orange Labs, France*
{*elie.chedemail, baptiste.olivier*}*@orange.com*
[2] *ENSAI, Campus de Ker Lann, Rue Blaise Pascal, BP 37203, 35172 Bruz*
{*basile.deloynes, fabien.navarro*}*@ensai.fr*

Résumé. L'analyse de signaux sur les graphes s'attache à étendre la théorie (analyse de Fourier) et les méthodologies (filtrage notamment) du champ classique à des signaux définis sur les noeuds d'un graphe. De plus en plus populaire de par la flexibilité de la structure de graphe, ce champ de recherche trouve de nombreux domaines d'applications (réseaux de télécommunications, réseaux sociaux, chimie organique, neurologie, apprentissage profond). L'approche mise en oeuvre consiste à appliquer une procédure de seuillage dans un domaine transformé bien choisi dans lequel une représentation parcimonieuse du signal est présumée. La calibration des seuils est obtenue en minimisant l'estimateur sans biais du risque dû originellement à Stein et adapté à la transformation choisie. Le coeur de cette communication est de proposer une approche permettant le passage à l'échelle des grands graphes à l'aide d'une approximation par polynômes de Chebyshev du calcul fonctionnel. La mise en oeuvre de cette approche est illustrée numériquement sur un grand graphe (dépassant le million de noeuds).

Mots-clés. Estimateur de Stein sans biais du risque, traitement du signal sur les graphes, estimation par Monte Carlo

Abstract. Graph Signal Analysis focuses on extending the theory (Fourier analysis) and methodologies (such as filtering) of the classical field to signals defined on the vertices of a graph. Increasingly popular because of the flexibility of the underlying structure, this research area can be applied in many context such as telecommunications networks, social networks, organic chemistry, neurology or deep learning. The approach implemented in the sequel consists in applying a thresholding procedure in a well-chosen transformed domain in which the signal is presumed sparsely represented. The threshold calibration is obtained by minimizing the unbiased estimator of the risk originally due to Stein and adapted to the chosen transformation. The core of this paper is to propose an approach that scales to large graphs using a Chebyshev polynomial approximation of the functional computation. The implementation of this approach is illustrated numerically on a large graph (exceeding one million nodes).

Keywords. Stein Unbiased Risk Estimation, Graph Signal Processing, Monte Carlo estimation

1 Introduction

Les données acquises à partir de systèmes interactifs à grande échelle, tels que les réseaux informatiques, écologiques, sociaux, financiers ou biologiques, sont de plus en plus répandues et accessibles. Au sein de l'apprentissage statistique moderne, la représentation, le traitement ou l'analyse efficace de ces données structurées à grande échelle, telles que les graphes ou les réseaux, sont quelques-uns des problèmes clés (*cf.* Nickel et al. (2015); Bronstein et al. (2017)). Le domaine émergent du traitement des signaux sur graphes met en évidence les liens entre les domaines que sont le traitement du signal et de la théorie spectrale des graphes (par exemple Shuman et al. (2013); Ortega et al. (2018) pour l'illustration de telles interactions). Le rapide développement de cette thématique est illustré par la revue récente Dong et al. (2020) dans laquelle sont en outre évoquées les perspectives de cette thématique ainsi que son rôle dans certaines des premières conceptions des architectures de réseaux neuronaux sur graphes (GNN).

Basé sur une analogie entre le laplacien d'un graphe et l'opérateur de Laplace-Beltrami, une notion de transformée en ondelettes dans le contexte des graphes a été proposées dans Hammond et al. (2011). Cette construction est par ailleurs étroitement liée à celle plus générale proposée dans Coifman and Maggioni (2006) qui s'applique notamment au cas des variétés différentiables. Une construction légèrement différente proposée dans Göbel et al. (2018) permet de définir une transformée multi-échelle semi-orthogonale de sorte que l'énergie du signal est préservée dans le domaine transformé, ouvrant ainsi la porte à des méthodes de débruitages numériquement efficaces. Plus récemment, dans de Loynes et al. (2021), une méthode de sélection automatique du paramètre de seuillage a été introduite en adaptant l'estimateur sans biais du risque de Stein (SURE) à ce type de transformées semi-orthogonales. Par construction, une telle approche nécessite la diagonalisation du laplacien ce qui est rédhibitoire en grande dimension. Les difficultés à lever se situent à la fois au moment de la transformation et lors de la phase d'optimisation du SURE dont le terme de divergence fait apparaître des poids caractéristiques de la transformée. La première difficulté a déjà été traitée dans Hammond et al. (2011) en proposant une transformée rapide à base d'approximations par polynômes de Chebyshev. Le deuxième écueil, quant à lui, sera résolu par Monte Carlo en profitant de l'interprétation en terme de covariance des poids. La faisabilité d'une telle stratégie est illustrée par une comparaison numérique avec le DFS Fussed Lasso introduit dans Padilla et al. (2017). Les résultats numériques corroborent ceux de de Loynes et al. (2021) sur la pertinence de l'analyse multi-échelle face aux méthodes de pénalisations.

2 Débruitage de signaux sur un graphe

Dans toute la suite, on considère $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ un graphe non orienté composé d'un ensemble $\mathcal{V}$ de n sommets, d'un ensemble $\mathcal{E}$ d'arêtes pondérées et d'une matrice de poids W. La matrice laplacienne est donnée par $\mathcal{L} = D - W$ avec D la matrice diagonale

des degrés ($D_{i,i} = \sum_{j\in\mathcal{V}} W_{i,j}$). Par ailleurs, on considère un signal $f \in \mathbb{R}^n$ défini sur les sommets ainsi que le modèle additif de débruitage suivant :

$$\tilde{f} = f + \xi,$$

où $\xi \sim \mathcal{N}(0, \sigma^2)$. Le but du débruitage est de construire un estimateur $\hat{f}$ de f qui ne dépend que des observations $\tilde{f}$.

Une façon de construire un estimateur non linéaire est de considérer une procédure de seuillage dans un espace transformé défini à l'aide d'une *frame*, c'est-à-dire une famille $\mathfrak{F} = \{r_i\}_{i\in I}$ de vecteurs de $\mathbb{R}^n$ telle qu'il existe $A, B > 0$ satisfaisant pour tous les $f \in \mathbb{R}^n$

$$A\|f\|_2^2 \leq \sum_{i\in I} |\langle f, r_i\rangle|^2 \leq B\|f\|_2^2.$$

Une *frame* est dite *ajustée* (ou *étroite*) si $A = B$.

La matrice $\mathcal{L}$ est symétrique, sa décomposition spectrale est donnée par $\mathcal{L} = \sum_\ell \lambda_\ell \langle \chi_\ell, \cdot\rangle \chi_\ell$, où $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n = 0$ désignent les valeurs propres (ordonnées) de la matrice $\mathcal{L}$, et $(\chi_\ell)_{1\leq\ell\leq n}$ sont les vecteurs propres associés. Ainsi, pour toute fonction $\rho : \mathrm{sp}(\mathcal{L}) \to \mathbb{R}$ définie sur le spectre de $\mathcal{L}$, on a $\rho(\mathcal{L}) = \sum_\ell \rho(\lambda_\ell)\langle \chi_\ell, \cdot\rangle \chi_\ell$.

Les valeurs propres λ_ℓ, $\ell = 1, \ldots, n$ s'interprètent comme les fréquences fondamentales du graphe et $\rho(\mathcal{L})$ comme un opérateur de filtrage en termes d'analyse du signal.

La *frame* ajustée considérée ici est construite à partir d'une partition de l'unité finie $(\psi_j)_{j=0,\ldots,J}$ du compact $[0, \lambda_1]$ définie comme suit : soit $\omega : \mathbb{R}^+ \to [0, 1]$ une fonction quelconque à support dans $[0, 1]$, satisfaisant $\omega \equiv 1$ sur $[0, b^{-1}]$, pour $b > 1$, et

$$\psi_0(x) = \omega(x), \quad \psi_j(x) = \omega(b^{-j}x) - \omega(b^{-j+1}x) \ \text{ pour } \ j = 1, \ldots, J, \ \text{ où } \ J = \left\lfloor \frac{\log \lambda_1}{\log b} \right\rfloor + 2.$$

Il est montré dans Göbel et al. (2018) que l'ensemble de vecteurs suivants est une *frame* ajustée :

$$\mathfrak{F} = \left\{ \sqrt{\psi_j}(\mathcal{L})\delta_i, j = 0, \ldots, J, i = 1, \ldots, n \right\}.$$

La transformée en ondelettes d'un signal $f \in \mathbb{R}^n$ est donnée par

$$\mathcal{W}f = \left(\sqrt{\psi_0}(\mathcal{L})f^T, \ldots, \sqrt{\psi_J}(\mathcal{L})f^T \right)^T \in \mathbb{R}^{n(J+1)},$$

et sa transformée inverse $\mathcal{W}^*$ est donnée par

$$\mathcal{W}^* \left(\eta_0^T, \eta_1^T, \ldots, \eta_J^T\right)^T = \sum_{j\geq 0} \sqrt{\psi_j}(\mathcal{L})\eta_j.$$

Étant donné le laplacien et une *frame* donnée, le débruitage dans ce contexte peut être résumé comme suit :

- Analyse : calculer la transformée $\mathcal{W}\tilde{f}$;
- Seuillage : appliquer un opérateur de seuillage donné (par ex. *dur* ou *doux*) aux coefficients $\mathcal{W}\tilde{f}$;
- Synthèse : appliquer la transformée inverse pour obtenir une estimation $\hat{f}$ de f.

Les performances dépendent fortement du choix du paramètre de seuillage. Le SURE a été étendue pour cette *frame* afin de sélectionner un seuil optimal en minimisant ce critère dans de Loynes et al. (2021). Le principal inconvénient de cette approche est qu'elle nécessite le calcul de la diagonalisation complète du laplacien si bien que son application se limite aux graphes de taille modérée. Les difficultés à lever se situent à la fois au moment de la transformation et lors de la phase d'optimisation du SURE dont le terme de divergence fait apparaître des poids caractéristiques de la transformée. La première difficulté a déjà été traitée dans Hammond et al. (2011) en proposant une transformée rapide à base d'approximations par polynômes de Chebyshev. Le deuxième écueil, quant à lui, sera résolu par Monte Carlo en profitant de l'interprétation en terme de covariance des poids. Ces points font l'objet de la section suivante.

3 SURE adapté aux grands graphes

Les polynômes de Chebyshev de première espèce $T_k(x)$ d'ordre k peuvent être calculés par la relation de récurrence $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$, avec $T_0(x) = 1$ et $T_1(x) = x$. Ces polynômes forment une base orthogonale de l'espace de Hilbert $\mathbb{L}^2([-1,1], dy/\sqrt{1-y^2})$. Une ondelette (ou filtre) ρ peut donc être approchée par le développement tronqué d'ordre $K-1$,

$$\rho_K(\mathcal{L}) = \sum_{i=0}^{K-1} \theta_i(\rho) T_i(\mathcal{L}),$$

où $\theta_i(\rho)$ est le coefficient de Chebyshev associé à $T_i(\mathcal{L})$, le i-ème polynôme de Chebyshev évalué en $\tilde{\mathcal{L}} = 2\mathcal{L}/\lambda_1 - I_n$. En suivant Hammond et al. (2011), pour tout filtre ρ défini sur $\mathrm{sp}(\mathcal{L})$ et tout signal f sur le graphe $\mathcal{G}$, l'approximation $\rho_K(\mathcal{L})f$ fournie est proche de $\rho(\mathcal{L})f$ avec une complexité en temps $O(|\mathcal{E}|K)$ raisonnable lorsque la matrice $\mathcal{L}$ est creuse.

D'après de Loynes et al. (2021), le SURE pour un processus général de seuillage $h : \mathbb{R}^{n(J+1)} \to \mathbb{R}^{n(J+1)}$ est donné par l'identité suivante

$$\mathbf{SURE}(h) = -n\sigma^2 + \|h(\widetilde{F}) - \widetilde{F}\|^2 + 2\sum_{i,j=1}^{n(J+1)} \gamma_{i,j}^2 \partial_j h_i(\widetilde{F}), \tag{1}$$

où $\gamma_{i,j}^2 = (\mathcal{W}\mathcal{W}^*)_{i,j}$. Dans de Loynes et al. (2021), les poids $\gamma_{i,j}^2$, $i,j = 1,\ldots,n(J+1)$, sont calculés à partir de la réduction complète de la matrice laplacienne qui n'est plus réalisable

pour les grands graphes. En outre, il est clair d'après l'interprétation probabiliste donnée dans de Loynes et al. (2021) que

$$\forall i, j = 1, \ldots, n(J+1), \quad \gamma_{i,j}^2 = \mathbb{E}[(\mathcal{W}\varepsilon)_i (\mathcal{W}\varepsilon)_j]$$

où $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ sont n variables aléatoires indépendantes et identiquement distribuées (*i.i.d.*) de moyenne nulle et de variance égale à un. Ainsi, en tirant parti de cette identité, les poids peuvent être estimés comme suit

- Générer N variables aléatoires *i.i.d.* $(\varepsilon_{i,k})_{i=1,\ldots,n,k=1,\ldots,N}$ telles que $\mathbb{E}[\varepsilon_{i,k}] = 0$ et $\mathbb{V}(\varepsilon_{i,k}) = 1$;
- Calculer

$$\hat{\gamma}_{i,j}^2 = \frac{1}{N} \sum_{k=1}^{N} \left(\sum_{p=1}^{n} \mathcal{W}_{i,p} \varepsilon_{p,k} \right) \left(\sum_{p=1}^{n} \mathcal{W}_{j,p} \varepsilon_{p,k} \right).$$

4 Simulations

Nous comparons notre méthode à celle du *DFS Fused Lasso* introduite dans Padilla et al. (2017) et Rudin et al. (1992) par des simulations numériques réalisées avec les packages R **igraph** (Csardi and Nepusz 2006), **genlasso** (Arnold and Tibshirani 2020) et **gasper** (de Loynes et al. 2020). Ce dernier fourni une interface à la *SuiteSparse Matrix Collection*[1] (Davis and Hu 2011). Ici, nous utilisons le graphe des routes de l'état de Pennsylvanie[2] tiré de Leskovec et al. (2009) composé de 1088092 nœuds et de 1541898 arêtes. Sur celui-ci, deux classes de signaux synthétiques sont générées en s'inspirant de la méthodologie introduite dans Behjat et al. (2016). Selon deux paramètres $p \in (0, 1)$ et $k \in \mathbb{N}$, on produit un signal $f_{p,k} = W^k x_p / \lambda_1^k$ où W est la matrice de poids, x_p la réalisation d'une variable aléatoire suivant une loi de Bernoulli de paramètre p et λ_1 la plus grande valeur propre de $\mathcal{L}$. Pour ces simulations, nous avons généré deux signaux avec les paramètres $p = 0.001$, $k = 4$ et $p = 0.01$, $k = 10$ respectivement.

Table 1: SNR moyen calculé sur 10 réalisations pour chaque niveau de bruit.

	$f_{0.001,4}$				$f_{0.01,10}$			
SNR_{in}	0.61	6.63	12.65	18.67	2.09	8.12	14.13	20.16
MSE_{LD}	12.28	17.41	21.37	23.53	9.19	13.33	16.51	18.01
SURE_{LD}	12.28	17.41	21.37	23.53	9.18	13.33	16.51	18.01
MSE_{DFS}	7.92	13.10	18.41	24.27	6.64	11.12	15.98	18.46

Les performances sont comparées en termes de rapport signal sur bruit (SNR) pour différents niveaux de bruits et pour chaque signal synthétique $f_{p,k}$. L'analyse multi-échelle présente globalement des résultats très prometteurs en comparaison à *DFS Fused Lasso*.

[1] `https://sparse.tamu.edu/`

[2] `https://sparse.tamu.edu/SNAP/roadNet-PA`

Bibliographie

Arnold, T. B. and R. J. Tibshirani (2020). *R package genlasso: Path algorithm for generalized lasso problems*. Version 1.5.

Behjat, H., U. Richter, D. Van De Ville, and L. Sörnmo (2016). Signal-adapted tight frames on graphs. *IEEE Transactions on Signal Processing 64*(22), 6017–6029.

Bronstein, M. M., J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst (2017). Geometric deep learning: going beyond euclidean data. *IEEE Signal Process. Mag. 34*(4), 18–42.

Coifman, R. R. and M. Maggioni (2006). Diffusion wavelets. *Applied and Computational Harmonic Analysis 21*(1), 53 – 94. Special Issue: Diffusion Maps and Wavelets.

Csardi, G. and T. Nepusz (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, 1695.

Davis, T. A. and Y. Hu (2011). The university of florida sparse matrix collection. *ACM Transactions on Mathematical Software (TOMS) 38*(1), 1.

de Loynes, B., F. Navarro, and B. Olivier (2020). Gasper: Graph signal processing in r.

de Loynes, B., F. Navarro, and B. Olivier (2021). Data-driven thresholding in denoising with Spectral Graph Wavelet Transform. *J. Comput. Appl. Math. 389*, 113319.

Dong, X., D. Thanou, L. Toni, M. Bronstein, and P. Frossard (2020). Graph signal processing for machine learning: A review and new perspectives. *IEEE Signal Process. Mag. 37*(6), 117–127.

Göbel, F., G. Blanchard, and U. von Luxburg (2018). Construction of tight frames on graphs and application to denoising. In *Handbook of big data analytics*, Springer Handb. Comput. Stat., pp. 503–522. Springer, Cham.

Hammond, D. K., P. Vandergheynst, and R. Gribonval (2011). Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis 30*(2), 129–150.

Leskovec, J., K. J. Lang, A. Dasgupta, and M. W. Mahoney (2009). Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics 6*(1), 29–123.

Nickel, M., K. Murphy, V. Tresp, and E. Gabrilovich (2015). A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE 104*(1), 11–33.

Ortega, A., P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst (2018). Graph signal processing: Overview, challenges, and applications. *Proceedings of the IEEE 106*(5), 808–828.

Padilla, O. H. M., J. Sharpnack, J. G. Scott, and R. J. Tibshirani (2017). The dfs fused lasso: Linear-time denoising over general graphs. *J. Mach. Learn. Res. 18*, 176–1.

Rudin, L. I., S. Osher, and E. Fatemi (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena 60*(1), 259–268.

Shuman, D., S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Process. Mag. 3*(30), 83–98.