

Dictionnaires de localité-régularité pour les signaux sur graphe

Quentin LABORDE¹, Antoine MAZARGUIL¹, Laurent OUDRE¹

¹Université Paris Saclay, Université Paris Cité, ENS Paris Saclay, CNRS, SSA, INSERM, Centre Borelli,
F-91190, Gif-sur-Yvette, France

quentin.laborde@ens-paris-saclay.fr, antoine.mazarguil@ens-paris-saclay.fr,
laurent.oudre@ens-paris-saclay.fr

Résumé – Dans ce travail, nous tirons parti des principes d’incertitude afin d’introduire une nouvelle méthode d’apprentissage de dictionnaire pour les signaux sur graphe. Notre méthode prend en compte deux classes distinctes d’atomes, les atomes locaux spatialement sur le graphe et les atomes spectralement réguliers introduits avec des fonctions de pénalisation adaptées. En particulier, la prise en compte de la notion d’atomes dits de localité permet de modéliser des phénomènes locaux et interprétables.

Abstract – In this work, we leverage recent graph uncertainty principles to introduce a new dictionary learning method for graph signals. Our method considers two distinct classes of atoms, local atoms on the graph, and local atoms spectrally together with well suited penalty functions. Notably, the consideration of the notion of localized atoms on graphs allows to model local and interpretable phenomena.

1 Introduction

L’apprentissage par dictionnaire est un outil essentiel pour estimer une représentation linéaire des données, c’est-à-dire pour apprendre un sous-espace linéaire défini par une famille de vecteurs de base - appelés *atomes* — qui forment un dictionnaire. Alors que les méthodes classiques d’apprentissage linéaire se concentrent principalement sur les données vectorielles, l’étude de signaux sur des domaines topologiquement compliqués nécessite l’introduction d’un graphe comme outil flexible de représentation des données, adapté à la modélisation de la structure sous-jacente des signaux. L’étude de modèles de factorisation applicables à de telles données a donc motivé le développement de modèles d’apprentissage de dictionnaire pour les signaux sur graphe.

Dans ce travail, nous abordons la question de l’apprentissage de tels dictionnaires à partir de signaux sur graphe en considérant deux classes distinctes d’atomes : (i) les atomes spectralement réguliers — caractérisant des phénomènes continus de la structure sous-jacente du graphe — et (ii) les atomes spatialement locaux. La prise en compte des atomes locaux est en effet d’un grand intérêt pour l’analyse des séries temporelles multivariées, permettant de décomposer les signaux étudiés en une combinaison d’éléments de base localisés et donc plus facilement interprétables.

2 Etat de l’art

2.1 Signal sur graphe

Soit $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ un graphe symétrique pondéré et non orienté, composé d’un ensemble de noeuds $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$,

d’un ensemble de N arêtes $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$, ainsi que d’une matrice de poids $\mathbf{W} \in \mathbb{R}_+^{N \times N}$. Le graphe est de plus supposé connexe. Un signal réel sur le graphe \mathcal{G} est une fonction $x : \mathcal{V} \rightarrow \mathbb{R}$, représenté par un vecteur $\mathbf{x} \in \mathbb{R}^N$.

Nous introduisons également la matrice Laplacienne du graphe \mathcal{G} , définie comme $\mathbf{L} = \mathbf{D} - \mathbf{W}$, avec \mathbf{D} la matrice des degrés de \mathcal{G} et \mathbf{W} sa matrice de poids. La matrice Laplacienne permet classiquement de quantifier la *régularité* d’un signal sur graphe, définie comme suit

$$S_{\mathcal{G}}(\mathbf{x}) := \sum_{(i,j) \in \mathcal{E}} w_{i,j} (x_i - x_j)^2 = \mathbf{x}^T \mathbf{L} \mathbf{x} \quad (1)$$

En outre, deux notions de localisation du signal peuvent être introduites : la localisation dans le domaine spatial — à quel point les intensités du signal sont concentrées — et dans le domaine fréquentiel. Etant donné un nœud central $c \in \mathcal{V}$ et une distance entre chaque paire de nœuds $d^{\mathcal{G}} : |\mathcal{V}| \times |\mathcal{V}| \rightarrow \mathbb{R}_+$, l’étalement spatial [1] d’un signal \mathbf{x} autour de c est défini comme

$$\Delta_{loc}^2(\mathbf{x}, c) = \frac{1}{\|\mathbf{x}\|_2^2} \sum_{j=1}^N d^{\mathcal{G}}(c, v_j)^2 x_j^2 = \frac{1}{\|\mathbf{x}\|_2^2} \mathbf{x}^T \mathbf{P}_c \mathbf{x}, \quad (2)$$

avec $\mathbf{P}_c = \text{diag}(d^{\mathcal{G}}(c, v_1), \dots, d^{\mathcal{G}}(c, v_N))$. De même, l’étalement spectral d’un signal sur graphe peut être quantifié comme suit

$$\Delta_{sm}^2(\mathbf{x}) = \frac{1}{\|\mathbf{x}\|_2^2} \mathbf{x}^T \mathbf{L} \mathbf{x} \quad (3)$$

Compte tenu des notions d’étalement définies ci-dessus, la métrique d’incertitude sur graphe introduite par Agaskar [1] restreint l’ensemble des paires $(\Delta_{loc}^2(\mathbf{x}, c), \Delta_{sm}^2(\mathbf{x}))$ réalisables, ce qui implique qu’un signal ne peut pas être arbitrairement localisé simultanément dans les domaines spatial et spectral. Formellement, étant donné $\alpha > 0$, l’inégalité suivante s’applique

à tout signal non nul $\mathbf{x} \in \mathbb{R}^N$

$$\Delta_{loc}^2(\mathbf{x}, c) + \alpha \Delta_{sm}^2(\mathbf{x}) \geq q(\alpha) \quad (4)$$

avec $q(\alpha) > 0$ la plus petite valeur propre de la matrice $\mathbf{P} + \alpha \mathbf{L}$.

2.2 Apprentissage de dictionnaire sur graphe

Une *décomposition* d'un signal sur graphe \mathbf{x} par un dictionnaire est un vecteur $\mathbf{a} \in \mathbb{R}^M$, tel que

$$\mathbf{x} \approx \sum_{m=1}^M a_m \mathbf{d}^{(m)}. \quad (5)$$

Le dictionnaire \mathbf{D} est ainsi défini comme $(\mathbf{d}^{(1)}, \dots, \mathbf{d}^{(M)})$ l'ensemble de M vecteurs de \mathbb{R}^N — communément appelés *atomes* — tandis que le vecteur $\mathbf{a} = (a_1, \dots, a_M)$ est le *vecteur des activations* du signal \mathbf{x} . Étant donné $\mathbf{X} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(L)})$ un ensemble de L signaux de $\mathbb{R}^{N \times L}$, l'apprentissage d'un dictionnaire de taille M revient à considérer le problème d'optimisation suivant

$$\mathbf{D}^*, \mathbf{A}^* = \underset{\mathbf{D}, \mathbf{A}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 + f(\mathbf{D}) + g(\mathbf{A}) \quad (6)$$

avec $\|\cdot\|_F$ la norme de Frobenius et $f(\mathbf{D}), g(\mathbf{A})$ des pénalisations imposées au dictionnaire ainsi qu'à la matrice des activations appris.

Le problème d'optimisation (6) n'est généralement pas convexe et la recherche d'une solution locale s'effectue par le biais d'une procédure de minimisation alternée :

- Étape (AU) d'*actualisation des activations* : étant donné un dictionnaire \mathbf{D} , il s'agit d'apprendre la matrice des activations, telle que

$$\mathbf{A}^* = \underset{\mathbf{A}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 + g(\mathbf{A}) \quad (7)$$

- Étape (DU) d'*actualisation du dictionnaire* : étant donné une matrice des activations \mathbf{A} , il s'agit d'apprendre le dictionnaire optimal, tel que

$$\mathbf{D}^* = \underset{\mathbf{D}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 + f(\mathbf{D}) \quad (8)$$

2.3 Pénalités et contraintes

De multiples termes de pénalisation g et f ont été introduits dans la littérature. Nous présentons brièvement les principaux.

Pénalisation du problème AU. La sparsité de \mathbf{A} est généralement recherchée. Cependant, les approches qui imposent strictement une norme L_0 ne sont pas convexes, rendant le problème NP-difficile. Une solution courante [4] consiste à recourir à la norme L_1 , qui tend à produire des activations sparses tout en simplifiant la recherche d'une solution, telle que

$$g^{(L_1)}(\mathbf{A}) = \alpha_p \|\mathbf{A}\|_1 \quad (9)$$

avec $\alpha_p > 0$ un paramètre de pénalisation.

Une seconde approche, introduite par Ramamurthy [5], consiste à intégrer une connaissance *a priori* de la structure de graphe lors de la mise à jour de la matrice des activations. Ainsi, il s'agit de considérer un second graphe $\mathcal{G}^x = (\mathcal{V}^x, \mathcal{E}^x, \mathbf{W}^x)$ dont chaque nœud porte un signal $\mathbf{x}^{(l)}$ de l'ensemble \mathbf{X} , tandis que les poids $\mathbf{W}_{i,j}^x$ fournissent une mesure de similarité entre chaque paire de signaux $\mathbf{x}^{(i)}$ et $\mathbf{x}^{(j)}$. La fonction de pénalisation associée à cette approche est formulée comme suit

$$\begin{aligned} g^{(\mathcal{G}^x)}(\mathbf{A}) &= \alpha_s \sum_{(i,j) \in \mathcal{E}^x} \mathbf{W}_{i,j}^x \|\mathbf{a}^{(i)} - \mathbf{a}^{(j)}\|_2^2 \\ &= \alpha_s \operatorname{Tr}(\mathbf{A} \mathbf{L}^x \mathbf{A}^T), \end{aligned} \quad (10)$$

avec $\alpha_s > 0$ un paramètre de pénalisation et \mathbf{L}^x la matrice Laplacienne de \mathcal{G}^x .

Pénalisation du problème DU. Une première contrainte populaire est la pénalité de *régularité* sur les atomes plutôt que sur les signaux eux-mêmes. Étant donné un graphe \mathcal{G} , cette contrainte, introduite par Yankelevsky [2], est définie comme suit

$$f^{(sm)}(\mathbf{D}) = \beta_s \sum_{m=1}^M S_{\mathcal{G}}(\mathbf{d}^{(m)}) = \beta_s \operatorname{Tr}(\mathbf{D}^T \mathbf{L} \mathbf{D}) \quad (11)$$

avec $\beta_s > 0$ un paramètre de pénalisation. L'ajout de cette contrainte permet de guider l'apprentissage vers des atomes réguliers par rapport à la structure du graphe sous-jacent.

Une approche alternative consiste à considérer les atomes d'une famille spécifique, en imposant une structure particulière au dictionnaire et en apprenant les paramètres de cette structure. La structure en question incorpore généralement des propriétés souhaitables du dictionnaire telles que l'invariance par translation ou une *consistance* minimale. [6]. En particulier, citons la méthode introduite par Thanou [3] qui propose de structurer le dictionnaire appris comme un ensemble de sous-dictionnaires, chacun étant un polynôme de la matrice Laplacienne du graphe.

3 Modèle d'apprentissage proposé

Nous présentons maintenant notre variation au problème (6), motivée par les principes d'incertitude introduits section 2.1.

3.1 Problèmes d'optimisation

Étape AU. À l'instar de Ramamurthy [5], notre méthode utilise la pénalité d'activation de graphe $g^{(\mathcal{G}^x)}$, ce qui conduit à l'étape AU suivante

$$\mathbf{A}^* = \underset{\mathbf{A}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 + \alpha_s \operatorname{Tr}(\mathbf{A} \mathbf{L}^x \mathbf{A}^T) \quad (12)$$

Cette pénalisation est retenue car favorisant la dissimilarité entre atomes et réduisant les dépendances linéaires entre eux. Afin de construire un graphe de similarité \mathcal{G}^x , la matrice de poids \mathbf{W}^x est générée, pour $i, j = 1, \dots, L$, comme

$$\mathbf{W}_{i,j}^x = \exp\left(-\frac{\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|^2}{\sigma^2}\right). \quad (13)$$

Par suite, la matrice \mathbf{L}^x associée au graphe \mathcal{G}^x peut être facilement obtenue à partir de la matrice des poids \mathbf{W}^x .

Etape DU. Le problème DU considéré introduit des dictionnaires mixtes, *i.e.* composés de plusieurs types d'atomes. Dans la continuité des principes d'incertitude introduits section 2.1, nous proposons ici d'apprendre des dictionnaires avec deux types d'atomes : des atomes locaux sur le graphe, et des atomes réguliers. Ainsi, étant donné M_l atomes de localité et M_s atomes de régularité, la décomposition suivante est effectuée

$$\mathbf{D} = \left(\mathbf{D}^{(loc)}, \mathbf{D}^{(sm)} \right) = \left(\mathbf{d}^{(1)}, \dots, \mathbf{d}^{(M_l+M_s)} \right), \quad (14)$$

avec $\mathbf{D}^{(loc)} \in N \times M_l$ le sous-dictionnaire de localité et $\mathbf{D}^{(sm)} \in N \times M_s$ le sous-dictionnaire de régularité.

Par suite, il apparait naturel de contraindre le sous-dictionnaire de régularité $\mathbf{D}^{(sm)}$ par l'utilisation de la pénalisation $f^{(sm)}$ — voir Equation (11). Cependant, contraindre l'apprentissage du sous-dictionnaire de localité $\mathbf{D}^{(loc)}$ nécessite l'introduction d'une fonction de perte appropriée, qui est définie, étant donné un graphe \mathcal{G} , une distance de graphe $d^{\mathcal{G}}$ — *e.g.* la distance géodésique de graphe [7] — et un ensemble (c_1, \dots, c_{M_l}) de M_l nœuds de centralité de \mathcal{G} , comme

$$f^{(loc)}(\mathbf{D}) = \beta_l \sum_{i=1}^{M_l} \text{Loc}_{c_i}(\mathbf{d}^{(i)}) \quad (15)$$

avec un paramètre $\beta_l > 0$ et, pour $i = 1, \dots, M_l$, $\text{Loc}_{c_i}(\mathbf{x}) = \mathbf{x}^T \mathbf{P}_i \mathbf{x}$, avec $\mathbf{P}_i = \text{Diag}(d^{\mathcal{G}}(c_i, v_1)^2, \dots, d^{\mathcal{G}}(c_i, v_N)^2)$. Une question importante concerne la position des centres. Si l'on dispose d'une connaissance *a priori* de ces centres, il est possible de guider l'apprentissage du dictionnaire en les imposant de manière fixe. Dans le cas contraire, les nœuds de centralité peuvent être échantillonnés aléatoirement.

Finalement, étant donné une matrice $\mathbf{A} \in \mathbb{R}^{(M_l+M_s) \times L}$, le problème d'apprentissage de localité-régularité (LSDL) est donné par

$$\mathbf{D}^* = \underset{\mathbf{D}^{(loc)}, \mathbf{D}^{(sm)}}{\text{argmin}} \left\| \mathbf{X} - [\mathbf{D}^{(sm)}, \mathbf{D}^{(loc)}] \mathbf{A} \right\|_F^2 + f^{(loc)}(\mathbf{D}^{(loc)}) + f^{(sm)}(\mathbf{D}^{(sm)}) \quad (16)$$

3.2 Résolution

Etape AU. Le problème (12) est convexe en \mathbf{A} mais ne présente pas de solution explicite. Une recherche de minimum par descente de gradient est donc réalisée, telle que

$$\mathbf{A}^{(\tau)} = \mathbf{A}^{(\tau-1)} - \theta \nabla_{\mathbf{A}} \left(\left\| \mathbf{X} - \mathbf{D} \mathbf{A} \right\|_F^2 + g^{(\mathcal{G}^x)}(\mathbf{A}) \right) \quad (17)$$

avec $\theta > 0$ le pas de descente et

$$\nabla_{\mathbf{A}} \left(\left\| \mathbf{X} - \mathbf{D} \mathbf{A} \right\|_F^2 + g^{(\mathcal{G}^x)}(\mathbf{A}) \right) = 2 \alpha_s \mathbf{A} \mathbf{L}_x - 2 \mathbf{D}^T (\mathbf{X} - \mathbf{D} \mathbf{A}). \quad (18)$$

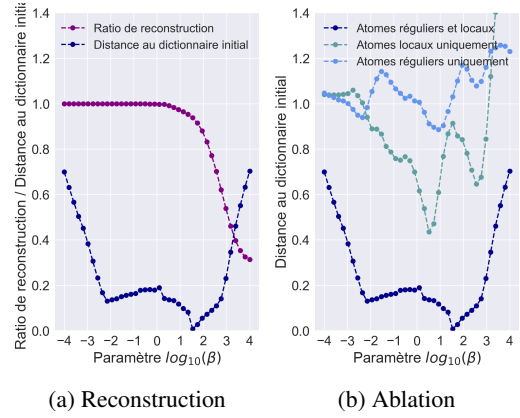


FIGURE 1 – La Figure 1a présente les taux de reconstruction et distance au dictionnaire initial \mathbf{D}^0 . La Figure 1b présente les distances au dictionnaire initial \mathbf{D}^0 de dictionnaires appris régulier (bleu clair), de localité (vert) et mixte (bleu foncé).

Etape DU. Afin de simplifier le problème (16), nous introduisons l'expression vectorisée $\mathbf{v}(\mathbf{D}) \in \mathbb{R}^{N(M_l+M_s)}$ de \mathbf{D} et $\mathbf{v}(\mathbf{X}) \in \mathbb{R}^{NL}$ celle de \mathbf{X} .

Par suite, l'étape d'actualisation du dictionnaire de localité-régularité peut être reformulée comme suit

$$\mathbf{v}(\mathbf{D}^*) = \underset{\mathbf{d} \in \mathbb{R}^{N(M_l+M_s)}}{\text{argmin}} \left\| \mathbf{v}(\mathbf{X}) - (\mathbf{A} \otimes \mathbf{I}_N)^T \mathbf{d} \right\|_2^2 + \mathbf{d}^T \mathbf{Q} \mathbf{d} \quad (19)$$

avec $\mathbf{Q} = \text{Diag}(\beta_l \mathbf{P}_1, \dots, \beta_l \mathbf{P}_{M_l}, \beta_s \mathbf{L}, \dots, \beta_s \mathbf{L})$ une matrice diagonale par bloc — \mathbf{L} étant répété M_s fois. Cette reformulation du problème (16) met en évidence les propriétés quadratiques du problème LSDL, dont la solution est donnée par

$$\mathbf{v}(\mathbf{D}^*) = \left((\mathbf{A} \mathbf{A}^T) \otimes \mathbf{I}_N + \mathbf{Q} \right)^{-1} (\mathbf{A} \otimes \mathbf{I}_N) \mathbf{v}(\mathbf{X}) \quad (20)$$

4 Applications numériques

4.1 Reconstruction de dictionnaire synthétique

Dans cette première partie, nous cherchons à reconstruire un dictionnaire ayant généré un ensemble de signaux synthétiques. Un graphe géométrique aléatoire \mathcal{G} est construit en connectant un ensemble de $N = 500$ points tirés uniformément sur $[0, 1]^2$ par une procédure de 5-plus proches voisins. La distance associée à \mathcal{G} est la distance géodésique sur graphe [7].

Un ensemble de $M_s = 5$ atomes réguliers est échantillonné à partir d'une distribution gaussienne de matrice de covariance \mathbf{L}^{-1} . Un ensemble de $M_l = 10$ nœuds est échantillonné aléatoirement et les signaux localisés autour des nœuds à une distance géodésique < 2 des centres sont générés à partir d'une distribution gaussienne. Les atomes sont ensuite concaténés pour former \mathbf{D}^0 . La matrice d'activation $\mathbf{A}^0 \in \mathbb{R}^{(M_s+M_l) \times L}$ est tirée à partir d'une distribution normale, avec $L = 1000$. Enfin, la matrice de données est donnée par $\mathbf{X} = \mathbf{D}^0 \mathbf{A}^0$.

