



HAL
open science

Echantillonnage de signaux sur graphes via des processus déterminantaux

Nicolas Tremblay, Simon Barthelme, Pierre-Olivier Amblard

► **To cite this version:**

Nicolas Tremblay, Simon Barthelme, Pierre-Olivier Amblard. Echantillonnage de signaux sur graphes via des processus déterminantaux. 2017. hal-01503736v1

HAL Id: hal-01503736

<https://hal.science/hal-01503736v1>

Preprint submitted on 7 Apr 2017 (v1), last revised 30 Jun 2017 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Échantillonnage de signaux sur graphes via des processus déterminantaux

Nicolas TREMBLAY, Simon BARTHELMÉ, Pierre-Olivier AMBLARD

GIPSA-LAB, CNRS, Grenoble, France
prenom.nom@gipsa-lab.fr

Résumé – Nous considérons l'échantillonnage de signaux à bande limitée k sur graphe, *i.e.*, les combinaisons linéaires des k premiers modes de Fourier du graphe. Il existe k nœuds du graphe qui permettent leur reconstruction parfaite, les exhiber nécessite cependant une diagonalisation partielle de la matrice laplacienne, trop coûteuse à grande dimension. Nous proposons une nouvelle méthode rapide d'échantillonnage basée sur des processus déterminantaux qui permet la reconstruction à partir d'un nombre d'échantillons de l'ordre de k . Ce faisant, nous proposons un nouvel algorithme général d'échantillonnage de processus déterminantaux, plus rapide d'un facteur k que l'algorithme de l'état de l'art.

Abstract – We consider the problem of sampling k -bandlimited graph signals, *i.e.*, linear combinations of the first k graph Fourier modes. We know that a set of k nodes embedding all k -bandlimited signals always exists, thereby enabling their perfect reconstruction after sampling. Unfortunately, to exhibit such a set, one needs to partially diagonalize the graph Laplacian, which becomes prohibitive at large scale. We propose a novel strategy based on determinantal point processes that side-steps partial diagonalisation and enables reconstruction with only $O(k)$ samples. While doing so, we exhibit a new general algorithm to sample determinantal process, faster than the state-of-the-art algorithm by an order k .

1 Introduction

Étant donnée une certaine classe de signaux, échantillonner consiste à mesurer un signal un nombre de fois suffisant pour une certaine tâche, e.g. à des fins de reconstruction. Pour les signaux définis sur des graphes, une classe de régularité souvent utilisée est celle des signaux à bande limitée k (cf la définition 1). Dans ce contexte, il existe deux types de méthodes d'échantillonnage : i) celles qui calculent les k premiers modes de Fourier du graphe et cherchent via des heuristiques k nœuds qui les discriminent tous [2], ii) celles qui s'affranchissent de ce calcul, et qui soit cherchent un nombre de nœuds proche de k et résolvent des problèmes combinatoires qui ne passent pas à l'échelle [3], soit s'autorisent un nombre de nœuds de l'ordre de $O(k \log k)$ et permettent de passer à l'échelle via un échantillonnage iid adapté au graphe [4], soit via des marches aléatoires sur graphe spécifiques [5].

Contributions. Nous proposons une méthode basée sur les processus ponctuels déterminantaux (PPD) [6], ces processus aléatoires connus pour favoriser la diversité des échantillons selon un certain sens choisi. Nous proposons un nouvel algorithme général pour échantillonner des PPD, plus rapide d'un facteur k que l'état de l'art. Puis, nous montrons comment cet algorithme peut être utilisé et approché rapidement dans le cadre de l'échantillonnage des signaux à bande limitée sur graphe.

2 Notations et objectif

Les matrices sont en majuscule, e.g. K ; les vecteurs en minuscules et en gras, e.g. \mathbf{x} ; les ensembles en cursive, e.g. \mathcal{A} ; $K_{\mathcal{A},\mathcal{B}}$ est la restriction de K aux lignes (resp. colonnes) in-

dexées par les éléments de \mathcal{A} (resp. \mathcal{B}); enfin : $K_{\mathcal{A}} = K_{\mathcal{A},\mathcal{A}}$. On note δ_s le vecteur dont les entrées sont nulles sauf en s .

On considère un graphe \mathcal{G} composé de N nœuds interconnectés selon la matrice d'adjacence $W \in \mathbb{R}^{N \times N}$ tel que $W_{ij} = W_{ji} \geq 0$ représente le poids associé au lien connectant i à j . Notons D la matrice diagonale des degrés de chaque nœud : l'entrée (i, i) de cette matrice s'écrit $D_{ii} = \sum_j W_{ij}$. Le laplacien combinatoire associé à \mathcal{G} s'écrit $L = D - W \in \mathbb{R}^{N \times N}$. Cette matrice est semi-définie positive et se diagonalise en $L = U\Lambda U^T$, où $U = (\mathbf{u}_1 | \mathbf{u}_2 | \dots | \mathbf{u}_N) \in \mathbb{R}^{N \times N}$ est la matrice de ses vecteurs propres et $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N) \in \mathbb{R}^{N \times N}$ la matrice diagonale de ses valeurs propres associées, que l'on range dans l'ordre croissant $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$. Par analogie au traitement du signal discret classique, \mathbf{u}_i est considéré comme le i -ème mode de Fourier du graphe, et λ_i sa fréquence associée [7]. Étant donné $k \in \mathbb{N}^*$, on définit $U_k = (\mathbf{u}_1 | \dots | \mathbf{u}_k) \in \mathbb{R}^{N \times k}$ la concaténation des k premiers modes de Fourier du graphe. Un signal à bande limitée se définit ainsi :

Définition 1 (Signal à bande limitée k). *Un signal $\mathbf{x} \in \mathbb{R}^N$ défini sur les nœuds d'un graphe \mathcal{G} est à bande limitée $k \in \mathbb{N}^*$ si $\mathbf{x} \in \text{span}(U_k)$, *i.e.*, $\exists \boldsymbol{\alpha} \in \mathbb{R}^k$ tel que $\mathbf{x} = U_k \boldsymbol{\alpha}$.*

En notant m le nombre de mesures, échantillonner consiste à choisir un ensemble de m nœuds $\mathcal{A} = \{s_1, \dots, s_m\}$. Notons $M = (\delta_{s_1} | \delta_{s_2} | \dots | \delta_{s_m})^T \in \mathbb{R}^{m \times N}$ la matrice de mesure associée : le signal $\mathbf{x} \in \mathbb{R}^N$ mesuré en \mathcal{A} s'écrit

$$\mathbf{y} = M\mathbf{x} + \mathbf{n} \in \mathbb{R}^m, \quad (1)$$

où $\mathbf{n} \in \mathbb{R}^m$ est un bruit de mesure. Étant donné que \mathbf{x} est supposé être à bande limitée k , on le reconstruit à partir de sa

Algorithm 1 Échantillonner un m -PPD à noyau K [8, Sec 2.4]

Entrée : K, m
 $\mathcal{S} \leftarrow \emptyset$, définir $p_0 = \text{diag}(K) \in \mathbb{R}^N$
 $p \leftarrow p_0$
for $n = 1, \dots, m$ **do** :
• Tirer s_n avec probabilité $\mathbb{P}(s) = p(s) / \sum_i p(i)$
• $\mathcal{S} \leftarrow \mathcal{S} \cup \{s_n\}$
• Mettre à jour $p : \forall i \quad p(i) = p_0(i) - K_{\mathcal{S},i}^T K_{\mathcal{S}}^{-1} K_{\mathcal{S},i}$
end for
Sortie : $\mathcal{A} \leftarrow \mathcal{S}$.

mesure \mathbf{y} en calculant $\mathbf{x}_{\text{rec}} = U_k(\text{MU}_k)^\dagger \mathbf{y}$, où $(\text{MU}_k)^\dagger$ est le pseudo-inverse de Moore-Penrose de $\text{MU}_k \in \mathbb{R}^{m \times k}$. Notons $\sigma_1 \leq \dots \leq \sigma_k$ les valeurs singulières de MU_k . Le signal \mathbf{x} est parfaitement reconstitué (au bruit prêt) à partir de \mathbf{y} si $U_k^T \text{M}^T \text{MU}_k$ est inversible, i.e., si $\sigma_1^2 > 0$. Dans ce cas :

$$\mathbf{x}_{\text{rec}} = U_k (U_k^T \text{M}^T \text{MU}_k)^{-1} U_k^T \text{M}^T \mathbf{y} \quad (2)$$

$$= \mathbf{x} + U_k (U_k^T \text{M}^T \text{MU}_k)^{-1} U_k^T \text{M}^T \mathbf{n}. \quad (3)$$

Parmi tous les choix possibles de \mathcal{A} (et donc de M) qui vérifient $\sigma_1^2 > 0$, lesquels sont optimaux ? Il existe plusieurs définitions d’optimalité [5], nous nous intéressons ici à la suivante :

$$\mathcal{A}^{\text{MV}} = \arg \max_{\mathcal{A} \text{ s.t. } |\mathcal{A}|=k} \prod_{i=1}^k \sigma_i^2, \quad (4)$$

où “MV” signifie “volume maximal” : en effet, en maximisant le produit des valeurs singulières, on maximise le déterminant de $U_k^T \text{M}^T \text{MU}_k$, c’est-à-dire le volume formé par les k lignes échantillonnées de U_k . Trouver \mathcal{A}^{MV} est NP-complet [1]. Notre objectif est de s’en approcher, c’est-à-dire : trouver un ensemble de taille k (comme \mathcal{A}^{MV}) ou proche de k tel qu’on ait la meilleure reconstruction possible des signaux à bande limitée k .

3 Processus déterminantaux

Notons $[N]$ l’ensemble des sous-ensembles de $\{1, \dots, N\}$ et $K \in \mathbb{R}^{N \times N}$ une matrice semi-définie positive (SDP). On convient que $\det(\emptyset) = 1$.

Définition 2 (m -PPD). *Considérons un processus ponctuel, i.e., un processus qui tire aléatoirement un ensemble $\mathcal{A} \in [N]$. Ce processus est un m -PPD à noyau K si :*

- i) $\mathbb{P}(\mathcal{A}) = 0$, pour tout \mathcal{A} tel que $|\mathcal{A}| \neq m$.
- ii) $\mathbb{P}(\mathcal{A}) = \frac{1}{Z} \det(K_{\mathcal{A}})$, pour tout \mathcal{A} tel que $|\mathcal{A}| = m$, où Z est la constante de normalisation.

Proposition 1. *Si le rang de K est supérieur ou égal à m , alors l’algorithme 1 échantillonne un m -PPD à noyau K .*

Démonstration. Notons \mathcal{S} l’ensemble des n échantillons obtenus à l’issue de l’étape n de la boucle de l’algorithme 1, et $\bar{\mathcal{S}}$ l’ensemble des $n - 1$ échantillons obtenus à l’issue de l’étape $n - 1$. On a donc : $\mathcal{S} = \bar{\mathcal{S}} \cup \{s_n\}$. Enfin, notons $p_n(i)$ la valeur de $p(i)$ à l’issue de la n -ème itération de la boucle.

En utilisant le complément de Schur, on a : $\forall n \in [1, m], \forall i$,

$$\det(K_{\bar{\mathcal{S}} \cup \{i\}}) = \left(K_{i,i} - K_{\bar{\mathcal{S}},i}^T K_{\bar{\mathcal{S}}}^{-1} K_{\bar{\mathcal{S}},i} \right) \det(K_{\bar{\mathcal{S}}}) \quad (5)$$

$$= p_{n-1}(i) \det(K_{\bar{\mathcal{S}}}). \quad (6)$$

Algorithm 2 Échantillonner un m -PPD – nouvel algorithme

Entrée : $K = [k_1, \dots, k_N], m$
 $\mathcal{S} \leftarrow \emptyset$, définir $p = \text{diag}(K) \in \mathbb{R}^N$
for $n = 1, \dots, m$ **do** :
• Tirer s_n avec probabilité $\mathbb{P}(s) = p(s) / \sum_i p(i)$
• $\mathcal{S} \leftarrow \mathcal{S} \cup \{s_n\}$
• Calculer $\mathbf{f}_n = \mathbf{k}_{s_n} - \sum_{l=1}^{n-1} \mathbf{f}_l f_l(s_n)$
• Normaliser $\mathbf{f}_n \leftarrow \mathbf{f}_n / \sqrt{f_n(s_n)}$
• Mettre à jour $p : \forall i \quad p(i) \leftarrow p(i) - f_n(i)^2$
end for
Sortie : $\mathcal{A} \leftarrow \mathcal{S}$.

À partir de (6), et sachant que i) K étant SDP : $\forall \mathcal{S}, \det(K_{\mathcal{S}}) \geq 0$, ii) le rang de K est supérieur ou égal à m , on peut montrer que $p_n(i) \geq 0$ et $\sum_i p_n(i) \neq 0$, c’est-à-dire qu’à chaque itération de la boucle, la probabilité $\mathbb{P}(s)$ est bien définie. La boucle étant répétée m fois, la sortie de l’algorithme, notée \mathcal{A} , est nécessairement de taille m . Enfin, montrons que $\mathbb{P}(\mathcal{A})$ est proportionnel à $\det(K_{\mathcal{A}})$. Par construction de \mathcal{A} :

$$\mathbb{P}(\mathcal{A}) = \prod_{l=1}^m \mathbb{P}(s_l | s_1, s_2, \dots, s_{l-1}) \propto \prod_{l=1}^m p_{l-1}(s_l). \quad (7)$$

Or, en réécrivant (6) pour $i = s_n$, et par itération sur $\bar{\mathcal{S}}$, on peut écrire : $\det(K_{\mathcal{A}}) = \prod_{l=1}^m p_{l-1}(s_l)$ qui est donc proportionnel à $\mathbb{P}(\mathcal{A})$ d’après (7). Ce qui termine la preuve. \square

\mathcal{A}^{MV} , l’ensemble à approcher, est l’ensemble le plus probable du m -PPD (avec $m = k$) à noyau $K_k = U_k U_k^T$. Si nous avons les ressources pour calculer U_k , une première stratégie d’échantillonnage est donc l’Alg. 1 appliqué à K_k . Dans le cas contraire, nous proposons dans la suite un nouvel algorithme d’échantillonnage de m -PPD, de complexité inférieure d’un facteur m , qui permet d’appliquer des techniques d’approximation polynomiale, évitant ainsi toute étape de diagonalisation.

4 Approximation via des filtres sur graphe

4.1 Réécriture de l’algorithme d’échantillonnage

Proposition 2. *L’algorithme 2 est équivalent à l’algorithme 1 : il échantillonne aussi un m -PPD à noyau K .*

Démonstration. Considérons $\mathcal{S}, \bar{\mathcal{S}}, p_n(i)$ définis comme dans la preuve précédente. Nous allons montrer que les $p_n(i)$ dans les boucles des deux algorithmes sont égaux. Dans l’Alg. 2 : $p_n(i) = p_{n-1}(i) - f_n(i)^2 = p_0(i) - \sum_{l=1}^n f_l(i)^2$. En comparant avec $p_n(i)$ obtenu dans l’Alg. 1, il suffit de montrer que :

$$\forall n \forall i \quad \sum_{l=1}^n f_l(i)^2 = K_{\bar{\mathcal{S}},i}^T K_{\bar{\mathcal{S}}}^{-1} K_{\bar{\mathcal{S}},i}. \quad (8)$$

Nous allons montrer plus généralement que :

$$\forall n \forall i, j \quad \sum_{l=1}^n f_l(i) f_l(j) = K_{\bar{\mathcal{S}},i}^T K_{\bar{\mathcal{S}}}^{-1} K_{\bar{\mathcal{S}},j}. \quad (9)$$

Pour ce faire, nous proposons une récurrence.

Initialisation. C’est vrai pour $n = 1$, où \mathcal{S} est réduit à $\{s_1\}$:

$$\forall i, j \quad K_{\mathcal{S},i}^T K_{\mathcal{S}}^{-1} K_{\mathcal{S},j} = K_{s_1,i} K_{s_1,j} / K_{s_1,s_1} = f_1(i) f_1(j).$$

Hypothèse. Supposons que (9) est vraie à l'étape $n - 1$, i.e. :

$$\forall i, j \quad \sum_{l=1}^{n-1} f_l(i)f_l(j) = \mathbf{K}_{\mathcal{S},i}^T \mathbf{K}_{\mathcal{S}}^{-1} \mathbf{K}_{\mathcal{S},j}. \quad (10)$$

Réurrence. Montrons qu'elle est également vraie à l'étape n . En utilisant l'identité de Woodbury sur $\mathbf{K}_{\mathcal{S}}^{-1}$, on montre que :

$$\mathbf{K}_{\mathcal{S},i}^T \mathbf{K}_{\mathcal{S}}^{-1} \mathbf{K}_{\mathcal{S},j} = \mathbf{K}_{\mathcal{S},i}^T \mathbf{K}_{\mathcal{S}}^{-1} \mathbf{K}_{\mathcal{S},j} + \frac{z_n(i)z_n(j)}{z_n(s_n)}, \quad (11)$$

où $z_n(i) = \mathbf{K}_{s_n,i} - \mathbf{K}_{\mathcal{S},s_n}^T \mathbf{K}_{\mathcal{S}}^{-1} \mathbf{K}_{\mathcal{S},i}$. En remplaçant $\mathbf{K}_{\mathcal{S},i}^T \mathbf{K}_{\mathcal{S}}^{-1} \mathbf{K}_{\mathcal{S},j}$ par $\sum_{l=1}^{n-1} f_l(i)f_l(j)$ grâce à (10), il nous reste à montrer que :

$$\forall i, j \quad f_n(i)f_n(j) = \frac{z_n(i)z_n(j)}{z_n(s_n)}. \quad (12)$$

Or, par construction dans l'Algorithme 2, $f_n(i)$ s'écrit :

$$\forall i \quad f_n(i) = \frac{\mathbf{K}_{s_n,i} - \sum_{l=1}^{n-1} f_l(i)f_l(s_n)}{\sqrt{\mathbf{K}_{s_n} - \sum_{l=1}^{n-1} f_l(s_n)^2}}. \quad (13)$$

En utilisant une seconde fois l'hypothèse (10), on montre que :

$$\forall i \quad f_n(i) = \frac{z_n(i)}{\sqrt{z_n(s_n)}}, \quad (14)$$

ce qui prouve (12) et termine la preuve. \square

L'algorithme 2 est un nouvel algorithme général pour échantillonner un m -PPD. Sa complexité est en $O(Nm^2)$, alors que la complexité de l'algorithme 1 est en $O(Nm^3)$. Notons que des idées similaires pour réduire la complexité d'un facteur m existent dans la littérature (par exemple dans [9]) mais sous des formes un peu cachées, et, à notre connaissance, n'ont jamais été vraiment explicitées dans le cas discret.

4.2 Approcher l'échantillonnage d'un PPD

En plus d'être plus rapide, nous allons voir que la manière dont l'algorithme 2 fait appel à \mathbf{K} permet de tirer profit d'approximations polynomiales usuellement utilisées lors d'opérations de filtrage sur graphe. Rappelons que dans notre cas, nous nous intéressons au m -PPD associé au noyau projectif \mathbf{K}_k . Or : $\mathbf{K}_k = \mathbf{U}h_k(\Lambda)\mathbf{U}^T$, où $h_k(\Lambda) = \text{diag}(h_k(\lambda_1), \dots, h_k(\lambda_N))$ et $h_k(\lambda)$ est tel que $h_k(\lambda) = 1$ si $\lambda \leq \lambda_k$ et $h_k(\lambda) = 0$ sinon. En traitement du signal sur graphes, \mathbf{K}_k est un filtre passe-bas idéal de fréquence de coupure λ_k .

Approximation polynomiale. Considérons le polynôme de Chebychev \tilde{h}_k de degré r qui approche au mieux h_k :

$$\forall \lambda \in [0, \lambda_N] \quad \tilde{h}_k(\lambda) = \sum_{l=1}^r \alpha_l \lambda^l \simeq h_k(\lambda).$$

$$\text{On a : } \mathbf{K}_k \simeq \mathbf{U}\tilde{h}_k(\Lambda)\mathbf{U}^T = \sum_{l=1}^r \alpha_l \mathbf{U}\Lambda^l \mathbf{U}^T = \sum_{l=1}^r \alpha_l \mathbf{L}^l = \tilde{h}_k(\mathbf{L}).$$

Filtrage rapide sur graphe. On ne calcule jamais explicitement $\tilde{h}_k(\mathbf{L})$ qui est en général dense de taille $N \times N$. En revanche, étant donné un signal \mathbf{x} défini sur le graphe, le signal

Algorithm 3 Échantillonnage approché d'un m -PPD à noyau $\mathbf{K} = h(\mathbf{L})$

Entrée : $\mathbf{L}, h(\lambda), r, m$

Calculer λ_N , la plus grande valeur propre de \mathbf{L}

Calculer le polynôme \tilde{h} de degré r qui approche h sur $[0, \lambda_N]$

Estimer $\mathbf{p} = \text{diag}(\tilde{h}(\mathbf{L})) \in \mathbb{R}^N$ comme vu dans la Sec. 4.2

for $n = 1, \dots, m$ **do** :

• Tirer $s_n \leftarrow \text{argmax}(\mathbf{p})$

• Calculer $\mathbf{f}_n = \tilde{h}(\mathbf{L})\boldsymbol{\delta}_{s_n} - \sum_{l=1}^{n-1} \mathbf{f}_l f_l(s_n)$

• Normaliser $\mathbf{f}_n \leftarrow \mathbf{f}_n / \sqrt{f_n(s_n)}$

• Mettre à jour $p(i) \leftarrow p(i) - f_n(i)^2$

end for

Output : $\mathcal{A} = \{s_1, \dots, s_m\}$

filtré par h_k , $\mathbf{K}_k \mathbf{x}$, est approché par $\tilde{h}_k(\mathbf{L})\mathbf{x} = \sum_{l=1}^r \alpha_l \mathbf{L}^l \mathbf{x}$, qui peut se calculer via r multiplications matrice-vecteur (on calcule d'abord $\mathbf{L}\mathbf{x}$, puis $\mathbf{L}(\mathbf{L}\mathbf{x})$, etc.) si bien que le nombre d'opérations nécessaires pour filtrer un signal est de l'ordre de $r|E|$ où $|E|$ est le nombre de liens du graphe.

Estimation de la diagonale de \mathbf{K}_k . Soit $\mathbf{R} \in \mathbb{R}^{N \times n}$ une matrice contenant n signaux aléatoires, dont toutes les entrées sont tirées d'une gaussienne de moyenne nulle et de variance $1/n$. La $i^{\text{ème}}$ valeur de la diagonale de \mathbf{K}_k , $p_0(i)$, est approchée par la norme de la $i^{\text{ème}}$ ligne de $\tilde{h}_k(\mathbf{L})\mathbf{R}$, i.e. :

$$p_0(i) \simeq \left\| \boldsymbol{\delta}_i^T \tilde{h}_k(\mathbf{L})\mathbf{R} \right\|^2. \quad (15)$$

En appliquant le lemme de Johnson-Lindenstrauss, on peut montrer qu'un nombre de signaux aléatoires $n = O(\log N)$ est suffisant pour avoir une estimation convenable [10].

Algorithme d'approximation. On souhaite approcher l'algorithme 2 avec entrée \mathbf{K}_k sans avoir à calculer \mathbf{U}_k . Pour ce faire, on suit l'algorithme 3 en lui donnant comme entrée $\mathbf{L}, h(\lambda) = h_k(\lambda), r = 50$ et m le nombre d'échantillons souhaité. Au lieu d'accéder directement à la diagonale d'un noyau \mathbf{K} connu, on l'estime à l'aide de l'équation (15). Puis, au lieu d'accéder aux colonnes \mathbf{k}_s directement, on les estime via le filtrage rapide associé : $\mathbf{k}_s \simeq \tilde{h}_k(\mathbf{L})\boldsymbol{\delta}_s$. Au vu des erreurs d'approximation successives, l'écart entre les $\{\mathbf{f}_l\}$ calculés au cours des algorithmes 2 et 3 ne cesse de s'accroître au fur et à mesure de la boucle. Pour stabiliser l'algorithme 3 : i) le nouvel échantillon est celui qui maximise \mathbf{p} ; ii) et on s'autorise $m > k$ (ce qui n'a pas de sens pour l'algorithme 2 appliqué à \mathbf{K}_k qui est de rang k). Aussi, en pratique, on fait quelques modifications mineures : i) après la mise à jour de \mathbf{p} , on force $p(s_i) = 0$ pour tous les échantillons s_i déjà choisis ; ii) lors de la normalisation de \mathbf{f}_n , si $f_n(s_n) \leq 0$ dues aux erreurs d'approximations, alors on normalise \mathbf{f}_n par $\sqrt{\|\mathbf{f}_n\|/N}$.

5 Simulations

Trois graphes. Nous considérons le graphe routier du Minnesota ($N = 2642$), le graphe correspondant au maillage 3D d'un objet (graphe "Bunny" avec $N = 2503$), et une réalisation d'un

1. pour avoir $h_k(\lambda)$, il faut λ_k , que nous estimons en suivant [4, Sec. 4.2].

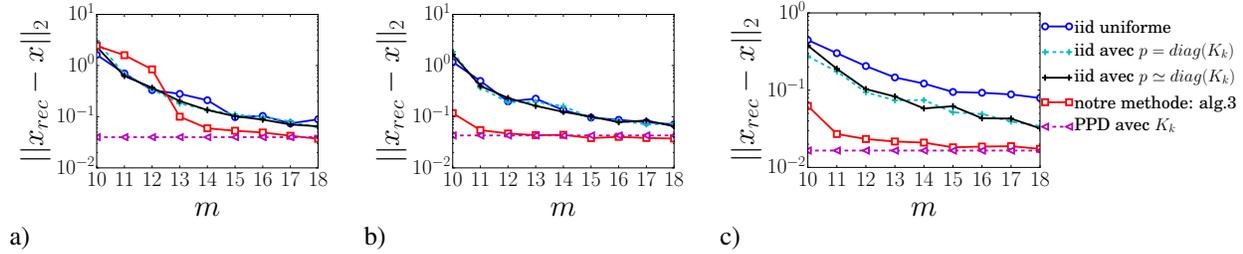


FIGURE 1 – Performance médiane obtenue en reconstruisant 100 signaux à bande limitée 10 sur a) le graphe du Minnesota, b) le graphe “Bunny”, c) une réalisation d’un MSB. Les méthodes en pointillés doivent calculer U_k pour échantillonner, les autres non. La méthode via le PPD à noyau K_k ne peut échantillonner que $m = 10$ nœuds : la droite horizontale associée est artificielle et sert uniquement à la comparaison.

graphe aléatoire avec communautés issue du modèle stochastique par blocs (MSB) avec $N = 1000$, 10 communautés de même taille et de paramètre de structure $\epsilon = \epsilon_c/4$, où ϵ_c correspond à la limite où les structures en communautés ne sont plus détectables ; et plus ϵ est petit, plus la structure en communautés est forte. Voir par exemple [5, sec. 4] pour des détails sur ϵ ; [4, Fig. 1] pour des illustrations des deux premiers graphes.

Pour générer les signaux à bande limitée. Dans tous ces exemples, nous choisissons $k = 10$. Chaque signal x à bande limitée k est généré en calculant U_k , puis en calculant $x = U_k \alpha$ où $\alpha \in \mathbb{R}^k$ est tiré aléatoirement selon une gaussienne $\mathcal{G}(0, 1)$, puis en normalisant x pour qu’il soit de norme 1.

Les différentes méthodes d’échantillonnage comparées. On compare 5 méthodes d’échantillonnage : i) l’échantillonnage iid uniforme sans remise, ii) l’échantillonnage iid selon $p = \text{diag}(K_k)$ sans remise (comme dans [4]), iii) la même stratégie mais avec une distribution $p \simeq \text{diag}(K_k)$ estimée via (15), iv) l’échantillonnage dépendant en suivant l’algorithme 3, et v) l’échantillonnage selon le PPD idéal, c’est-à-dire en suivant l’algorithme 1 avec la légère modification qu’à chaque itération de la boucle on tire l’échantillon qui maximise la probabilité (cela donne de meilleures performances de reconstruction dans les cas testés). Les quatre premières méthodes peuvent échantillonner un nombre quelconque de nœuds m ; la dernière échantillonne nécessairement k nœuds.

Le bruit de mesure $n \in \mathbb{R}^m$ est tiré selon une Gaussienne de moyenne nulle et d’écart-type σ , et est ajouté à la mesure du signal, comme dans (1). On fixe $\sigma = 10^{-3}$.

Reconstruction des signaux échantillonnés. Quelle que soit la méthode choisie, on reconstruit les signaux avec (10), c’est-à-dire avec la connaissance de U_k . Il existe des moyens de reconstruction qui ne nécessitent pas U_k (voir [4]) ; mais nous souhaitons ici comparer exclusivement l’échantillonnage.

Discussion des résultats. La Fig. 1 montre que notre méthode approche la performance du PPD à noyau K_k plus rapidement que les autres, dès $m = O(k)$. Pour le graphe du Minnesota, à petit m , notre performance est moindre. Néanmoins, nous retrouvons une meilleure performance pour d’autres choix de k . Nous avons mené de nombreuses autres expériences sur d’autres réalisations du MSB, avec d’autres valeurs de k , ϵ et σ , et des communautés de taille hétérogène : nous trouvons toujours des comportements similaires.

6 Conclusion

Notre objectif initial était de réécrire l’algorithme 1 sous une forme qui nous permettait d’utiliser efficacement les approximations polynomiales du noyau K_k . Nous proposons l’algorithme 2 qui remplit cette objectif, mais qui s’avère également être un algorithme d’intérêt général pour les m -PPD. Nous proposons aussi l’algorithme 3, qui permet d’approcher tout PPD dont le noyau est de la forme $K = h(L)$ où L est une matrice diagonalisable (de préférence parcimonieuse), sans jamais calculer explicitement ce noyau potentiellement dense, et en évitant tout calcul de diagonalisation de L . Nous illustrons avec succès l’intérêt de cet algorithme pour l’échantillonnage de signaux sur graphe à bande limitée k .

Remerciements. Ce travail a été en partie soutenu par le LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01), l’ANR GenGP (ANR-16-CE23-0008), le LIA CNRS/Melbourne Univ Geodesic.

Références

- [1] A. Çivril, M. Magdon-Ismaïl. *On selecting a maximum volume sub-matrix of a matrix and related problems*. Theoretical Computer Science, no. 47, vol. 410, p. 4801–4811, 2009.
- [2] S. Chen, R. Varma, A. Sandryhaila, J. Kovacevic. *Discrete Signal Processing on Graphs : Sampling Theory*. IEEE Transactions on Signal Processing, no. 24, vol. 63, p. 6510–6523, 2015.
- [3] A. Anis, A. Gadde, A. Ortega. *Efficient Sampling Set Selection for Band-limited Graph Signals Using Graph Spectral Proxies*. IEEE Transactions on Signal Processing, no. 14, vol. 64, p. 3775–3789, 2016.
- [4] G. Puy, N. Tremblay, R. Gribonval, P. Vandergheynst. *Random sampling of bandlimited signals on graphs*. Applied and Computational Harmonic Analysis, in press, 2016.
- [5] N. Tremblay, P-O. Amblard, S. Barthélémy. *Graph sampling with determinantal processes*. arXiv preprint, 1703.01594, 2017.
- [6] A. Kulesza, B. Taskar. *Determinantal point processes for machine learning*. Found. and Trends in Mach. Learn., no. 2, vol. 5, p. 123–286, 2012.
- [7] D. Shuman, S. Narang, P. Frossard, A. Ortega, P. Vandergheynst. *The emerging field of signal processing on graphs : Extending high-dimensional data analysis to networks and other irregular domains*. IEEE Signal Processing Magazine, no. 3, vol. 30, p. 83–98, 2013.
- [8] R. Bardenet, A. Hardy. *Monte Carlo with determinantal point processes*. arXiv preprint, 1605.00361, 2016.
- [9] P. Hennig, R. Garnett. *Exact Sampling from Determinantal Point Processes*. arXiv preprint, 1609.06840, 2016.
- [10] N. Tremblay, G. Puy, R. Gribonval, P. Vandergheynst. *Accelerated spectral clustering using graph filtering of random signals*. ICASSP, 2016.