Un processus ponctuel déterminantal pour la sélection d'attributs

Ayoub Belhadji, Rémi Bardenet, Pierre Chainais

Un processus ponctuel déterminantal pour la sélection d'attributs

Ayoub BELHADJI, Rémi BARDENET, Pierre CHAINAIS
Université de Lille, CNRS, Centrale Lille, Inria, UMR 9189 – CRIStAL,
59651 Villeneuve d'Ascq, France
ayoub.belhadji@centralelille.fr, remi.bardenet@gmail.com
pierre.chainais@centralelille.fr

Résumé – La réduction de dimension est une tâche récurrente de l'analyse des signaux en grande dimension. Les axes principaux de l'analyse en composantes principales sont difficilement interprétables. Lorsqu'on souhaite préserver l'interprétabilité des dimensions, la sélection d'attributs est préférable, mais implique a priori une optimisation combinatoire très coûteuse. Cet article contourne la difficulté en proposant un nouvel algorithme de sélection aléatoire d'attributs, ou de manière équivalente, de colonnes de la matrice des données. On utilise pour cela un processus ponctuel déterminantal (PPD) sur les indices des colonnes, dont le noyau est contrôlé par la structure des données.

Abstract – Dimension reduction is often a preamble to the analysis of high-dimensional signals. Principal component analysis projects data onto directions of large variance, but these directions are difficult to interpret. Feature selection avoids this loss of interpretability, at the price of an expensive combinatorial optimization. To avoid paying this price, we propose a novel algorithm for random feature selection, i.e., random column subset selection in the design matrix. The crux is a determinantal point process on the column indices, with a data-dependent kernel.

1 Introduction

La réduction de dimension est une tâche récurrente de l'analyse de données. Les techniques les plus connues sont l'analyse en composantes principales (ACP) et la sélection d'attributs. L'analyse en composantes principales projette les données sur les directions de plus grande variance, les directions principales. Ces nouveaux attributs, représentés par les axes principaux, sont difficilement interprétables comparés aux attributs initiaux. Lorsqu'on souhaite préserver la sémantique de ces derniers, on pratique plutôt la sélection d'attributs.

Si l'on appelle $\boldsymbol{X} \in \mathbb{R}^{N\times d}$ la matrice rassemblant les données sous la forme de N observations exprimées sur d attributs, la sélection d'attributs consiste à choisir un sous-ensemble S d'indices de colonnes de la matrice $\boldsymbol{X}$. Indépendamment de la tâche d'apprentissage à réaliser ensuite, on peut mesurer la qualité d'une sélection S d'attributs à l'aide de l'erreur d'approximation $\|\boldsymbol{X} - \Pi_S \boldsymbol{X}\|$; ici $\|\cdot\|$ est une norme matricielle, et Π_S est la projection sur l'espace engendré par les colonnes de $\boldsymbol{X}$ d'indice S. Trouver naïvement k colonnes qui optimisent ce critère requiert d'explorer les C_d^k combinaisons possibles, ce qui est habituellement hors d'atteinte quand $d \gg 1$. Une alternative à l'exploration combinatoire est de sous-échantillonner aléatoirement des colonnes. Les qualités attendues d'un tel algorithme sont alors *(i)* une erreur d'approximation moyenne faible, i.e. proche de celle de l'ACP de cardinal comparable et *(ii)* un coût d'échantillonnage polynomial.

Cet article propose un nouvel algorithme de sélection aléatoire d'attributs qui possède ces deux qualités. On utilise pour cela un processus ponctuel déterminantal (PPD) sur les indices des colonnes de $\boldsymbol{X}$. Les PPD sont des distributions de probabilité qui produisent des ensembles de points en interaction répulsive. Un PPD est paramétré par une matrice que l'on appelle noyau. En choisissant un noyau lié à la décomposition en valeurs singulières de $\boldsymbol{X}$, nous obtenons alors des garanties fortes sur l'erreur moyenne d'approximation de la sélection d'attributs. Les résultats détaillés sont à retrouver dans l'article long [8].

Le reste du présent article s'organise comme suit. En Partie 2, on présente quelques-uns des algorithmes de sélection existants. En Partie 3, on introduit les PPD. En Partie 4, on introduit un PPD particulier défini via la décomposition en valeurs singulières de $\boldsymbol{X}$ et on liste quelques-unes de ses propriétés théoriques. Enfin, en Partie 5, on valide ces résultats par des simulations.

2 La sélection aléatoire de colonnes

Dans cette partie, on rappelle quelques algorithmes de l'état de l'art après avoir expliqué la forme des résultats théoriques.

2.1 Comment et à quoi se comparer

On compare les algorithmes de sélection de colonnes à l'ACP qui utiliserait un nombre k de directions principales. L'ACP résume les données $\boldsymbol{X}$ par leur meilleure approximation $\Pi_k \boldsymbol{X}$ de rang k. L'adjectif « meilleur » s'entend ici au sens de la norme $\|\cdot\|_2$ ou de la norme de Frobenius $\|\cdot\|_{\mathrm{Fr}}$, deux normes matricielles usuelles [7] donnant le même minimiseur

$$\Pi_k \boldsymbol{X} = \arg\min_{\mathrm{rg}\boldsymbol{P}=k} \|\boldsymbol{X} - \boldsymbol{P}\|_2 = \arg\min_{\mathrm{rg}\boldsymbol{P}=k} \|\boldsymbol{X} - \boldsymbol{P}\|_{\mathrm{Fr}}. \quad (1)$$

En sélection de colonnes, on souhaite comparer $\boldsymbol{X}$ à sa projection sur l'espace engendré par les colonnes sélectionnées.

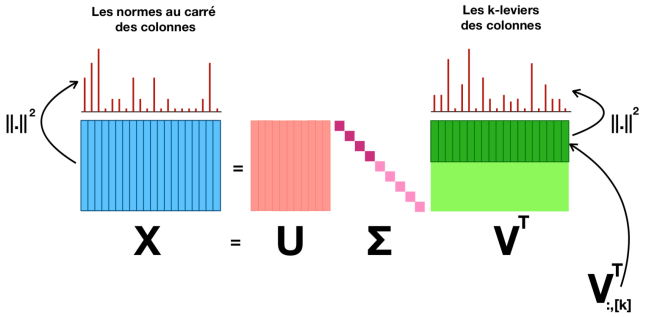

FIGURE 1 – La différence entre les k-leviers et les normes au carré des colonnes.

Concrètement, pour un ensemble $S \subset [d] = \{1, \dots, d\}$ d'indices de colonnes, on restreint le minimum de (1) aux produits $\boldsymbol{P} = \boldsymbol{X}_{:,S}\boldsymbol{B}$, avec $\boldsymbol{B}$ de taille $|S| \times d$. On note alors $\Pi^2_{S,k}\boldsymbol{X}$ et $\Pi^{\mathrm{Fr}}_{S,k}\boldsymbol{X}$ les minimiseurs obtenus pour la norme 2 et de Frobenius, respectivement. Notons que contrairement à la PCA, le choix de la norme a une influence sur la projection.

L'enjeu de la sélection de colonnes est de trouver S de petit cardinal tel que $\|\boldsymbol{X} - \Pi^{\nu}_{S,k}\boldsymbol{X}\|_{\nu}$ est comparable à $\|\boldsymbol{X} - \Pi_k\boldsymbol{X}\|_{\nu}$, pour $\nu \in \{2, \mathrm{Fr}\}$.

2.2 Échantillonnage multinomial

L'algorithme de sélection de colonnes le plus simple consiste à tirer chaque colonne indépendamment avec une probabilité p_i, qui traduit l'importance de la colonne $i \in [d]$ dans la matrice $\boldsymbol{X}$. Une possibilité est d'utiliser directement les normes des colonnes [4]. Une autre possibilité, plus efficace, est de tirer parti des *k-leviers*, ou *k-leverage scores* en anglais [1].

Définition 1 (k-leviers)**.** *Soit $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\intercal \in \mathbb{R}^{N\times d}$ la décomposition en valeurs singulières de $\boldsymbol{X}$. Pour $j \in [d]$, le k-levier de la j-ème colonne de $\boldsymbol{X}$ est défini par*

$$\ell_j^k = \sum_{i=1}^{k} V_{j,i}^2. \tag{2}$$

La Définition 1 est illustrée Figure 1. Le k-levier ℓ_j^k est une mesure de l'alignement du j-ème axe canonique avec les k premières directions principales, cf. Figure 2. Nous reviendrons sur cette interprétation en Partie 4.1. En utilisant une distribution multinomiale de paramètre $\boldsymbol{\ell}^k$ pour sélectionner des colonnes, on favorise donc des colonnes qui sont individuellement proches des axes principaux.

Théorème 1 ([1])**.** *On tire s colonnes i.i.d. de distribution multinomiale de paramètre $\boldsymbol{\ell}^k$. Si $s \geq 3200k^2\epsilon^{-2}$, alors*

$$\mathbb{P}\left(\|\boldsymbol{X} - \Pi^{\mathrm{Fr}}_{S,k}\boldsymbol{X}\|^2_{\mathrm{Fr}} \leq (1+\epsilon)\|\boldsymbol{X} - \Pi_k\boldsymbol{X}\|^2_{\mathrm{Fr}}\right) \geq 0.7. \tag{3}$$

L'échantillonnage multinomial par k-leviers permet donc d'avoir des erreurs d'approximation relatives à l'ACP de rang k arbitrairement petites, mais au prix de conserver un nombre de colonnes largement supérieur à k.

2.3 Échantillonnage volumique

Un autre algorithme consiste à sélectionner exactement k colonnes de façon non indépendante pour éviter toute redondance. C'est le cas de l'échantillonnage volumique, ou *Volume Sampling (VS)*.

Théorème 2 ([3])**.** *Soit S un sous-ensemble de $[d]$ tiré selon*

$$\mathbb{P}_{\mathrm{VS}}(S) \propto \mathrm{Det}(\boldsymbol{X}_{:,S}^\intercal \boldsymbol{X}_{:,S})\, \mathbb{1}_{\{|S|=k\}}. \tag{4}$$

Alors

$$\mathbb{E}_{\mathrm{VS}}\, \|\boldsymbol{X} - \Pi^{\mathrm{Fr}}_{S,k}\boldsymbol{X}\|^2_{\mathrm{Fr}} \leq (k+1)\|\boldsymbol{X} - \Pi_k\boldsymbol{X}\|^2_{\mathrm{Fr}}, \tag{5}$$

et

$$\mathbb{E}_{\mathrm{VS}}\, \|\boldsymbol{X} - \Pi^{2}_{S,k}\boldsymbol{X}\|^2_{2} \leq (d-k)(k+1)\|\boldsymbol{X} - \Pi_k\boldsymbol{X}\|^2_{2}. \tag{6}$$

L'échantillonnage multinomial en Partie 2.2 sélectionnait les colonnes les plus importantes de la matrice $\boldsymbol{X}$ sans tenir compte de leur complémentarité éventuelle. L'échantillonnage volumique favorise quant à lui les sous-ensembles S qui correspondent à des colonnes *diverses* dans $\mathbb{R}^N$. Pour comprendre cette diversité, supposons que les colonnes de $\boldsymbol{X}$ sont des échantillons i.i.d. de moyenne nulle. L'équation (4) dit alors que l'échantillonnage volumique sélectionne S proportionnellement au déterminant de la matrice de covariance des colonnes $\boldsymbol{X}_{:,S}$ indexées par S.

Nous analysons maintenant l'échantillonnage volumique à travers la lentille des processus ponctuels déterminantaux.

3 Processus ponctuels déterminantaux

On rappelle dans cette section quelques notions de la théorie des processus ponctuels déterminantaux discrets (PPD ; [5]).

Définition 2 (PPD)**.** *Soit $\boldsymbol{K} \in \mathbb{R}^{d\times d}$ une matrice semi-définie positive. Un sous-ensemble aléatoire noté $Y \subset [d]$ est tiré d'un PPD de noyau marginal $\boldsymbol{K}$ si et seulement si*

$$\forall S \subset [d], \quad \mathbb{P}(S \subset Y) = \mathrm{Det}(\boldsymbol{K}_S), \tag{7}$$

avec $\boldsymbol{K}_S = [\boldsymbol{K}_{i,j}]_{i,j\in S}$. Par convention $\mathrm{Det}(\boldsymbol{K}_\emptyset) = 1$.

Étant donnée une matrice $\boldsymbol{K}$, (7) ne permet pas nécessairement de définir correctement une distribution de probabilité. Une condition suffisante est que $\boldsymbol{K}$ soit symétrique et de spectre inclus dans $[0, 1]$. Si la matrice $\boldsymbol{K}$ est symétrique, le PPD correspondant est une distribution répulsive dans le sens suivant : pour tout $i, j \in [d]$,

$$\begin{aligned}
\mathbb{P}(\{i,j\} \subset Y) &= K_{i,i}K_{j,j} - K_{i,j}^2 && (8)\\
&= \mathbb{P}(\{i\} \subset Y)\,\mathbb{P}(\{j\} \subset Y) - K_{i,j}^2 && (9)\\
&\leq \mathbb{P}(\{i\} \subset Y)\,\mathbb{P}(\{j\} \subset Y). && (10)
\end{aligned}$$

Dans le cas extrême où le spectre de $\boldsymbol{K}$ est inclus dans $\{0, 1\}$, le noyau et le PPD correspondants sont dits *de projection*. Les PPD de projection ont la particularité de tirer des échantillons Y qui ont toujours le même cardinal avec probabilité 1, égal au nombre des valeurs propres de $\boldsymbol{K}$ égales à 1. Ces PPD

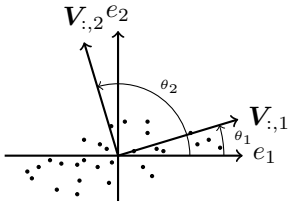

FIGURE 2 – Interprétation géométrique des k-leviers dans le cas $k = 1 : \ell_1^k = \cos^2\theta_1$, $\ell_2^k = \cos^2\theta_2$.

de projection apparaissent en filigrane dans l'échantillonnage volumique de la Partie 2.3, qui peut-être vu comme un mélange de PPD de projection. En effet, si on note $\lambda_1 \geq \cdots \geq \lambda_d$ les valeurs propres de $\boldsymbol{X}^\intercal \boldsymbol{X}$ et r le rang de $\boldsymbol{X}$, on a [5] :

$$\mathbb{P}_{\mathrm{VS}}(Y = S) = \sum_{\substack{T\subset[r]\\|T|=k}} \mu_T \left[\frac{1}{k!}\operatorname{Det}\left(\boldsymbol{V}_{T,S}\boldsymbol{V}_{T,S}^\intercal\right)\right] \tag{11}$$

avec

$$\mu_T \propto \prod_{i\in T}\lambda_i, \quad \sum_{\substack{T\subset[r]\\|T|=k}} \mu_T = 1. \tag{12}$$

Lorsque d est grand devant N, le coût du calcul de la matrice $\boldsymbol{V}$ est en Nd^2. Une fois $\boldsymbol{V}$ calculée, le coût d'échantillonnage d'un PPD du mélange (11) est $\mathcal{O}(dk^2)$. Une analyse plus détaillée du coût d'échantillonnage est fournie dans [5] ou [2].

4 Un PPD bien choisi

On considère l'algorithme de sélection de colonnes suivant : on tire les colonnes S avec une probabilité

$$\mathbb{P}(S) \propto \operatorname{Det} \boldsymbol{V}_{S,[k]}(\boldsymbol{V}_{S,[k]})^\intercal \tag{13}$$

Cet échantillonnage correspond au PPD de projection de noyau $\boldsymbol{K} = \boldsymbol{V}_{:,[k]}(\boldsymbol{V}_{:,[k]})^\intercal$, qui est aussi le PPD de plus grand poids dans le mélange (11).

4.1 Interprétation géométrique

Pour $j \in [d]$, soit $\boldsymbol{e}_j$ le j-ème vecteur de la base canonique de $\mathbb{R}^d$ et soit θ_j l'angle entre $\boldsymbol{e}_j$ et le sous-espace $\mathcal{P}_k = \mathrm{Vect}(\boldsymbol{V}_{:,[k]})$. Soit $\Pi_{\mathcal{P}_k}\boldsymbol{e}_j$ la projection orthogonale de $\boldsymbol{e}_j$ sur $\mathcal{P}_k$. Alors

$$\cos^2(\theta_j) := \frac{(\boldsymbol{e}_j, \Pi_{\mathcal{P}_k}\boldsymbol{e}_j)^2}{\|\Pi_{\mathcal{P}_k}\boldsymbol{e}_j\|^2} = (\boldsymbol{e}_j, \Pi_{\mathcal{P}_k}(\boldsymbol{e}_j)) \tag{14}$$

$$= (\boldsymbol{e}_j, \sum_{i=1}^k V_{j,i}\boldsymbol{V}_{:,i}) = \sum_{i=1}^k V_{j,i}^2 = \ell_j^k. \tag{15}$$

Un k-levier élevé indique que $\boldsymbol{e}_j$ est presque aligné avec $\mathcal{P}_k$. La sélection des colonnes qui ont un k-levier élevé comme dans le Théorème 1 peut donc être vue comme le remplacement du sous-espace principal $\mathcal{P}_k$ par un sous-espace formé par des colonnes qui sont individuellement alignées avec $\mathcal{P}_k$. Pour éviter la redondance entre colonnes, on peut mesurer la ressemblance entre $\mathcal{P}_k$ et $\mathcal{P}_S = \mathrm{Vect}(\boldsymbol{e}_j, j \in S)$ d'une façon naturelle à l'aide des *angles principaux* [7, Section 6.4.4]. Formellement,

$$\cos^2(\mathcal{P}_k, \mathcal{P}_S) := \operatorname{Det}(\boldsymbol{V}_{S,[k]})^2. \tag{16}$$

Ainsi, le PPD de noyau (4) est une généralisation naturelle du tirage multinomial avec les k-leviers : selon (7), chaque colonne est toujours marginalement tirée proportionnellement à son k-levier, mais on ajoute la contrainte que le sous-ensemble $\mathcal{P}_S$ soit une bonne approximation de $\mathcal{P}_k$.

4.2 Bornes sous hypothèse de parcimonie

Avant de présenter les résultats, on définit la parcimonie des k-leviers comme le nombre de k-leviers non-nuls :

$$p = \left|\{j \in [d], \ell_j^k \neq 0\}\right|. \tag{17}$$

La décroissance des valeurs singulières $\sigma_1 \geq \cdots \geq \sigma_r$ de $\boldsymbol{X}$ est caractérisée par

$$\beta = \sigma_{k+1}^2 \left(\frac{1}{d-k}\sum_{j\geq k+1}\sigma_j^2\right)^{-1} \in [1, d-k]. \tag{18}$$

Ces quantités contrôlent la qualité de l'approximation sous la loi du PPD de noyau marginal (4). La parcimonie p des k-leviers reflète le nombre de coordonnées initiales qui influent le calcul des axes principaux. La constante β reflète l'importance de σ_{k+1}^2 relativement à la moyenne des valeurs singulières au carré $(\sigma_i^2)_{i\geq k+1}$. Une grande valeur de β indique une coupure entre les directions principales d'ordre $> k+1$ relativement à celles d'ordre $\leq k+1$. Notons que cette constante β n'intervient ci-après que dans les garanties en norme de Frobenius. Nous énonçons maintenant nos résultats.

Proposition 1. *Sous la loi du PPD de noyau marginal $\boldsymbol{K}$,*

$$\mathbb{E}_{\mathrm{DPP}}\|\boldsymbol{X} - \Pi_{S,k}^2\boldsymbol{X}\|_2^2 \leq (1 + (p-k)k)\|\boldsymbol{X} - \Pi_k\boldsymbol{X}\|_2^2 \tag{19}$$

et

$$\mathbb{E}_{\mathrm{DPP}}\|\boldsymbol{X} - \Pi_{S,k}^{\mathrm{Fr}}\boldsymbol{X}\|_{\mathrm{Fr}}^2 \leq \left(1 + \beta\frac{p-k}{d-k}k\right)\|\boldsymbol{X} - \Pi_k\boldsymbol{X}\|_{\mathrm{Fr}}^2. \tag{20}$$

Notre borne spectrale (19) est meilleure que la borne spectrale (6) de l'échantillonnage volumique d'un facteur $(p-k)/(d-k)$. Plus le profil des k-leviers est parcimonieux, meilleure est la qualité de l'approximation. Quand il n'y a pas de coupure de spectre ($\beta \approx 1$), la borne (20) est meilleure que la borne de Frobenius (5) de l'échantillonnage volumique.

4.3 Cas de la parcimonie approchée

En général, aucun des k-leviers n'est strictement nul et la constante p est égale à d. Toutefois, il est courant qu'un grand nombre de k-leviers soient très petits, voir [8] pour des exemples concrets. De la même manière que l'ACP tire parti d'une dimension effective petite d'un problème, nous souhaitons un résultat qui profite de cette quasi-parcimonie des k-leviers et qui rende compte de l'amélioration observée expérimentalement par rapport à l'échantillonnage volumique.

On définit la parcimonie effective de la façon suivante. Soit π une permutation de $[d]$ telle que les k-leviers soient ordonnés :

$$\ell_{\pi_1}^k \geq \ell_{\pi_2}^k \geq \ldots \geq \ell_{\pi_d}^k. \tag{21}$$

Soit $\delta \in [d]$, et $T_\delta = [\pi_\delta, \dots, \pi_d]$. Soit $\theta \geq 1$ et

$$p_{\text{eff}}(\theta) = \min \left\{ q \in [d] \Big/ \sum_{i \leq q} \ell^k_{\pi_i} \geq k - 1 + \frac{1}{\theta} \right\}. \quad (22)$$

Sachant que par construction, le vecteur $\boldsymbol{\ell}^k$ somme à k, le paramètre θ contrôle dans (22) le défaut de somme cumulée des $p_{\text{eff}}(\theta)$ plus grands k-leviers.

Proposition 2. *Soit $\mathcal{A}_\theta$ l'événement $\{S \cap T_{p_{\text{eff}}(\theta)} = \emptyset\}$. Alors,*

$$\mathbb{P}_{\text{DPP}}(\mathcal{A}_\theta) \geq \frac{1}{\theta}, \quad (23)$$

$$\mathbb{E}_{\text{DPP}}\left[\|\boldsymbol{X} - \Pi^2_{S,k}\boldsymbol{X}\|_2^2 \,\middle|\, \mathcal{A}_\theta\right] \leq (1 + (p_{\text{eff}}(\theta) - k + 1)(k - 1 + \theta))\|\boldsymbol{X} - \Pi_k \boldsymbol{X}\|_2^2, \quad (24)$$

et

$$\mathbb{E}_{\text{DPP}}\left[\|\boldsymbol{X} - \Pi^{\text{Fr}}_{S,k}\boldsymbol{X}\|_{\text{Fr}}^2 \,\middle|\, \mathcal{A}_\theta\right] \leq \left(1 + \beta \frac{(p_{\text{eff}}(\theta) + 1 - k)}{d - k}(k - 1 + \theta)\right) \|\boldsymbol{X} - \Pi_k \boldsymbol{X}\|_{\text{Fr}}^2. \quad (25)$$

Dans la Proposition 2, la parcimonie effective $p_{\text{eff}}(\theta)$ remplace la parcimonie p de la Proposition 1. L'idée est de conditionner le PPD à ne pas contenir les colonnes qui ont des k-leviers faibles, représentées par l'ensemble de « mauvais » indices $T_{p_{\text{eff}}(\theta)}$. En pratique, le conditionnement peut se réaliser à l'aide d'une étape de réjection : on tire des sous ensemble $S \sim \text{DPP}(\boldsymbol{K})$ jusqu'à réaliser la condition $S \cap T_{p_{\text{eff}}}(\theta) = \emptyset$. Le temps d'attente moyen est alors contrôlé par (23).

Enfin, on a présenté ici des garanties en termes de normes matricielles de l'erreur d'approximation. Lorsqu'on connaît les étapes d'apprentissage qui vont être réalisées après la réduction de dimension, l'objectif peut changer. Nous proposons dans [8] des garanties sur l'erreur de prédiction d'une régression linéaire après sélection de colonnes.

5 Simulations numériques

Cette partie illustre les résultats théoriques à l'aide de simulations numériques sur des jeux de donnés synthétiques. On compare l'erreur d'approximation $\mathbb{E}\|\boldsymbol{X} - \Pi^{\text{Fr}}_S \boldsymbol{X}\|^2_{\text{Fr}}$ divisée par celle de l'ACP pour l'échantillonnage volumique et l'échantillonnage avec PPD de noyau (4). Pour cela, on génère des matrices $\boldsymbol{X}$ de petite taille $N \times d = 100 \times 20$, si bien qu'on peut énumérer toutes les combinaisons de $k = 5$ colonnes. On génère 200 telles matrices aléatoires $\boldsymbol{X}$ avec des profils de k-leviers parcimonieux, à l'aide d'un algorithme dont la description détaillée est fournie dans l'article [8]. Sur la Figure 3, pour chaque matrice on représente deux croix : en bleu l'erreur d'approximation de l'échantillonnage volumique, en rouge celle du PPD. En trait plein bleu, la borne théorique (5) de l'échantillonnage volumique est indépendante du niveau de parcimonie p, tandis que la borne théorique du PPD dépend linéairement de p. On remarque que les bornes

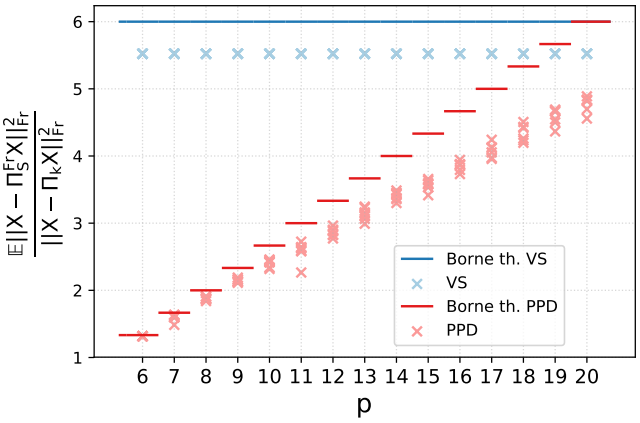

FIGURE 3 – L'erreur d'approximation $\mathbb{E}\|\boldsymbol{X} - \Pi^{\text{Fr}}_{S,k}\boldsymbol{X}\|^2_{\text{Fr}}$ et la borne théorique pour l'échantillonnage volumique (bleu) et le PPD (rouge), en fonction de la parcimonie p. La valeur 1 en ordonnée correspond à l'erreur de l'ACP.

théoriques reflètent bien les comportements empiriques des deux approches. La présence de parcimonie est favorable au PPD, comme attendu. D'autres simulations et exemples réels sont disponibles dans [8].

6 Conclusion

On a proposé un nouvel algorithme de sélection aléatoire de colonnes utilisant un processus ponctuel déterminantal inspiré de l'ACP. On a présenté des garanties théoriques de cet algorithme, qui dépassent l'état de l'art dans les cas où les k-leviers sont parcimonieux. Les expériences numériques confirment la précision des bornes obtenues.

Références

[1] Drineas, P. and Mahoney, M. W. and Muthukrishnan, S. *Relative-error CUR matrix decompositions*. SIAM J MATRIX ANAL A, 2008.

[2] Tremblay, N. and Barthelme, S. and Amblard, P.O. *Optimized algorithms to sample determinantal point processes*. ArXiv preprint, arXiv :1802.08471, 2018.

[3] Deshpande, A. and Rademacher, L. and Vempala, S. and Wang, G. *Matrix Approximation and Projective Clustering via Volume Sampling*. Proc. ACM-SIAM SODA, 2006.

[4] Drineas, P. and Frieze, A. and Kannan, R. and Vempala, S. and Vinay, V. *Clustering Large Graphs via the Singular Value Decomposition*. Mach. Learn., 2004.

[5] Kulesza, A. and Taskar, B. *Determinantal point processes for machine learning*. FTML, 2012.

[6] Macchi, O. *The coincidence approach to stochastic point processes*. Adv. Appl. Probab., 1975.

[7] Golub, G. H. and Van Loan, C. F. *Matrix Computations (3rd Ed.)* Johns Hopkins University Press, 1996

[8] Belhadji, A. and Bardenet, R. and Chainais, P. *A determinantal point process for column subset selection*. ArXiv preprint, arXiv :1812.09771, 2018.