

Estimation statistique d’une large famille de distances entre matrices de covariance dans le régime des grandes matrices aléatoires

Malik TIOMOKO^{1,2}, Romain COUILLET^{2,1}

¹Laboratoire des Signaux et Systèmes, CentraleSupélec, 91192 Gif-Sur-Yvette

²Gipsa-Lab, 11 Rue des Mathématiques, 38400 Saint-Martin-d’Hères

malik.tiomoko@gipsa-lab.grenoble-inp.fr, tiomoko_malik@yahoo.fr
romain.couillet@gipsa-lab.grenoble-inp.fr, romain.couillet@centralesupelec.fr

Résumé – De nombreuses applications en traitement du signal et des données, dans des domaines aussi variés que l’imagerie hyperspectrale ou les signaux cérébraux, reposent sur l’estimation de distances entre matrices de covariance. En pratique, les estimateurs classiques substituent les covariances par leurs estimateurs empiriques, idéalement obtenus à partir de nombreuses observations. Cependant, dans les applications modernes où les données sont potentiellement peu nombreuses mais de grandes dimensions, ces estimateurs sont biaisés et induisent des erreurs d’approximation parfois dramatiques. Dans cet article, nous nous reposons sur des outils avancés de la théorie des matrices aléatoires pour fournir des estimateurs consistents de la distance entre matrices de covariance pour une grande famille de métrique, avec une emphase particulière sur la très populaire distance de Fisher. Une application au regroupement spectral de matrices de covariances démontre la puissance de nos estimateurs.

Abstract – Many machine learning and signal processing applications, in fields as diverse as hyperspectral image or brain signal processing, rely on the statistical estimation of the distance between covariance matrices. In practice, standard estimates simply replace the unknown population covariances by sample covariances ideally obtained from numerous independent observations. However in modern applications where data are possibly few and large dimensional, those estimators are biased and induce dramatic approximation errors. In this article, based on advanced tools in random matrix theory, we provide consistent estimates of the distance between covariance matrices for a large family of metrics, with a particular emphasis on the popular Fisher distance. An application to covariance-based spectral clustering supports the strength of our estimators.

1 Introduction

Les statistiques mesurant la distance entre distributions de probabilité sont des outils très utilisés en apprentissage automatique et en traitement du signal. En particulier, pour des applications où l’information de discrimination des distributions est contenue dans les statistiques de deuxième ordre des distributions multivariées (matrices de covariance), la distance se résume en une métrique entre matrices de covariance. De nombreuses applications relèvent de ce contexte (classification d’images hyperspectrales ou SAR, de signaux EEG, etc.).

Si on définit par C_1 et C_2 deux matrices de covariance de taille $p \times p$, on constate que la plupart des distances usuelles, que l’on dénotera génériquement $D(C_1, C_2)$, s’expriment comme des fonctionnelles des valeurs propres de la matrice $C_1^{-1}C_2$ (distance de Fisher, distance de Bhattacharyya, divergence de Kullback Leibler ou de Rényi entre gaussiennes centrées) ou des valeurs propres de la matrice C_1C_2 (distance de Wasserstein entre deux gaussiennes centrées). Dans l’hypothèse où le nombre d’échantillons n_1 et n_2 de données ayant C_1 et C_2 pour covariance est très grand devant p , la loi des grands nombres garantit que $D(\hat{C}_1, \hat{C}_2)$ est un estimateur consistant pour $D(C_1, C_2)$ avec $\hat{C}_a = \frac{1}{n_a} \sum_{i=1}^{n_a} x_i^{(a)} x_i^{(a)T}$ pour $a \in \{1, 2\}$ la matrice de covariance empirique des n_a

échantillons centrés $x_i^{(a)}$. Cependant, cet estimateur est fortement biaisé lorsque $n_1, n_2 \sim p$.

A l’aide d’outils de la théorie des matrices aléatoires, cet article propose une formule générale d’un estimateur “universel” des distances $D(C_1, C_2)$ qui est consistant dans la limite où $p, n_1, n_2 \rightarrow \infty$ avec $p/n_1 \rightarrow c_1 > 0$ et $p/n_2 \rightarrow c_2 > 0$.

Ces résultats s’inspirent des travaux de Mestre [1] sur l’estimation de fonctionnelles des valeurs propres $\frac{1}{p} \sum_{i=1}^p f(\lambda_i(C))$ de matrices de covariance C . La procédure que nous suivons ici est la suivante : (i) la quantité d’intérêt $D(C_1, C_2)$ est exprimée comme une intégrale complexe faisant intervenir la transformée de Stieltjes de la mesure des valeurs propres de $C_1^{-1}C_2$ (ou C_1C_2); (ii) cette mesure est reliée asymptotiquement à la mesure limite des valeurs propres de $\hat{C}_1^{-1}\hat{C}_2$ (ou $\hat{C}_1\hat{C}_2$) en utilisant les travaux de [2]; (iii) on obtient alors un estimateur sous forme d’intégrale complexe qu’il s’agit d’évaluer pour chaque fonction f d’intérêt. Cependant, au contraire de [1] qui s’intéresse à des fonctions f simples, les distances $D(C_1, C_2)$ d’intérêt ici impliquent des logarithmes et racines carrées qui demandent un travail fin d’analyse complexe (notamment un traitement précis des “coupures”).

Reproductibilité. Codes Matlab de nos algorithmes :
<https://github.com/maliktiomoko/RMTEstimCovDist>.

Métriques	ℓ_i	$f(z)$
d_R^2	ℓ_i^-	$\log^2(z)$
d_B^2	ℓ_i^-	$-\frac{1}{4}\log(z) + \frac{1}{2}\log(1+z) - \frac{1}{2}\log(2)$
δ_{KL}	ℓ_i^-	$\frac{1}{2}z - \frac{1}{2}\log(z) - \frac{1}{2}$
$\delta_{\alpha R}$	ℓ_i^-	$\frac{-1}{2(\alpha-1)}\log(\alpha + (1-\alpha)z) + \frac{1}{2}\log(z)$
d_W	ℓ_i^+	$-2\sqrt{z}$

TABLE 1 – Métriques et fonctions $f(z)$ associées.

2 Modèle et Hypothèses

Pour $a \in \{1, 2\}$, considérons n_a vecteurs $X_a = [x_1^{(a)}, \dots, x_{n_a}^{(a)}]$ indépendants et identiquement distribués avec $x_i^{(a)} = C_a^{\frac{1}{2}} \tilde{x}_i^{(a)}$ où $\tilde{x}_i^{(a)} \in \mathbb{R}^p$ est de moyenne nulle, de matrice de covariance identité et de moment d'ordre quatre fini. Pour des raisons d'analyse, nous considérons par ailleurs l'hypothèse suivante d'accroissement des valeurs de n_a et p .

Hypothèse 1 Lorsque $n_a \rightarrow \infty$, $p/n_a \rightarrow c_a \in (0, 1)$ et $\limsup_a \max\{\|C_a\|^{-1}, \|C_a\|\} < \infty$ avec $\|\cdot\|$ la norme opérateur.

Les quantités d'intérêt dans cet article sont les distances entre les matrices de covariance C_1 et C_2 :

$$D(C_1, C_2, f) = \frac{1}{p} \sum_{i=1}^p f(\ell_i) \quad (1)$$

où $\ell_i = \ell_i^-$ sont les valeurs propres de la matrice $C_1^{-1}C_2$ ou $\ell_i = \ell_i^+$ celles de C_1C_2 , selon la distance considérée. Les distances de Fisher-Rao d_R , de Bhattacharyya d_B , de Wasserstein d_W ainsi que les divergences de Kullbach Leibler d_{KL} ou de Rényi $\delta_{\alpha R}$ entrent dans ce cadre. La Table 1 présente les fonctions f correspondant à ces distances.¹

Pour estimer (1), nous connectons la mesure spectrale $\nu_p = \frac{1}{p} \sum_{i=1}^p \delta_{\ell_i}$ des valeurs propres de $C_1^{-1}C_2$ (ou C_1C_2) à la mesure spectrale $\mu_p = \frac{1}{p} \sum_{i=1}^p \delta_{\lambda_i}$ des valeurs propres $\lambda_i = \lambda_i^-$ de $\hat{C}_1^{-1}\hat{C}_2$ (ou $\lambda_i = \lambda_i^+$ de $\hat{C}_1\hat{C}_2$) à travers leurs transformées de Stieltjes respectives, définies, pour une mesure de probabilité θ à support dans \mathbb{R} et pour tout $z \in \mathbb{C} \setminus \text{supp}(\theta)$, par

$$m_\theta(z) = \int \frac{\theta(dt)}{t-z}.$$

En anticipation des calculs d'intégration complexe, notons que $m_{\mu_p}(z) = \frac{1}{p} \sum_{i=1}^p \frac{1}{\lambda_i - z}$ est une fonction rationnelle.

3 Principaux Résultats

3.1 Forme intégrale de l'estimateur proposé

Le Théorème 1 présente le nouvel estimateur proposé sous une forme intégrale, valable pour toute fonction f admettant une extension complexe.

1. La distance d_W intègre en fait le terme additionnel $\frac{1}{p} \text{tr}(C_1 + C_2)$ qui est facilement estimé par ailleurs.

Théorème 1 Soit $\Gamma \subset \{z \in \mathbb{C}, \text{real}[z] > 0\}$ un contour encerclant toutes les valeurs propres de $\hat{C}_1^{-1}\hat{C}_2$ (respectivement $\hat{C}_1\hat{C}_2$). Alors, sous l'Hypothèse 1,

$$D(C_1, C_2; f) - \hat{D}(X_1, X_2; f) \xrightarrow{p.s.} 0$$

$$\hat{D}(X_1, X_2; f) = \frac{1}{2\pi i} \oint_{\Gamma} f \left(\frac{\varphi_p(z)}{\psi_p(z)} \right) \left[\frac{\varphi_p'(z)}{\varphi_p(z)} - \frac{\psi_p'(z)}{\psi_p(z)} \right] \frac{\psi_p(z) dz}{c_2}.$$

avec $\psi_p(z) = 1 - \frac{p}{n_2} - \frac{p}{n_2} z m_{\mu_p}(z)$ et $\varphi_p(z) = z(1 - \frac{p z m_{\mu_p}(z)}{n_1})$ (respectivement $\varphi_p(z) = \frac{z}{1 - \frac{p}{n_1} - \frac{p}{n_1} z m_{\mu_p}(z)}$).

En dépit du caractère obscur de cette intégrale, on lit très bien l'influence individuelle et croisée des données $x_i^{(1)}$ et $x_i^{(2)}$ à travers les fonctions $\varphi_p(z)$ et $\psi_p(z)$ (qui dépendent toutes deux de $m_{\mu_p}(z)$ qui "croise" les données mais qui reposent aussi indépendamment sur n_1 et n_2). Il est aussi intéressant de noter que la formule générale ne diffère, pour les estimateurs des fonctionnelles de ℓ_i^- ou ℓ_i^+ , que par l'expression de $\varphi_p(z)$.

Esquisse de preuve. Les idées de la preuve sont similaires pour les cas C_1C_2 et $C_1^{-1}C_2$; nous ne traitons ici que le cas $C_1^{-1}C_2$. Par le théorème d'intégration de Cauchy, la métrique souhaitée s'écrit

$$\begin{aligned} D(C_1, C_2, f) &= \frac{1}{p} \sum_{i=1}^p f(\lambda_i) = \int f(t) \nu_p(dt) \\ &= \frac{1}{2\pi i} \int \left[\oint_{\Gamma_{\nu_p}} \frac{f(z)}{z-t} dz \right] \nu_p(dt) \\ &= \frac{1}{2\pi i} \oint_{\Gamma_{\nu_p}} f(z) m_{\nu_p} dz \quad (2) \end{aligned}$$

avec Γ_{ν_p} un contour encerclant le support de ν_p . Pour aller au delà (2), l'idée est d'établir un lien entre m_{ν_p} (qui reste inaccessible, $C_1^{-1}C_2$ étant inconnu) et m_{μ_p} (qui est observable car lié à $\hat{C}_1^{-1}\hat{C}_2$). On procède ici en remarquant que $\hat{C}_1^{-\frac{1}{2}}\hat{C}_2\hat{C}_1^{-\frac{1}{2}}$ peut être vu comme une matrice de covariance empirique (en conditionnant sur \hat{C}_1) dont la matrice de population est elle-même une matrice de covariance empirique indépendante. En utilisant deux fois [2], qui relie les mesures spectrales limites de matrices de population et de covariance empirique, on établit ce lien à travers une transformée de Stieltjes intermédiaire m_{ζ_p} (avec ζ_p la mesure spectrale de $\hat{C}_1^{-\frac{1}{2}}\hat{C}_2\hat{C}_1^{-\frac{1}{2}}$)

$$m_{\mu_p}(z) = \varphi_p(z) m_{\zeta_p}(\varphi_p(z)) + o_p(1) \quad (3)$$

$$m_{\nu_p} \left(\frac{z}{\Psi_p(z)} \right) = m_{\zeta_p}(z) \Psi_p(z) + o_p(1) \quad (4)$$

où $\Psi_p(z) \equiv 1 - \frac{p}{n_2} - \frac{p}{n_2} z m_{\zeta_p}(z)$. Le résultat du théorème s'obtient alors en appliquant deux changements de variables successifs en (2) ($z \mapsto z/\Psi_p(z)$ et $z \mapsto \varphi_p(z)$) en utilisant ces deux relations. La clé du résultat repose cependant sur un contrôle précis de la validité de ces changements de variables et du contour Γ résultant. Les détails techniques sont disponibles dans [3]. Il apparait en particulier que les hypothèses $c_1, c_2 < 1$ sont cruciales pour la validité du changement de variable.

Notons que le Théorème 1 peut être étendu au cas où l'une des matrices, disons C_1 , est connue, ce qui correspond à prendre la limite $c_1 \rightarrow 0$. On obtient alors le corollaire suivant :

Corollaire 1 *Sous l'Hypothèse 1 avec $c_1 \rightarrow 0$, pour Γ un contour encerclant les valeurs propres de $C_1^{-1}\hat{C}_2$ (resp., $C_1\hat{C}_2$),*

$$D(C_1, C_2; f) - \hat{D}(X_2; f) \xrightarrow{p.s.} 0$$

$$\hat{D}(X_2; f) = \frac{1}{2\pi i} \oint_{\Gamma} F(-m_{\tilde{\mu}_p}(z)) dz.$$

avec $F'(z) = f(\frac{1}{z})$, $m_{\tilde{\mu}_p}(z) = \frac{p}{n_2} m_{\mu_p}(z) - (1 - \frac{p}{n_2}) \frac{1}{z}$ et ici $m_{\mu_p}(z) = \frac{1}{p} \sum_{i=1}^p \frac{1}{\lambda_i - z}$ où $\lambda_i = \lambda_i(C_1^{-1}\hat{C}_2)$ (resp., $\lambda_i = \lambda_i(C_1\hat{C}_2)$).

Grâce à ce résultat, nous proposons dans [4] un nouvel estimateur de la matrice de covariance (C_2) basé sur l'approximation $\arg \min_M D(M, C_2; f) \simeq \arg \min_M \hat{D}(X_2; f)$.

3.2 Expression analytique de l'estimateur

En plus d'être peu intuitif, l'estimateur du Théorème 1 requiert une évaluation numérique de l'intégrale complexe et un choix arbitraire d'un contour, qui sont autant de limitations pratiques.

Lorsque $f(z)$ est une fonction élémentaire (z , $\log(z)$, $\log^2(z)$, $\log(1+z)$, \sqrt{z} dont les combinaisons linéaires forment les distances listées en Table 1), il est possible d'obtenir une expression analytique de ces estimateurs. Comme l'intégrale du Théorème 1 ne fait intervenir que des fonctions rationnelles, à l'exception éventuelle de f elle-même, le cas $f(z) = z$ est aisément traité par calcul de résidu. Les cas impliquant logarithmes ou racines carrées sont plus délicats et requièrent une analyse fine des dites "coupures" de l'intégrande.

Pour donner un exemple particulièrement éclairant, nous discuterons ici du cas $f(z) = \log^2(z)$ pour $C_1^{-1}C_2$, cas directement lié à la distance de Fisher, populaire en géométrie de l'information. Le Théorème 2 fournit une forme explicite de l'estimateur pour ce cas. Dans [3, 4] nous fournissons les résultats pour toutes les fonctions f listées précédemment, permettant ainsi que couvrir toutes les métriques discutées jusqu'ici.

Théorème 2 *Soit $\lambda = (\lambda_1, \dots, \lambda_p)^T$ ($\lambda_i = \lambda_i^+$) dans l'ordre croissant et $\Lambda = \text{diag}(\lambda)$. Dénotons $\eta \in \mathbb{R}^p$ les valeurs propres dans l'ordre croissant de $\Lambda + \frac{1}{n_1 - p} \sqrt{\lambda} \sqrt{\lambda}^T$ et $\xi \in \mathbb{R}^p$ les valeurs propres dans l'ordre croissant de $\Lambda - \frac{1}{n_2} \sqrt{\lambda} \sqrt{\lambda}^T$. Alors, sous l'Hypothèse 1,*

$$\int \log^2(t) \nu_p(dt) - \left[\frac{1}{p} \sum_{i=1}^p \log^2((1 - c_1)\lambda_i) - \frac{2}{p} (\Delta_{\xi}^{\eta})^T N_{1,p} \right. \\ \left. + 2 \frac{c_1 + c_2 - c_1 c_2}{c_1 c_2} \left\{ (\Delta_{\xi}^{\eta})^T M (\Delta_{\lambda}^{\eta}) + (\Delta_{\lambda}^{\eta})^T r \right\} \right. \\ \left. - 2 \frac{1 - c_2}{c_2} \left\{ \frac{1}{2} \log^2((1 - c_1)(1 - c_2)) + (\Delta_{\xi}^{\eta})^T r \right\} \right] \xrightarrow{p.s.} 0$$

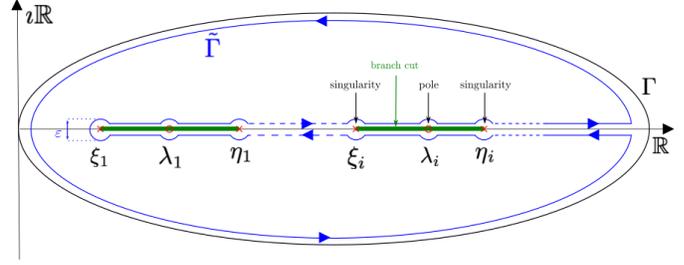


FIGURE 1 – Déformation de Γ pour exclure les coupures de l'intérieur du contour.

où nous définissons $1_p \in \mathbb{R}^p$ le vecteur composé de 1, Δ_a^b le vecteur tel que $(\Delta_a^b)_i = b_i - a_i$ pour $i, j \in \{1, \dots, p\}$, et

$$M_{ij} = \begin{cases} \frac{\frac{\lambda_i}{\lambda_j} - 1 - \log(\frac{\lambda_i}{\lambda_j})}{(\lambda_i - \lambda_j)^2} & , i \neq j \\ \frac{1}{2\lambda_i^2} & , i = j \end{cases},$$

$$N_{ij} = \begin{cases} \frac{\log(\frac{\lambda_i}{\lambda_j})}{\lambda_i - \lambda_j} & , i \neq j \\ \frac{1}{\lambda_i} & , i = j. \end{cases}, \text{ and } r_i = \frac{\log((1 - c_1)\lambda_i)}{\lambda_i}.$$

Esquisse de preuve. Les coupures de la fonction complexe $\log^2(\omega)$ interviennent lorsque ω est réel négatif. L'analyse fonctionnelle de $\frac{\varphi_p}{\psi_p}$ (l'argument de \log^2 dans le Théorème 1) révèle des coupures au niveau des segments $[\xi_i, \eta_i]$ où ξ_i et η_i sont les zéros respectifs de $\psi_p(z)$ et $\varphi_p(z)$. Ces zéros sont entrelacés avec les λ_i comme illustré dans la Figure 1. Ces segments se trouvant à l'intérieur du contour d'intégration Γ , il est nécessaire de déformer Γ en un contour $\tilde{\Gamma}$ qui les exclut de l'intérieur de Γ comme montré dans la figure. Parmi les parties du nouveau contour à intégrer, les plus intéressantes sont les segments (réels) qui longent les coupures. Les intégrales réelles résultant font intervenir des formes du type $\int \frac{\log(x-a)}{x-b} dx$ dont la solution est liée à la fonction dilogarithme $\text{Li}_2(x)$ [5]. En exploitant les identités fonctionnelles du dilogarithme (on a par exemple $\text{Li}_2(z) + \text{Li}_2(\frac{1}{z}) = -\frac{\pi^2}{6} - \frac{1}{2} \log^2(-z)$), on parvient à une forme simplifiée. Dans la limite des grands p , les $\text{Li}_2(x)$ résultants s'approximent au premier ordre et donnent finalement accès au résultat présenté.

4 Simulations et Applications

4.1 Confirmation des résultats

Nous confirmons dans un premier temps la consistance de l'estimateur proposé à travers des simulations sur données synthétiques (vecteurs gaussiens). La Table 2 compare les performances de l'estimateur du Théorème 2 à la méthode classique consistant à substituer C_a par \hat{C}_a dans l'expression de la distance. Bien que l'estimateur soit construit dans le régime de l'Hypothèse 1, à savoir pour p, n grands, il est remarquable que l'estimateur démontre des performances accrues dès $p = 2$.

p	$D_F(C_1, C_2)$	Classique	Proposé
2	0.0980	0.1002	0.0973
4	0.1456	0.1520	0.1461
8	0.1694	0.1820	0.1703
16	0.1812	0.2081	0.1845
32	0.1872	0.2363	0.1886
64	0.1901	0.2892	0.1920
128	0.1916	0.3955	0.1934
256	0.1924	0.6338	0.1942
512	0.1927	1.2715	0.1953

(error > 50%) (error > 100%) (error > 500%)

TABLE 2 – Estimateur proposé et classique de la distance de Fisher $D(C_1, C_2, \log^2)$ pour $[C_1^{-\frac{1}{2}}C_2C_1^{-\frac{1}{2}}]_{ij} = .3^{|i-j|}$, $x_i^{(a)} \sim \mathcal{N}(0, C_a)$; $n_1 = 1024$ et $n_2 = 2048$, pour différents p . Moyennage obtenu sur 10 000 essais.

4.2 Application à la classification

Nous démontrons ici les performances de l'estimateur proposé dans un contexte de classification par matrices de covariances. Précisément, considérons $m = 200$ jeux de données matricielles X_1, \dots, X_m à classer. Chaque $X_i \in \mathbb{R}^{p \times n_i}$ est généré à partir de n_i échantillons i.i.d. de taille p . La moitié $1 \leq i \leq m/2$ de ces données X_i ont des colonnes de covariance $C^{(i)} = C_1 = I_p$ et l'autre moitié des colonnes de covariance $[C^{(i)}]_{ab} = [C_2]_{ab} = 0.05^{|a-b|}$. Il s'agit ici de regrouper les m échantillons X_i de manière non-supervisée à partir du regroupement spectral basé sur le noyau

$$K = \exp\left(-\frac{1}{2}D(C^{(i)}, C^{(j)}; \log^2)\right).$$

Les $C^{(i)}$ étant inconnus, on estime K via

$$\check{K}_{ij} = \exp\left(-\frac{1}{2}\check{D}(X_i, X_j; \log^2)\right)$$

pour $\check{D}(X_i, X_j; \log^2)$ soit l'estimateur classique, soit l'estimateur $\hat{D}(X_i, X_j; \log^2)$ proposé ici. La Figure 2 représente les deux vecteurs propres dominants de \check{K} dans ces deux cas, lorsque les nombres d'échantillons n_i des X_i sont tous égaux à $n_i = 512$ à l'exception de $n_m = 256$ et pour des matrices de covariance C_1, C_2 très proches (et donc difficiles à discriminer). On remarque dans cette figure que l'estimateur classique induit une isolation de l'échantillon associé à X_m et une diminution (induite par cet échantillon isolé) de la distance entre les données X_i de classes différentes. La méthode proposée ici maintient une distance suffisante entre les deux classes pour permettre leur regroupement par la méthode des k-moyennes.

5 Remarques de Conclusion

Nous avons présenté ici une méthode améliorée pour l'estimation de diverses distances entre matrices de covariances communément utilisées en traitement du signal et des données. Particulièrement adaptée au traitement de grandes données ($p \sim n_1, n_2 \gg 1$), cette méthode démontre de grandes performances

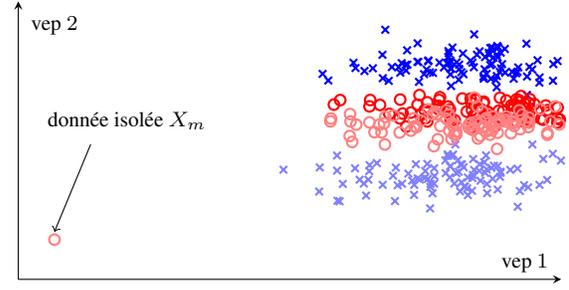


FIGURE 2 – Deux premiers vecteurs propres (vep 1 et vep 2) de \check{K} pour les estimateurs classique (cercles rouges) et proposé (croix bleues); $n_1 = \dots = n_{m-1} = 512$ et $n_m = 256$.

même sur des jeux de données de petite taille. Cependant, tout comme les estimateurs classiques, le cas $p > n_1, n_2$ (où \hat{C}_1, \hat{C}_2 ne sont pas inversibles) nous demeure inaccessible. Des approches par régularisation de \hat{C}_1, \hat{C}_2 ou par réduction pertinente de dimension pourraient apporter une solution.

Par ailleurs, de nombreuses métriques étudiées ici ont pour origine des divergences entre distributions gaussiennes *centrées*. L'extension de notre étude à des distances entre gaussiennes *non centrées* nécessite l'estimation de formes quadratiques impliquant également les vecteurs propres des C_a . Des outils plus avancés de la théorie des matrices aléatoires sont nécessaires pour traiter ces cas, dont l'importance en traitement des données est plus fondamentale que les distances étudiées ici.

Remerciements

Ces travaux sont soutenus par le projet ANR RMT4GRAPH (ANR-14-CE28-0006) et par la Chaire IDEX GSTATS de l'Université Grenoble Alpes.

Références

- [1] X. Mestre, "On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices," vol. 56, no. 11, pp. 5353–5368, Nov. 2008.
- [2] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, 1995.
- [3] Romain Couillet, Malik Tiomoko, Steeve Zozor, and Eric Moisan, "Random matrix-improved estimation of covariance matrix distances," *arXiv preprint arXiv:1810.04534*, 2018.
- [4] Malik Tiomoko, Florent Bouchard, Guillaume Ginholac, and Romain Couillet, "Random matrix improved covariance estimation for a large class of metrics," *arXiv preprint arXiv:1902.02554*, 2019.
- [5] Don Zagier, "The dilogarithm function," in *Frontiers in number theory, physics, and geometry II*, pp. 3–65. Springer, 2007.