

Distribution matricielle t -Wishart : géométrie d'information, estimation et application pour la classification de signaux EEG

Imen AYADI Florent BOUCHARD Frederic PASCAL

Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes,
91190, Gif-sur-Yvette, France

Résumé – Ce papier s'intéresse à la distribution t -Wishart définie sur l'ensemble des matrices symétriques définies positives. Il contient deux contributions. Tout d'abord, l'étude de la géométrie de l'information associée à la distribution t -Wishart avec à la clé un estimateur du maximum de vraisemblance basé sur l'optimisation riemannienne. Par ailleurs, la distribution t -Wishart est utilisée pour développer une nouvelle méthode de classification des matrices de covariance. Le classifieur proposé est obtenu en exploitant le cadre de l'analyse discriminante et en fournissant une nouvelle règle de décision. Les performances de l'estimateur proposé sont évaluées sur des données simulées. De plus, l'intérêt pratique de notre nouveau classifieur est démontré grâce à des expériences numériques sur données électroencéphalographiques (EEG) réelles. Plus précisément, la technique proposée est plus performante que le classifieur de référence en EEG, *i.e.*, celui de distance minimale à la moyenne (*minimum distance to mean*, MDM) sur un jeu de données EEG standard d'imagerie moteur (*motor imagery*, MI).

Abstract – This paper deals with the t -Wishart distribution defined on the set of positive definite symmetric matrices. It contains two contributions. First, the information geometry associated with the t -Wishart distribution is studied and a maximum likelihood estimator based on Riemannian optimization is derived. In addition, this article proposes a new method to classify covariance matrices using the t -Wishart distribution. The proposed classifier is obtained by leveraging the discriminant analysis framework and by providing an original decision rule. The performance of proposed estimator are evaluated on simulated data. Moreover, the practical interest of our new classifier is assessed on real electroencephalographical (EEG) data. More precisely, the proposed classifier performs better than the reference minimum distance to mean (MDM) classifier on a standard motor imagery (MI) EEG dataset.

1 Introduction

Les matrices de covariance jouent un rôle très important dans diverses applications du traitement du signal et de l'apprentissage automatique, telles que le traitement radar et d'images [20, 15], l'analyse de signaux biomédicaux [4, 8], *etc.* Dans ces applications, les statistiques sur l'ensemble des matrices de covariance [11], qui est la variété \mathcal{S}_p^{++} des matrices symétriques positives définies (SPD) de taille $p \times p$, sont primordiales. En effet, les statistiques sur \mathcal{S}_p^{++} sont exploitées pour la classification [8] ou l'inférence bayésienne [5]. Dans ce contexte, la distribution la plus classique sur \mathcal{S}_p^{++} est la distribution de Wishart [11], notée $\mathcal{W}(n, \Sigma)$, où n est le degré de liberté (on se restreint dans ce papier au cas $n \geq p$) et $\Sigma \in \mathcal{S}_p^{++}$ est le centre de la loi. La distribution de Wishart $\mathcal{W}(n, \Sigma)$ correspond à la distribution des matrices de covariance empiriques de n vecteurs aléatoires de \mathbb{R}^p tirés indépendamment selon la distribution gaussienne multivariée $\mathcal{N}_p(0, \Sigma)$. La fonction de densité de probabilité associée à cette distribution est, à un facteur de normalisation près,

$$f^{\mathcal{W}}(\mathcal{S}|\Sigma) \propto |\Sigma|^{-\frac{n}{2}} |\mathcal{S}|^{\frac{n-p-1}{2}} \exp\left(-\frac{1}{2} \text{tr}(\Sigma^{-1}\mathcal{S})\right), \quad (1)$$

où $|\cdot|$ et $\text{tr}(\cdot)$ désignent respectivement les opérateurs de déterminant et de trace.

L'hypothèse gaussienne des données ne tient pas dans de nombreux cas pratiques. En effet, dans certaines applications, telles que le traitement radar, les données sont intrinsèquement non gaussiennes (voir *e.g.*, [16]). Dans d'autres, en raison du

bruit et des valeurs aberrantes, il est plus probable que les données suivent des distributions à queue lourde. Dans de telles situations, il est courant de remplacer la distribution gaussienne multivariée par une distribution elliptique multivariée [14] pour modéliser les données. De manière analogue, il est possible d'étendre la distribution de Wishart sur \mathcal{S}_p^{++} par les distributions dites Wishart elliptiques [9]. Dans cet article, nous nous intéressons en particulier à la distribution t -Wishart, qui permet entre autre de mieux prendre en compte des matrices de covariance aberrantes.

La question de l'estimation du centre de ces distributions n'a pas été traitée malgré son importance pour les applications. Nous rappelons que, dans le cas de la distribution de Wishart, on sait que l'estimateur du maximum de vraisemblance pour le centre de N échantillons $\{\mathcal{S}_i\}_{i=1}^N$ indépendants et identiquement distribués (iid) selon $\mathcal{W}(n, \Sigma)$ est

$$\hat{\Sigma}^{\mathcal{W}} = \frac{1}{nN} \sum_{i=1}^N \mathcal{S}_i. \quad (2)$$

La première contribution de cet article est de proposer un estimateur du maximum de vraisemblance pour le centre de la distribution t -Wishart. Pour dériver cet estimateur, la géométrie de l'information de la distribution t -Wishart sur \mathcal{S}_p^{++} et l'optimisation riemannienne [1] sont exploitées. L'estimateur proposé est validé sur des données simulées.

La deuxième contribution consiste à construire une nouvelle méthode de classification utilisant la distribution t -Wishart dans le cadre de l'analyse discriminante. Le classifieur résultant est testé sur une base de données électroencéphalogra-

phiques (EEG) et comparé à la méthode de l'état de l'art de la distance minimale à la moyenne (MDM).

2 Distribution t -Wishart et sa géométrie de l'information

La distribution de t -Wishart généralise la distribution de Wishart de la même manière que la t -distribution multivariée généralise la distribution gaussienne multivariée. La fonction de densité de probabilité associée à la variable aléatoire $\mathbf{S} \in \mathcal{S}_p^{++}$ suivant la distribution de t -Wishart, notée $t\text{-}\mathcal{W}(n, \Sigma, \nu)$ est, à un facteur de normalisation près,

$$f^{t\text{-}\mathcal{W}}(\mathbf{S}|\Sigma) \propto |\Sigma|^{-\frac{n}{2}} |\mathbf{S}|^{\frac{n-p-1}{2}} \left(1 + \frac{\text{tr}(\Sigma^{-1}\mathbf{S})}{\nu}\right)^{-\frac{\nu+np}{2}}, \quad (3)$$

où $\nu > 0$ représente le degré de liberté (ddl) de la distribution.

Afin d'obtenir la métrique d'information de Fisher et l'estimateur du maximum de vraisemblance (EMV) de la t -Wishart, la log-vraisemblance de (3) est nécessaire. Étant donné des échantillons iid $\{\mathbf{S}_i\}_{i=1}^N$, l'opposée de la log-vraisemblance correspondant à (3) est, à une constante additive près,

$$\mathcal{L}^{t\text{-}\mathcal{W}}(\Sigma) = \frac{nN}{2} \log |\Sigma| + \frac{\nu + np}{2} \sum_{i=1}^N \log \left(1 + \frac{\text{tr}(\Sigma^{-1}\mathbf{S}_i)}{\nu}\right). \quad (4)$$

L'espace des paramètres de la t -Wishart est \mathcal{S}_p^{++} . L'espace tangent en $\Sigma \in \mathcal{S}_p^{++}$ peut être identifié avec \mathcal{S}_p , l'espace vectoriel des matrices symétriques. La géométrie optimale de \mathcal{S}_p^{++} par rapport à la t -Wishart est celle induite par la métrique d'information de Fisher correspondante. Pour $\Sigma \in \mathcal{S}_p^{++}$, ξ et $\eta \in \mathcal{S}_p$, elle est donnée par [17]

$$\langle \xi, \eta \rangle_{\Sigma} = \mathbb{E}[\text{D}^2 \mathcal{L}(\Sigma)^{t\text{-}\mathcal{W}}[\xi, \eta]]. \quad (5)$$

où $\text{D}^2 \mathcal{L}^{t\text{-}\mathcal{W}}(\Sigma)[\xi, \eta]$ est la dérivée directionnelle seconde de $\mathcal{L}^{t\text{-}\mathcal{W}}$ à Σ par rapport aux directions ξ et η . Lors de la dérivation de la métrique de Fisher, il est habituel de choisir $N = 1$.

La métrique d'information de Fisher de la t -Wishart est fournie dans la Proposition 1 (voir [2] pour la preuve).

Proposition 1 (Métrique de Fisher). *La métrique d'information de Fisher de la t -Wishart est, pour $\Sigma \in \mathcal{S}_p^{++}$, ξ et $\eta \in \mathcal{S}_p$,*

$$\langle \xi, \eta \rangle_{\Sigma} = \alpha \text{tr}(\Sigma^{-1}\xi\Sigma^{-1}\eta) + \beta \text{tr}(\Sigma^{-1}\xi) \text{tr}(\Sigma^{-1}\eta),$$

$$\text{où} \quad \alpha = \frac{n}{2} \frac{\nu + np}{\nu + np + 2}, \quad \beta = \frac{n}{2} \left(\alpha - \frac{n}{2}\right).$$

La géométrie de \mathcal{S}_p^{++} munie d'une métrique riemannienne de la forme donnée dans la Proposition 1 est bien connue et se retrouve, par exemple, dans [6]. En particulier, la géodésique (généralisation d'une droite sur une variété) $\gamma : [0, 1] \rightarrow \mathcal{S}_p^{++}$ issue de $\Sigma \in \mathcal{S}_p^{++}$ dans la direction $\xi \in \mathcal{S}_p$ est

$$\gamma(t) = \Sigma \text{Exp}(t\Sigma^{-1}\xi), \quad (6)$$

où $\text{Exp}(\cdot)$ désigne l'exponentielle matricielle. De là, on peut déduire la distance d'information de Fisher de la distribution t -Wishart. Elle est définie comme la longueur de la géodésique

joignant Σ et $\bar{\Sigma}$ selon la métrique de la Proposition 1 et son carré est égal à

$$\delta^2(\Sigma, \bar{\Sigma}) = \alpha \|\text{Log}(\Sigma^{-\frac{1}{2}}\bar{\Sigma}\Sigma^{-\frac{1}{2}})\|_2^2 + \beta (\log |\Sigma^{-1}\bar{\Sigma}|)^2. \quad (7)$$

où $\|\cdot\|_2$ est la norme de Frobenius et $\text{Log}(\cdot)$ désigne le logarithme matriciel.

3 Apprentissage des paramètres de la distribution t -Wishart

Pour obtenir l'EMV du centre, il faut résoudre le problème d'optimisation

$$\hat{\Sigma} = \underset{\Sigma \in \mathcal{S}_p^{++}}{\text{argmin}} \mathcal{L}^{t\text{-}\mathcal{W}}(\Sigma). \quad (8)$$

Comme pour la t -distribution multivariée, aucune solution n'est connue sous forme explicite et il faut employer un algorithme itératif. Dans ce travail, nous exploitons la géométrie de \mathcal{S}_p^{++} fournie dans la Section 2 pour résoudre (8) via l'optimisation riemannienne [1].

Pour ce faire, la première étape est de définir le gradient riemannien, noté $\nabla \mathcal{L}^{t\text{-}\mathcal{W}}(\Sigma)$, de $\mathcal{L}^{t\text{-}\mathcal{W}}$ à $\Sigma \in \mathcal{S}_p^{++}$ associé à la métrique de Fisher de la t -Wishart. Il est défini comme le seul vecteur tangent tel que, pour tout $\xi \in \mathcal{S}_p$,

$$\langle \nabla \mathcal{L}^{t\text{-}\mathcal{W}}(\Sigma), \xi \rangle_{\Sigma} = \text{D} \mathcal{L}^{t\text{-}\mathcal{W}}(\Sigma)[\xi]. \quad (9)$$

On peut montrer que

$$\nabla \mathcal{L}^{t\text{-}\mathcal{W}}(\Sigma) = -\frac{\nu + np + 2}{n} \sum_{i=1}^N \frac{\mathbf{S}_i - n\Sigma}{\nu + \text{tr}(\Sigma^{-1}\mathbf{S}_i)} \quad (10)$$

Le gradient riemannien est suffisant pour définir une direction de descente de $\mathcal{L}^{t\text{-}\mathcal{W}}$ à Σ et définir l'algorithme de descente de gradient riemannien [1]. Cependant, si l'on veut employer des méthodes d'optimisation plus sophistiquées telles que le gradient conjugué ou l'algorithme de Broyden-Fletcher-Goldfarb-Shanno (BFGS) [7], on doit définir un opérateur de transport de vecteurs. Un tel opérateur permet de transporter un vecteur tangent d'un point sur l'espace tangent d'un autre point [1]. Le choix le plus naturel est celui correspondant au transport parallèle sur \mathcal{S}_p^{++} [13]. Le transport du vecteur tangent ξ de Σ sur l'espace tangent en $\bar{\Sigma}$ est $\mathcal{T}_{\Sigma \rightarrow \bar{\Sigma}}(\xi) = (\bar{\Sigma}\Sigma^{-1})^{1/2} \xi (\Sigma^{-1}\bar{\Sigma})^{1/2}$.

Une fois qu'une direction de descente est sélectionnée dans l'espace tangent, il reste à revenir sur la variété. Ceci est réalisé par une rétraction [1]. Dans notre cas, d'un point de vue numérique, la meilleure solution est de prendre l'approximation au second ordre des géodésiques (6). Étant donné $\Sigma \in \mathcal{S}_p^{++}$ et $\xi \in \mathcal{S}_p$, c'est [13]

$$R_{\Sigma}(\xi) = \Sigma + \xi + \frac{1}{2}\xi\Sigma^{-1}\xi. \quad (11)$$

Avec ces outils, plusieurs algorithmes d'optimisation riemannienne peuvent être utilisés pour résoudre (8). Par exemple, la séquence d'itérations $\{\Sigma_l\}$ et de directions de descente $\{\xi_l\}$ générées par un algorithme de gradient conjugué riemannien est :

$$\begin{cases} \Sigma_{l+1} = R_{\Sigma_l}(\xi_l) \\ \xi_l = t_l(-\nabla \mathcal{L}^{t\text{-}\mathcal{W}}(\Sigma_l) + s_l \mathcal{T}_{\Sigma_{l-1} \rightarrow \Sigma_l}(\xi_{l-1})), \end{cases} \quad (12)$$

où t_l est une taille de pas calculée via une recherche linéaire [1] et s_l peut être calculé avec la règle dans [12] par exemple.

4 Analyse discriminante basée sur la distribution t -Wishart

Dans cette section, un nouveau classifieur bayésien pour la classification des signaux est développé. Inspirée de l'Analyse Discriminante (LDA/QDA) dans le cas multivarié [10], la méthode introduite généralise le concept d'analyse discriminante aux matrices, similairement à [19]. Elle exploite la distribution t -Wishart pour modéliser la probabilité que les échantillons observés appartiennent à une classe donnée.

Formellement, chaque signal est représenté par une matrice $\mathbf{X} \in \mathbb{R}^{p \times n}$ où p correspond au nombre de capteurs et n au nombre d'échantillons temporels. On définit $\mathbf{S} = \mathbf{X}\mathbf{X}^T$, qui appartient par construction à \mathcal{S}_p^{++} . On considère un problème de classification avec K classes. Le classifieur proposé approche ce problème en supposant qu'une matrice \mathbf{S} appartenant à la classe $k \in \llbracket 1, K \rrbracket$ suit la distribution t - $\mathcal{W}(n, \Sigma_k, \nu)$ où Σ_k est le centre de la classe k . La proposition suivante introduit l'analyse discriminante de la t -Wishart, notée t -WDA. La preuve est omise faute d'espace et fournie dans [3].

Proposition 2 (t -WDA). *La règle de décision de l'analyse discriminante t -Wishart, pour une matrice de test \mathbf{S} , est*

$$\hat{y}(\mathbf{S}) = \operatorname{argmax}_{k \in \llbracket 1, K \rrbracket} \delta_k(\mathbf{S}), \quad (13)$$

avec la fonction discriminante

$$\delta_k(\mathbf{S}) = \log(\hat{\pi}_k) - \frac{n}{2} \log |\hat{\Sigma}_k| - \frac{\nu + np}{2} \log \left(1 + \frac{\operatorname{tr}(\hat{\Sigma}_k^{-1} \mathbf{S})}{\nu} \right), \quad (14)$$

où $\hat{\pi}_k$ est la proportion de la classe k dans l'ensemble d'apprentissage; $\hat{\Sigma}_k$ est le centre estimé des matrices de covariance de la classe k dans le jeu de données d'apprentissage, calculé grâce à (12), et ν est le degré de liberté du modèle¹.

Remarque 1. ν est un hyperparamètre et l'estimer pendant l'étape d'apprentissage dépasse le cadre de cet article. De plus, dans le modèle actuel, un unique ν est choisi pour toutes les classes pour simplifier. Il serait plus réaliste d'estimer différents ν_k pour chaque classe afin d'améliorer la discrimination, surtout lorsque les centres de classes sont proches.

5 Expériences numériques

5.1 Données simulées

Pour valider l'intérêt de l'estimateur du maximum de vraisemblance dérivé dans la Section 3, des expériences numériques sont menées sur des données simulées. On pose $n = 100$, $p = 16$ et on génère aléatoirement un centre $\Sigma \in \mathcal{S}_p^{++}$. Pour $N \in \{30, 70, 100, 300, 500\}$, nous tirons des échantillons iid $\{\mathbf{S}_i\}_{i=1}^N$ selon t - $\mathcal{W}(n, \Sigma, \nu)$ avec $\nu \in \{5, 100\}$. Pour chaque valeur du paramètre, 200 ensembles de données sont générés. Pour générer $\mathbf{S} \sim t$ - $\mathcal{W}(n, \Sigma, \nu)$, on génère $\mathbf{x} \in \mathbb{R}^{np}$ selon la loi $t_\nu(0, \mathbf{I}_n \otimes \Sigma)$ puis on construit $\mathbf{X} \in \mathbb{R}^{p \times n}$ telle que la vectorisation de \mathbf{X} est \mathbf{x} . Finalement, $\mathbf{S} = \mathbf{X}\mathbf{X}^T$.

Nous comparons deux algorithmes d'estimation différents : l'estimateur de Wishart $\hat{\Sigma}^{\mathcal{W}}$ défini dans (2); et l'estimateur du maximum de vraisemblance $\hat{\Sigma}^{t\text{-}\mathcal{W}}$. Ce dernier est calculé

¹ considéré ici comme un hyperparamètre

avec un algorithme de gradient conjugué riemannien tel que présenté dans (12). Pour évaluer l'erreur d'estimation, nous utilisons la distance induite par l'information de Fisher, donnée par (7), entre le vrai centre et les estimateurs.

La Figure 1 affiche la médiane, les quantiles à 5% et 95% des erreurs de chaque estimateur considéré. Lorsque les données suivent (ou sont proches de) la distribution de Wishart ($\nu=100$), on observe que l'EMV de la vraie distribution obtenu par optimisation riemannienne et l'estimateur de Wishart présentent des performances équivalentes. Dans un tel cas, $\hat{\Sigma}^{\mathcal{W}}$ est à préférer car il est beaucoup moins cher à calculer. Cependant, comme attendu, lorsque nous nous éloignons de la distribution de Wishart ($\nu=5$), $\hat{\Sigma}^{\mathcal{W}}$ ne donne plus de bons résultats et est largement surpassé par l'EMV.

5.2 Application sur les signaux EEG

Dans cette partie, nous montrons l'avantage pratique du classifieur t -WDA proposé par rapport au classifieur MDM de référence sur un jeu de données EEG standard d'imagerie motrice (MI). La MI est un paradigme des interfaces cerveau-ordinateur (*brain computer interfaces*, BCI). Un sujet simule mentalement une action physique, *i.e.*, il/elle imagine bouger sa main droite ou gauche, ses pieds, sa langue, *etc.* Dans ce travail, nous considérons le jeu de données BNCI2014001 [18]. Il contient les enregistrements EEG de 9 sujets. Les signaux sont acquis avec 22 électrodes et une fréquence d'échantillonnage de 250 Hz. Pour chaque sujet, deux sessions composées de 288 essais sont disponibles. Seules les quatre dernières secondes de l'essai sont prises en compte. Ainsi, pour chaque essai, nous avons $\mathbf{X} \in \mathbb{R}^{p \times n}$ avec $p = 22$ et $n = 4 \times 250 = 1000$.

Une classification intra-session est effectuée : une validation croisée à 5 blocs stratifiés est faite à chaque session, en la divisant en 228 échantillons d'apprentissage et 60 échantillons de test avec une représentation équilibrée des $K = 4$ classes. Pour t -WDA, le degré de liberté ν est fixé à 10. Les moyennes et les écarts types des précisions des tests pour chaque sujet sont indiqués dans la Figure 2.

On peut observer que le MDM est plus précis pour deux sujets (1 et 5) et a une précision similaire pour le sujet 2 par rapport à t -WDA. Le t -WDA donne les meilleurs résultats pour les six sujets restants. En faisant la moyenne sur tous les sujets, t -WDA a un gain de 1,78 % par rapport au MDM. Ainsi, t -WDA semble être le classificateur le plus puissant sur cet ensemble de données MI par rapport à MDM.

6 Conclusions

Cet article s'est intéressé à la distribution t -Wishart sur \mathcal{S}_p^{++} . La géométrie de l'information est obtenue et un estimateur du maximum de vraisemblance du centre est proposé en utilisant des outils de l'optimisation riemannienne. L'algorithme est validé sur des données simulées. Par ailleurs, une nouvelle méthode de classification basée sur t -Wishart est proposée, en utilisant l'analyse discriminante. Le nouveau classifieur est évalué sur des données électroencéphalographiques et montre une meilleure performance par rapport à la méthode de l'état de l'art. Ces travaux peuvent être étendus au cas de (1) degrés de libertés différents par classe et plus généralement à (2) l'analyse de nouvelles distributions matricielles elliptiques.

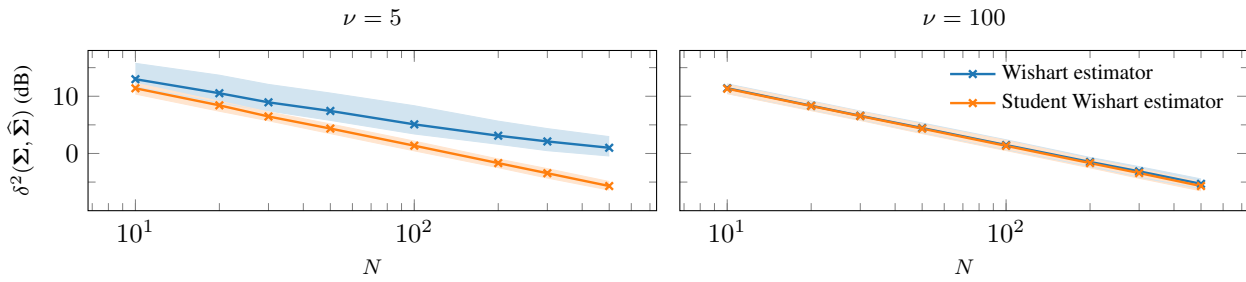


FIGURE 1 : Médiane (lines continues) et les quantiles à 5% et 95% (aires remplies) des erreurs de mesure de l'EMV de Wishart et de t -Wishart en fonction de la taille des échantillons N . Les médianes et les quantiles sont calculés sur 200 ensembles de données simulées $\{\mathcal{S}_i\}$ tirées selon une t -Wishart.

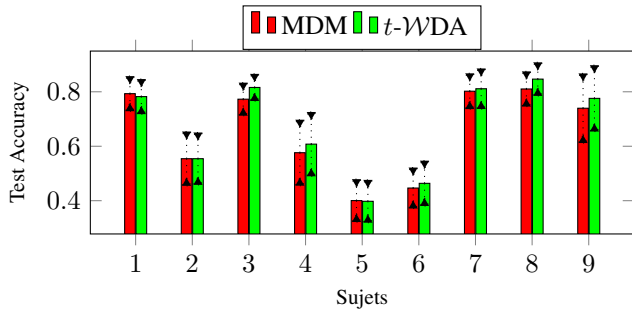


FIGURE 2 : Comparaison des classifieurs sur BNCI2014001. Les écart-types sont dessinés en trait discontinu. Le ddl pour t -WDA est $\nu = 10$.

Références

- [1] P.-A. ABSIL, R. MAHONY et R. SEPULCHRE : *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, USA, 2008.
- [2] I. AYADI, F. BOUCHARD et F. PASCAL : Elliptical Wishart distribution : maximum likelihood estimator from information geometry. *In International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023.
- [3] I. AYADI, F. BOUCHARD et F. PASCAL : t -Wda : A novel discriminant analysis applied to eeg classification. *In European Signal Processing Conference (EUSIPCO)*. IEEE, 2023.
- [4] A. BARACHANT, S. BONNET, M. CONGEDO et C. JUTTEN : Multiclass brain-computer interface classification by Riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, 2011.
- [5] O. BESSON, S. BIDON et J.-Y. TOURNERET : Covariance matrix estimation with heterogeneous samples. *IEEE Transactions on Signal Processing*, 56(3):909–920, 2008.
- [6] A. BRELOY, G. GINOLHAC, A. RENAUX et F. BOUCHARD : Intrinsic Cramér–Rao bounds for scatter and shape matrices estimation in CES distributions. *IEEE Signal Processing Letters*, 26(2):262–266, 2018.
- [7] A.G. BUCKLEY : A combined conjugate-gradient quasi-newton minimization algorithm. *Mathematical Programming*, 15(1):200–210, 1978.
- [8] S. CHEVALLIER, E. K. KALUNGA, Q. BARTHÉLEMY et E. MONACELLI : Review of Riemannian distances and divergences, applied to SSVEP-based BCI. *Neuroinformatics*, 19(1):93–106, 2021.
- [9] K.-T. FANG et T. W. ANDERSON : *Statistical inference in elliptically contoured and related distributions*. Allerton Press, 1990.
- [10] B. GHOJOGH et M. CROWLEY : Linear and quadratic discriminant analysis : Tutorial, 2019.
- [11] A. K. GUPTA et D. K. NAGAR : *Matrix variate distributions*. Chapman and Hall/CRC, 2018.
- [12] M. R. HESTENES et E. STIEFEL : Methods of conjugate gradients for solving linear equation. *Journal of research of the National Bureau of Standards*, 49(6):409, 1952.
- [13] B. JEURIS, R. VANDEBRIL et B. VANDEREYCKEN : A survey and comparison of contemporary algorithms for computing the matrix geometric mean. *Electronic Transactions on Numerical Analysis*, 39:379–402, 2012.
- [14] E. OLLILA, D. E. TYLER, V. KOIVUNEN et H. V. POOR : Complex elliptically symmetric distributions : Survey, new results and applications. *IEEE Transactions on signal processing*, 60(11):5597–5625, 2012.
- [15] F. PASCAL, Y. CHITOUR, J.-P. OVARLEZ, P. FORSTER et P. LARZABAL : Covariance structure maximum-likelihood estimates in compound Gaussian noise : Existence and algorithm analysis. *IEEE Transactions on Signal Processing*, 56(1):34–48, 2007.
- [16] K. J. SANGSTON et K. R/ GERLACH : Coherent detection of radar targets in a non-Gaussian background. *IEEE Transactions on Aerospace and Electronic Systems*, 30(2):330–340, 1994.
- [17] S. T. SMITH : Covariance, subspace, and intrinsic Cramér–Rao bounds. *IEEE Transactions on Signal Processing*, 53(5):1610–1630, 2005.
- [18] M. TANGERMANN, K.-R. MÜLLER, A. AERTSEN, N. BIRBAUMER, C. BRAUN, C. BRUNNER, R. LEEB, C. MEHRING, K.J. MILLER, G. MUELLER-PUTZ *et al.* : Review of the BCI competition IV. *Frontiers in neuroscience*, page 55, 2012.
- [19] G.Z. THOMPSON, R. MAITRA, W.Q. MEEKER et A.F. BASTAWROS : Classification with the matrix-variate t distribution. *Journal of Computational and Graphical Statistics*, 29(3):668–674, 2020.
- [20] O. TUZEL, F. PORIKLI et P. MEER : Pedestrian detection via classification on Riemannian manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1713–1727, 2008.