

Estimation de barycentres sur variétés de Stiefel : une approche par projection

Nils LAURENT¹ Florent BOUCHARD² Salem SAID³ Nicolas LE BIHAN¹

¹CNRS, Univ. Grenoble Alpes, Grenoble INP, GIPSA-lab, 38000 Grenoble France

²L2S, CentraleSupélec, Université Paris-Saclay, France

³CNRS, Univ. Grenoble Alpes, Grenoble INP, LJK, 38000 Grenoble France

Résumé – Dans cet article, nous nous intéressons à l’estimation du barycentre d’un ensemble de points sur la variété de Stiefel. Les méthodes de la littérature se basent sur des approximations d’opérations permettant de passer localement de la variété au plan tangent (*rétraction* et *rétraction inverse*) et proposent des algorithmes itératifs pour obtenir le centroïde d’un échantillon. Nous proposons l’utilisation d’une projection en place de l’opération de rétraction inverse et démontrons que la solution de l’équation de point fixe associée est donnée par une expression analytique, qui correspond simplement à la projection de la moyenne extrinsèque sur Stiefel. Une validation numérique montre la robustesse de notre approche ainsi que son faible coût calculatoire.

Abstract – In this paper, we are interested in the estimation of the barycentre of a set of points on the Stiefel variety. State-of-the-art methods are based on approximations of operations allowing to go locally from the variety to the tangent plane (*retraction* and *inverse retraction*) and propose iterative algorithms to obtain the centroid of a sample. We propose the use of a projection in place of the inverse retraction operation and show that the solution of the associated fixed point equation is given by an analytical expression, which simply corresponds to the projected extrinsic mean on Stiefel. A numerical validation shows the robustness of our approach as well as its low computational cost.

1 Introduction

Dans de nombreuses applications en traitement du signal, la représentation de séries temporelles dans un espace de dimension faible permet un traitement efficace et potentiellement rapide des données. Le choix de l’espace de représentation dépend souvent de l’application cible, mais d’une manière générale, il convient d’opter pour un espace de représentation possédant des propriétés adaptées à la tâche de traitement visée. Par exemple, en estimation/détection pour les signaux radars, la représentation des données par des matrices PSD est souvent choisie [2]. Pour les séries temporelles chaotiques, la projection sur variétés de dimension faible est également une technique largement utilisée [4] afin d’estimer l’ordre et la dimension du système dynamique permettant de générer les séries temporelles associées. La mise en correspondance entre une série temporelle et un point sur une variété peut donc se faire de plusieurs façons. Parmi les variétés matricielles classiquement considérées, certaines ont des spécificités qui nécessitent de redéfinir des outils standards d’analyse statistique comme le barycentre.

Dans cet article, nous nous intéressons à proposer un algorithme efficace pour le calcul du barycentre d’un nuage de points sur la variété de Stiefel. Pour cette variété, la difficulté réside dans l’approximation de l’application log (permettant de passer de la variété au plan tangent en un point de la variété) qui a été abordée suivant plusieurs approches [8]. Ici, nous proposons de considérer un calcul de barycentre basé sur une projection, et montrons que la solution obtenue est en fait celle de l’équation du point fixe considérée dans [5]. La solution que nous proposons est analytique et ne requiert donc pas d’algorithme itératif comme c’est le cas pour les algorithmes

proposés dans [5]. De plus, elle est plus robuste aux points-faux que les algorithmes connus et possède un coût de calcul bien moindre. Des simulations numériques illustrent les propriétés de l’approche proposée et la compare aux algorithmes itératifs mentionnés.

2 Variété de Stiefel

Nous rappelons quelques éléments de géométrie utiles pour la définition de barycentre et son calcul dans le cas de la variété matricielle de Stiefel. Ces éléments sont illustrés Figure 1. Le formalisme utilisé ici est emprunté à [1]. La variété de Stiefel $\text{St}_{n,p}$ est définie comme l’ensemble des matrices de taille $n \times p$ telles que

$$\text{St}_{n,p} = \{\mathbf{X} \in \mathbb{R}^{n \times p} : \mathbf{X}^T \mathbf{X} = \mathbf{I}_p\}. \quad (1)$$

Étant donné un élément $\mathbf{X} \in \text{St}_{p,k}$, l’espace tangent à $\text{St}_{n,p}$ en $\mathbf{X} \in \text{St}_{p,k}$, noté $T_{\mathbf{X}}\text{St}_{n,p}$ est donné par

$$T_{\mathbf{X}}\text{St}_{n,p} = \{\boldsymbol{\xi} \in \mathbb{R}^{n \times p} : \mathbf{X}^T \boldsymbol{\xi} + \boldsymbol{\xi}^T \mathbf{X} = \mathbf{0}\}. \quad (2)$$

Une fois muni d’une métrique $\text{St}_{p,k}$ est une variété riemannienne. Un choix direct est donné par la métrique euclidienne associée à son espace ambiant $\mathbb{R}^{n \times p}$. Elle s’exprime comme suit, pour $\mathbf{X} \in \text{St}_{n,p}$ et $\boldsymbol{\xi}, \boldsymbol{\eta} \in T_{\mathbf{X}}\text{St}_{n,p}$,

$$\langle \boldsymbol{\xi}, \boldsymbol{\eta} \rangle_{\mathbf{X}} = \text{tr}(\boldsymbol{\xi}^T \boldsymbol{\eta}). \quad (3)$$

Étant donné une matrice $\mathbf{Z} \in \mathbb{R}^{n \times p}$, sa projection orthogonale sur le plan tangent $T_{\mathbf{X}}\text{St}_{n,p}$ (plan tangent à $\mathbf{X} \in \text{St}_{n,p}$) est alors donnée par

$$P_{\mathbf{X}}(\mathbf{Z}) = \mathbf{Z} - \mathbf{X} \text{sym}(\mathbf{X}^T \mathbf{Z}), \quad (4)$$

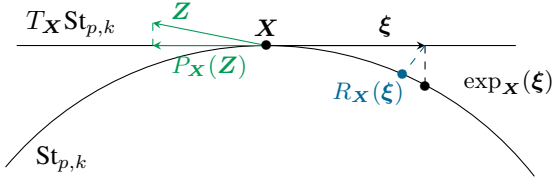


FIGURE 1 : Illustration de la variété de Stiefel $St_{p,k}$ et de l'espace tangent à un point $X \in St_{p,k}$. On présente également les concepts de : (i) projection orthogonale $P_X(Z)$ d'un point $Z \in \mathbb{R}^{p \times k}$ de l'espace ambiant sur l'espace tangent à X ; (ii) exponentielle riemannienne $\exp_X(\xi)$, qui permet de ramener un vecteur tangent ξ sur la variété de façon optimale ; et (iii) de rétraction $R_X(\xi)$, qui permet de ramener un vecteur tangent sur la variété de façon alternative.

où $\text{sym}(\cdot)$ renvoie la partie symétrique de son argument.

Dans cet article, nous nous intéressons au problème de l'estimation du barycentre d'un échantillon de taille N , $\{\mathbf{X}_k\}_{k=1,\dots,N}$ avec $\mathbf{X}_k \in St_{n,p}$. La moyenne riemannienne, $\mathbf{G} \in St_{n,p}$ de l'ensemble des N matrices \mathbf{X}_k peut s'obtenir de manière itérative, à partir d'une initialisation $\mathbf{G}^{(0)}$ par [6]

$$\mathbf{G}^{(i+1)} = \exp_{\mathbf{G}^{(i)}} \left(\frac{1}{N} \sum_{k=1}^N \log_{\mathbf{G}^{(i)}}(\mathbf{X}_k) \right), \quad (5)$$

où $\exp_X : T_X St_{n,p} \rightarrow St_{n,p}$ et $\log_X : St_{n,p} \rightarrow T_X St_{n,p}$ sont les applications exponentielles et logarithme sur Stiefel. Ces applications définissent un difféomorphisme au voisinage de X et constituent le choix le plus naturel pour obtenir un point sur la variété en fonction d'un vecteur tangent et inversement.

Bien que l'exponentielle riemannienne sur Stiefel admette une expression analytique [1, Section 5.4], ce n'est pas le cas pour le logarithme riemannien. Ce dernier ne peut être évalué que par la procédure itérative décrite dans [8], qui est coûteuse. La solution de l'équation (5) ne peut donc pas se calculer simplement.

3 R-barycentres

Une variante pour approximer le barycentre sur $St_{n,p}$ a été introduite par [5] qui permet d'éviter le calcul de l'application logarithme and définissant la notion de R -barycentre. L'approche proposée fait usage de *rétractions*, qui sont des approximations au premier ordre de l'application exponentielle [1, Chapitre 4]. Une rétraction $R_X : T_X St_{p,k} \rightarrow St_{p,k}$ évaluée en $\xi \in T_X St(n,p)$ satisfait

$$R_X(\xi) = X + \xi + o(\|\xi\|). \quad (6)$$

Bien que l'application exponentielle riemannienne soit elle-même une rétraction, il peut être intéressant d'en utiliser d'autres en pratique. En effet, l'application exponentielle peut être compliquée et coûteuse, voire numériquement instable. Plusieurs options de rétractions sont disponibles depuis leurs nombreuses utilisations en optimisation riemannienne [1]. L'intérêt, dans notre cas, est de remplacer le calcul du logarithme riemannien par le celui de l'inverse d'une rétraction. En revanche, cela requiert de définir une alternative à (5), par exemple en considérant un R -barycentre. Ce dernier est défini,

pour un ensemble de points $\{\mathbf{X}_k \in St_{n,p}\}$, et une initialisation $\mathbf{G}^{(0)} \in St_{n,p}$, par le point fixe associé à

$$\mathbf{G}^{(i+1)} = R_{\mathbf{G}^{(i)}} \left(\frac{1}{N} \sum_k R_{\mathbf{G}^{(i)}}^{-1}(\mathbf{X}_k) \right), \quad (7)$$

où $R_X^{-1} : St_{n,p} \rightarrow T_X St_{n,p}$ est la rétraction inverse de $R_X : T_X St_{n,p} \rightarrow St_{n,p}$. La rétraction inverse R^{-1} est une application qui satisfait, pour $X \in St_{n,p}$, $Y \in St_{n,p}$,

$$R_X(R_X^{-1}(Y)) = Y. \quad (8)$$

Des méthodes pour le calcul du point fixe associé à l'équation (7) ont été proposées pour plusieurs choix de rétractions [5]. Nous allons maintenant rappeler trois de ces rétractions, qui serviront de base pour la comparaison avec la méthode proposée.

Soit $\xi \in T_X St_{n,p}$ un élément du plan tangent au point X . La rétraction basée sur la décomposition polaire est donnée par $R_X(\xi) = \text{uf}(X + \xi)$, où uf est le facteur orthogonal (matrice orthogonale) de la décomposition polaire. De manière analogue, la rétraction basée sur la décomposition QR est donnée comme $R_X(\xi) = \text{qf}(X + \xi)$, où qf correspond au facteur Q (matrice unitaire dans la décomposition QR). Les inverses de ces deux rétractions peuvent être calculées par la résolution d'un système linéaire. Finalement, la troisième rétraction est celle dite *orthographique*, pour laquelle l'inverse est exprimée par $R_X^{-1}(Q) = Q - \frac{1}{2}X(Q^T X + X^T Q)$ et où l'on obtient $Q \in St_{n,p}$ et R_X par résolution d'une équation algébrique de Riccati.

Dans chacune des méthodes mentionnées, une expression simple existe soit pour R_X , soit pour R_X^{-1} , mais la résolution d'un système, potentiellement coûteux et instable, est nécessaire pour l'autre opération. De plus, l'existence d'une solution au système qui leur sont associé est seulement garantie dans un voisinage [5]. L'objectif ici est de proposer une méthode simple à implémenter, à la fois pour passer du plan tangent à la variété et de la variété au plan tangent.

Par la suite, nous proposons une variante à l'équation (7) pour la recherche d'un point fixe et la construction d'une méthode qui permet un calcul de sa solution. Nous démontrons que cette-dernière admet une expression analytique dont la mise en oeuvre est nettement plus efficace que les R -barycentres introduits précédemment.

4 Barycentre basé sur la projection

Dans cette section, nous proposons une approximation du barycentre en utilisant une rétraction R_X sans utiliser son inverse. À la place de R_X^{-1} , nous utilisons une application de lifting notée L_X^1 , définie par

$$L_X(Q) := P_X(Q - X). \quad (9)$$

où P_X est la projection orthogonale dans l'espace tangent de X , comme défini dans (4). Nous pouvons déjà remarquer que L_X correspond à la rétraction inverse orthographique. Cette

¹ L_X est une opération qui permet de projeter un point Q de la variété sur l'espace tangent d'un autre point X de la variété. Le logarithme riemannien tout comme les rétractions inverses sont des exemples de liftings. Ils sont cependant généralement difficiles à calculer et nous proposons ici un lifting plus simple basé sur la projection orthogonale sur le plan tangent.

relation entre le R-barycentre orthographique et le barycentre associé à notre méthode est commenté en fin de section.

L'application $L_{\mathbf{X}}$ nous permet de considérer l'alternative à (7) donnée par l'équation de point fixe

$$\mathbf{G} = R_{\mathbf{G}} \left(\frac{1}{N} \sum_{k=1}^N L_{\mathbf{G}}(\mathbf{X}_k) \right). \quad (10)$$

Dans la proposition suivante, nous montrons que la solution du problème de point fixe peut être calculé directement.

Proposition 1. *La moyenne extrinsèque \mathbf{X}^* , définie par*

$$\mathbf{X}^* := \text{uf} \left(\frac{1}{N} \sum_{k=1}^N \mathbf{X}_k \right) \quad (11)$$

est solution de l'équation de point fixe (10).

Démonstration. D'après l'équation (6), il est suffisant de montrer que

$$\frac{1}{N} \sum_{k=1}^N L_{\mathbf{X}^*}(\mathbf{X}_k) = 0. \quad (12)$$

Soit $F(\mathbf{X}) = \frac{1}{2N} \sum_{k=1}^N \|\mathbf{X} - \mathbf{X}_k\|^2$, où $\|\cdot\|$ est la norme de Frobenius. La moyenne extrinsèque \mathbf{X}^* est le minimiseur de $F(\mathbf{X})$, pour $\mathbf{X} \in \text{St}_{n,p}$. Ainsi, \mathbf{X}^* satisfait la condition d'optimalité du premier ordre [7]

$$P_{\mathbf{X}^*}(\nabla F(\mathbf{X}^*)) = 0. \quad (13)$$

Or,

$$\nabla F(\mathbf{X}^*) = \frac{1}{N} \sum_{k=1}^N (\mathbf{X}^* - \mathbf{X}_k).$$

Par linéarité de $P_{\mathbf{X}^*}$, en conséquence de (13), on a

$$\frac{1}{N} \sum_{k=1}^N P_{\mathbf{X}^*}(\mathbf{X}^* - \mathbf{X}_k) = 0. \quad (14)$$

Par définition $L_{\mathbf{X}^*} = P_{\mathbf{X}^*}(\mathbf{X}^* - \mathbf{X}_k)$, et finalement \mathbf{X}^* est le point fixe de (12). \square

La Proposition 1 nous montre que l'expression de la solution de l'équation (10) est analytique, et a la forme donnée par (11). Ceci n'est pas le cas pour le calcul du point fixe de l'équation (7), qui doit être fait de manière itérative.

Remarque : Le R-barycentre associé à la rétraction orthographique est une descente de gradient qui converge vers la solution donnée par l'équation (11). En effet, le membre gauche de l'équation (12) peut être identifié comme le gradient de (7), car $R_{\mathbf{X}}^{-1} = P_{\mathbf{X}}(\mathbf{Q} - \mathbf{X}) = L_{\mathbf{X}}$.

5 Validation numérique

Dans cette section, nous comparons la solution donnée par l'équation (11) avec les méthodes itératives proposées dans [5]. Cette comparaison est réalisée sur un échantillon à valeurs sur la variété de Stiefel $\text{St}_{n,p}$ pour différentes valeurs des paramètres n et p . Le cas particulier $n = 3, p = 1$ correspondant à la sphère dans \mathbb{R}^3 est utilisé pour certaines visualisations (Figure 2).

Pour générer un échantillon de N points \mathbf{X}_k ($k = 1, \dots, N$) sur $\text{St}_{n,p}$ de barycentre $\mathbf{C} \in \text{St}_{n,p}$, nous procédons comme suit. Le barycentre $\mathbf{C} \in \text{St}_{n,p}$ est pris aléatoirement sur la variété [3, Section 2.5], puis pour chaque $k \in \{1, \dots, N\}$, une matrice *antisymétrique* Ω_k de taille $n \times n$ est construite à partir des $\frac{n(n-1)}{2}$ entrées d'un vecteur normal centré réduit, i.e. un vecteur de loi $\mathcal{N}(\mathbf{0}, \mathbf{I}_{n(n-1)/2})$. Chaque point \mathbf{X}_k est obtenu par

$$\mathbf{X}_k = \exp_{\mathbf{C}}(\sigma \Omega_k) \mathbf{C}. \quad (15)$$

où $\sigma \in \mathbb{R}_+$ est un paramètre de dispersion autour du barycentre \mathbf{C} . Par la suite, les différentes estimations du barycentre sont notées ainsi : $\hat{\mathbf{X}}_{QR}$ et $\hat{\mathbf{X}}_P$ sont les estimées obtenues via les R-barycentres respectivement basés sur la décomposition QR et la décomposition polaire (P), tandis que l'estimée obtenue via la méthode proposée est notée \mathbf{X}^* (11). Les résultats d'estimation présentés par la suite sont basés sur la distance entre le barycentre \mathbf{C} et son estimation \mathbf{Y} , donnée par

$$\delta(\mathbf{C}, \mathbf{Y}) := \|\mathbf{C} - \mathbf{Y}\| \quad (16)$$

où $\|\cdot\|$ est la norme de Frobenius.

La Figure 2 présente des résultats obtenus pour la variété $\text{St}_{3,1}$ (i.e. la sphère unité) avec une taille d'échantillon fixe de $N = 30$. Dans ce cas de figure, les estimations $\hat{\mathbf{X}}_{QR}$ et $\hat{\mathbf{X}}_P$ coïncident. Pour une valeur faible de σ , Figure 2-(a), toutes les méthodes donnent des estimations proches. En revanche, lorsque la dispersion du nuage de points augmente, Figure 2-(b), les méthodes basées sur le R-barycentre sont plus sensibles que la méthode proposée basée sur un barycentre projeté. Cette robustesse (expérimentale) de l'approche par projection est observée également pour des valeurs de n et p croissantes.

La Figure 3 illustre la variabilité des estimées avec la taille de l'échantillon. Les résultats sont présents pour la variété $\text{St}_{20,4}$. Pour chaque méthode, les courbes donnent la distance médiane entre le *vrai* barycentre \mathbf{C} et son estimation au sens de (16). Les régions colorées donnent l'intervalle entre les quantiles 0.05 et 0.95, sur 500 réalisations de l'échantillon. Encore une fois, pour un nuage très concentré, Figure 3-(a), toutes les méthodes se comportent de manière similaire. Pour des valeurs de σ plus importantes, Figures 3-(b) et 3-(c), la méthode basée sur la projection proposée est la seule à avoir une erreur dont la décroissance est proportionnelle à la taille de l'échantillon.

Finalement, la Figure 4 illustre le fait que la méthode basée sur la projection a également l'avantage d'avoir un coût calculatoire bien moindre que les méthodes de type R-barycentre. Cela vient du fait que les méthodes fondées sur le R-barycentre ont besoin, pour chaque itération, de la résolution d'un système matriciel. Cet effet est évidemment accentué si n et p deviennent très grands. Au contraire, la méthode que nous proposons nécessite un calcul de barycentre dans l'espace ambiant et le calcul d'une décomposition polaire, ce qui évite les problèmes de convergence (vitesse et précision) et n'est pas sensible à l'initialisation.

À court terme, la suite des travaux présentés dans cet article consistera en la généralisation de ce type de barycentres projetés à d'autres variétés matricielles utilisées en traitement des données en grande dimension et l'application de ces techniques à des tâches de traitement du signal statistique.

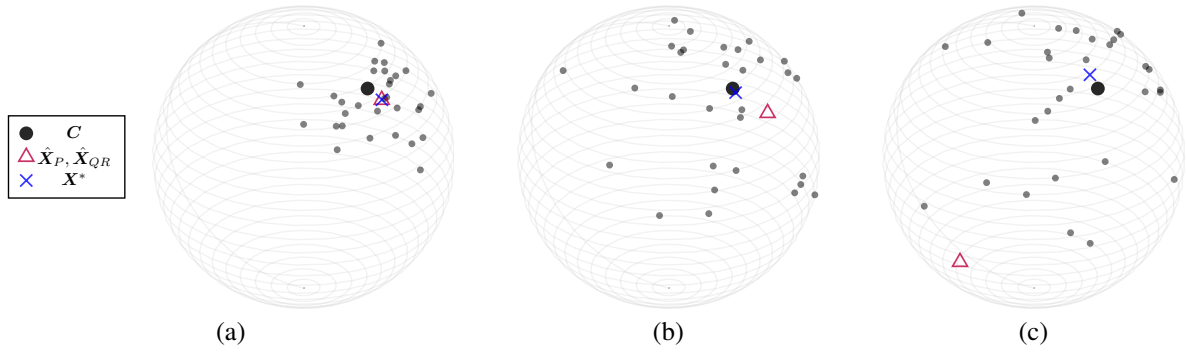


FIGURE 2 : Comparaison de l'estimation du barycentre par les méthodes R-barycentres QR (\hat{X}_{QR}) et Polaire (\hat{X}_P), et par notre méthode basée projection (X^*) pour un échantillon de taille 30 sur la sphère $St_{3,1}$, avec (a) : $\sigma = 0.3$; (b) : $\sigma = 0.5$; (c) : $\sigma = 0.6$.

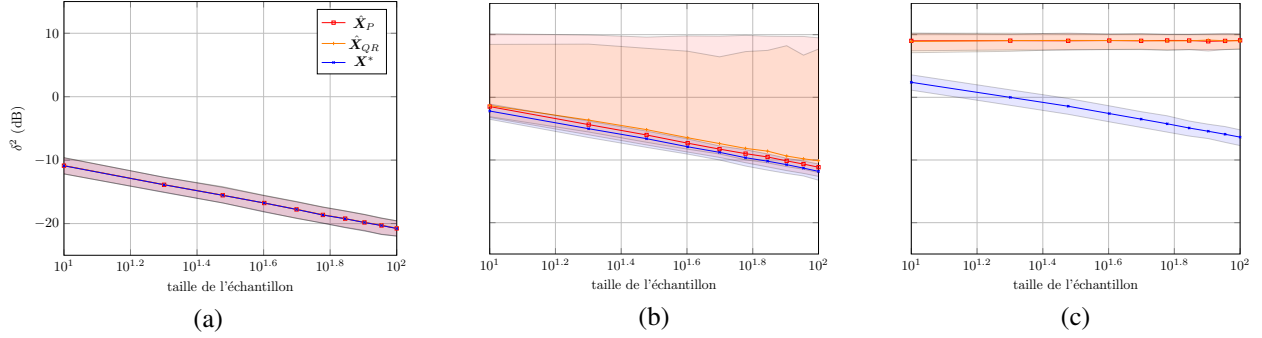


FIGURE 3 : Erreur δ^2 entre l'estimation du barycentre et C par les méthodes R-barycentres QR (\hat{X}_{QR}) et Polaire (\hat{X}_P), et par notre méthode basée projection (X^*) en fonction de la taille de l'échantillon N sur $St_{20,4}$, avec (a) : $\sigma = 0.1$; (b) : $\sigma = 0.2$; (c) : $\sigma = 0.3$.

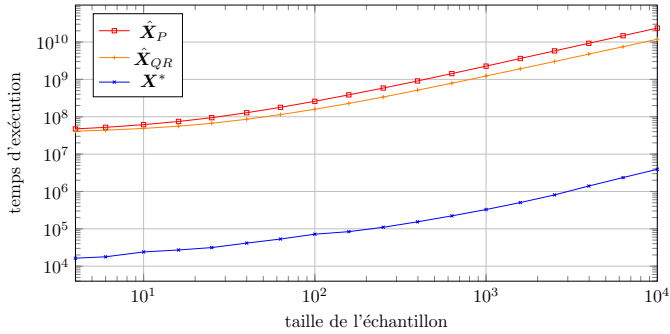


FIGURE 4 : Estimation du barycentre sur $St_{40,4}$. Vitesse d'exécution des algorithmes (en ns). Les algorithmes itératifs (R-barycentres) sont initialisés avec X^* .

Références

- [1] Pierre-Antoine ABSIL, Robert MAHONY et Rodolphe SEPULCHRE : Optimization algorithms on matrix manifolds. *In Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2009.
- [2] Sajid AHMED, John S THOMPSON, Yvan R PETILLOT et Bernard MULGREW : Unconstrained synthesis of covariance matrix for mimo radar transmit beampattern. *IEEE Transactions on Signal Processing*, 59(8):3837–3849, 2011.
- [3] Yasuko CHIKUSE : *Statistics on special manifolds*, volume 1. Springer, 2003.
- [4] Min HAN, Shoubo FENG, CL Philip CHEN, Meiling XU et Tie QIU : Structured manifold broad learning system : A manifold perspective for large-scale chaotic time series analysis and prediction. *IEEE Transactions on Knowledge and Data Engineering*, 31(9):1809–1821, 2018.
- [5] Tetsuya KANEKO, Simone FIORI et Toshihisa TANAKA : Empirical arithmetic averaging over the compact stiefel manifold. *IEEE Transactions on Signal Processing*, 61(4): 883–894, 2012.
- [6] Jonathan H MANTON : A globally convergent numerical algorithm for computing the centre of mass on compact lie groups. *In ICARCV 2004 8th Control, Automation, Robotics and Vision Conference, 2004.*, volume 3, pages 2211–2216. IEEE, 2004.
- [7] Stephen WRIGHT, Jorge NOCEDAL *et al.* : Numerical optimization. *Springer Science*, 35(67-68):7, 1999.
- [8] Ralf ZIMMERMANN et Knut HUPER : Computing the riemannian logarithm on the stiefel manifold : Metrics, methods, and performance. *SIAM Journal on Matrix Analysis and Applications*, 43(2):953–980, 2022.