



HAL
open science

Géométrie Riemannienne appliquée à la réduction de la dimension de signaux EEG pour les interfaces cerveau-machine

Pedro Luiz Coelho Rodrigues, Florent Bouchard, Marco Congedo, Christian Jutten

► **To cite this version:**

Pedro Luiz Coelho Rodrigues, Florent Bouchard, Marco Congedo, Christian Jutten. Géométrie Riemannienne appliquée à la réduction de la dimension de signaux EEG pour les interfaces cerveau-machine. GRETSI 2017 - XXVIème Colloque francophone de traitement du signal et des images, Cédric Richard; Marc Antonini, Sep 2017, Juan-Les-Pins, France. hal-01591252

HAL Id: hal-01591252

<https://hal.science/hal-01591252v1>

Submitted on 21 Sep 2017

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.

Géométrie Riemannienne appliquée à la réduction de la dimension de signaux EEG pour les interfaces cerveau-machine

Pedro L. C. RODRIGUES, Florent BOUCHARD, Marco CONGEDO, Christian JUTTEN

GIPSA-lab, CNRS, Université Grenoble Alpes, Grenoble INP
11, Rue des Mathématiques, 38400 Saint-Martin-d'Hères, France
pedro.rodrigues@gipsa-lab.fr

Résumé – Au cours des dernières années, le domaine des interfaces cerveau-machine (*brain-computer interfaces*, BCI) fondées sur l'électroencéphalographie (EEG) a connu un intérêt croissant. Les méthodes de l'état de l'art utilisent les concepts de la géométrie Riemannienne qui permettent de prendre en compte la structure intrinsèque des matrices de covariance des signaux EEG pour la classification dans différents paradigmes BCI. Cependant, elles montrent leurs limites lorsque le nombre d'électrodes augmente à cause de la complexité algorithmique et du mauvais conditionnement des matrices. Dans cet article, nous montrons qu'il peut être avantageux d'utiliser au préalable des méthodes de réduction de dimension afin d'augmenter les performances de classification. Nous évaluons deux algorithmes différents et comparons leurs performances face à des modèles sans réduction dans des tâches de classification de signaux EEG.

Abstract – In the past few years, there has been an increasing interest among the brain-computer interface (BCI) community in classification algorithms that respect the intrinsic geometry of the spatial covariance matrices of electroencephalographic (EEG) recordings. These methods are based on concepts of Riemannian geometry and, despite demonstrating good performance on several occasions, do not scale well when the number of electrodes increases. In this paper, we show that it may be advantageous to reduce the dimensions of the covariance matrices before performing the classification tasks. We evaluate two methods for reducing the dimension of the covariance matrices of EEG recordings and compare their classification performance on BCI data to models without any dimension reduction.

1 Introduction

Ces dernières années, de nouveaux algorithmes utilisant des concepts de la géométrie Riemannienne ont démontré des performances remarquables pour la classification de signaux électroencéphalographique (EEG) pour des interfaces cerveau-machine (*brain-computer interfaces*, BCI) [1]. Ces signaux, enregistrés à l'aide d'électrodes placées sur le cuir chevelu d'un sujet, permettent de caractériser des tâches cognitives et de les séparer en différentes classes. L'objectif est d'entraîner un modèle sur un ensemble d'essais pré-étiquetés (dont on connaît les classes) puis l'utiliser pour classifier des essais de classes inconnues.

L'idée centrale des algorithmes « géométriquement inspirés » est d'utiliser des matrices de covariance spatiale des signaux EEG comme descripteurs (*features*) du modèle statistique. Pour l'EEG, ces matrices appartiennent à la variété des matrices symétriques positives définies (SPD), qui possède une géométrie intrinsèque que les algorithmes de classification prennent en compte. Cependant, la complexité numérique de ces méthodes peut être problématique lorsque l'on considère des données EEG avec beaucoup de capteurs. En effet, elles reposent sur des décompositions en valeurs/vecteurs propres, dont la complexité de calcul est de l'ordre de n^3 , où n est le nombre d'électrodes.

Heureusement, la nature même des enregistrements EEG

nous permet de considérer un sous-espace des données tout en préservant la quasi-totalité de l'information. Ceci est rendu possible d'une part en raison de la forte corrélation statistique entre les signaux enregistrés par des électrodes proches, et d'autre part par le faible nombre de sources cérébrales indépendantes actives en relation à la tâche. En exploitant cette redondance, nous pouvons réduire les dimensions des matrices de covariance spatiale et utiliser des algorithmes basés sur la géométrie Riemannienne plus efficacement.

Dans cette communication, nous avons considéré deux méthodes non supervisées pour réduire la dimension de signaux EEG d'une base de données de BCI. La première est une extension de l'analyse en composantes principales (*principal component analysis*, PCA) aux signaux multivariés. L'autre est une méthode proposée par [2] et basée sur des concepts d'optimisation sur les variétés [3]. Nous avons appliqué ces deux méthodes comme étape de pré-traitement dans des séquences de classification et avons comparé leurs performances à celles de classification sans réduction de dimension préalable.

2 Méthodes

On note $X_k \in \mathbb{R}^{n \times T}$ le k -ème essai d'un enregistrement de T échantillons sur n électrodes EEG et y_k l'étiquette indiquant

sa classe. La covariance spatiale associée est définie par

$$C_k = \frac{1}{T} X_k X_k^T. \quad (1)$$

2.1 Géométrie Riemannienne des matrices symétriques définies positives

Lorsque l'on a un nombre suffisamment grand d'échantillons disponibles, les matrices de covariance estimées via (1) sont SPD et forment une variété notée \mathcal{S}_n^{++} . En équipant cette variété avec une métrique Riemannienne invariante par transformations affines, on obtient la distance géodésique

$$\delta(C_i, C_j) = \|\log(C_i^{-1/2} C_j C_i^{-1/2})\|_F, \quad (2)$$

qui apporte une notion de distance plus appropriée que la distance Euclidienne par rapport à la géométrie de \mathcal{S}_n^{++} [4]. En effet, on peut démontrer que la variété des matrices SPD a une courbure négative [4] et que certains concepts usuels de la géométrie Euclidienne n'y sont pas valables ; par exemple, la somme des angles d'un triangle n'est pas 180 degrés.

On définit le centre de masse M d'un ensemble de matrices de covariance $\{C_1, \dots, C_K\}$ pour la distance (2) par

$$M = \operatorname{argmin}_{M \in \mathcal{S}_n^{++}} \sum_{k=1}^K \delta^2(M, C_k). \quad (3)$$

Notons que M est le point de la variété qui minimise la dispersion (variance) de l'ensemble des matrices pour la distance (2). Si $n = 1$ (C_k est un scalaire strictement positif), M est la moyenne géométrique des matrices C_k . C'est pourquoi la matrice M est appelée moyenne géométrique de $\{C_1, \dots, C_K\}$ [1].

2.2 Réduction de la dimension

Pour réduire la dimension des matrices de covariance d'un ensemble $\{C_1, \dots, C_K\}$ dans \mathcal{S}_n^{++} , nous cherchons une transformation linéaire qui donne un nouvel ensemble noté $\{C_1^\downarrow, \dots, C_K^\downarrow\}$ dans \mathcal{S}_p^{++} ($p < n$) tout en conservant le maximum d'information selon un critère fixé. Nous sommes intéressés par un sous-espace de dimension p dans \mathbb{R}^n qui permet de conserver au maximum les propriétés géométriques des matrices $\{C_1, \dots, C_K\}$. Ce sous-espace est représenté par une matrice $W \in \mathbb{R}^{n \times p}$ dont les colonnes forment une base du sous-espace d'intérêt.

Nous utilisons W pour sélectionner des combinaisons linéaires d'électrodes de X_k via

$$X_k^\downarrow = W^T X_k,$$

dont les matrices de covariance sont

$$C_k^\downarrow = W^T C_k W \in \mathbb{R}^{p \times p}. \quad (4)$$

Sans perte de généralité, nous imposons que W est une matrice orthonormale. Notez que les matrices de dimension réduite sont toujours positives définies car W est de rang plein.

Nous considérons deux méthodes pour choisir la matrice W . La première est basée sur la PCA. Nous calculons la moyenne arithmétique de l'ensemble $\{C_1, \dots, C_K\}$,

$$\bar{C} = \frac{1}{K} \sum_{k=1}^K C_k, \quad (5)$$

et fixons les colonnes de W comme les vecteurs propres associés aux p plus grandes valeurs propres de \bar{C} . Cependant, cette approche ne prend pas en compte la géométrie intrinsèque des matrices de covariances dans \mathcal{S}_n^{++} . L'autre procédure, proposée par [2], est écrite sous la forme d'un problème d'optimisation,

$$\operatorname{argmax}_{W \in \mathbb{R}^{n \times p}} \sum_{k=1}^K \delta^2(W^T C_k W, W^T M W), \quad (6)$$

$$\text{t.q.} \quad W^T W = I_p,$$

où la fonction de coût choisie est une mesure de dispersion des matrices $\{C_1, \dots, C_K\}$ autour de leur moyenne géométrique.

Le problème (6) peut être résolu en utilisant des méthodes d'optimisation sur les variétés [3]. Nous avons utilisé les implémentations de la librairie `pymanopt` [5], écrite en Python, pour effectuer nos calculs. Nous avons fait appel aux routines basées sur la méthode de gradient conjugué adaptée à l'optimisation sur variétés. Les matrices W ont été traitées comme des éléments de la variété de Grassmann [3]. Le lecteur intéressé trouvera plus d'informations dans [2] et [3].

2.3 Séquences de classification

Nous classons chaque essai X_k à l'aide de l'algorithme MDM (*minimum distance to mean*) [1]. Cet algorithme comporte deux étapes : pendant la phase d'entraînement, on estime les moyennes géométriques des matrices de covariance de chaque classe connue. Ensuite, pour chaque matrice à étiquette inconnue, nous calculons sa distance aux moyennes géométriques correspondant aux différentes classes et l'affectons à la classe dont la moyenne géométrique est la plus proche selon la distance (2).

Nous considérons cinq séquences de classification différentes :

MDM : classification via MDM et sans réduction de dimension préalable.

PCA + MDM : réduction de dimension avec la PCA et classification via MDM.

HRD + MDM : réduction de dimension via la minimisation de (6) et classification via MDM.

SELg + MDM : réduction de dimension en choisissant les électrodes les plus proches des régions actives dans le cerveau lors des tâches cognitives de chaque classe (SELg : *good selection*). Classification via MDM.

SELb + MDM : réduction de dimension en choisissant des électrodes mal localisées et qui ne donnent pas beaucoup d'information pour discriminer chaque classe (SELb : *bad selection*). Classification via MDM.

Les performances de classification de chaque séquence sont évaluées via validation croisée.

2.4 Base de données

Nos analyses ont été faites sur une base de données contenant des enregistrements de signaux EEG correspondant à de l'imagerie motrice (*motor imagery*, MI) [6]. Pour chaque essai, le sujet a été prié d'imaginer pendant deux secondes qu'il était en train de lever sa main gauche ou sa main droite. La base de données contient les résultats de 109 sujets avec 64 électrodes. Chaque sujet a réalisé 44 essais (22 pour chaque classe). Avant d'effectuer les tâches de réduction et de classification, nous avons filtré les signaux entre 8 et 30 Hz avec un filtre IIR Butterworth d'ordre 5 et estimé leurs matrices de covariance comme décrit dans (1).

Lors de la sélection d'électrodes pour SELg + MDM et SELb + MDM, nous avons pris en compte que le cortex moteur est la région la plus active pour la MI. Nous avons choisi les 12 électrodes les plus proches du cortex moteur pour SELg + MDM, tout en gardant une symétrie gauche/droite. Les noms des électrodes sont : F3, Fz, F4, FC1, FC2, C3, Cz, C4, CP1, CP2, P3, P4. Les 12 électrodes de SELb + MDM ont été choisies sans préserver la symétrie et ne couvrent pas le cortex moteur du sujet. Les électrodes sont FPz, FP1, AFz, AF3, AF7, F7, F5, F3, F1, FT7, FC5, FC3.

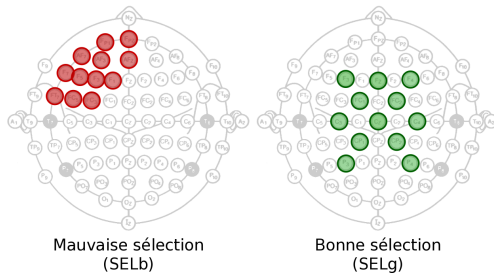


FIGURE 1 – Schéma avec la localisation des 64 électrodes de la base Physionet. Les électrodes à gauche sont celles qui ont été sélectionnées pour la séquence de classification SELb + MDM et à droite pour SELg + MDM.

3 Résultats et discussion

Nous avons comparé les performances des séquences de classification décrites dans la Section 2.3. La Figure 2 montre les résultats de classification avec HRD+MDM, PCA+MDM et MDM pour trois valeurs de la dimension p des matrices de covariance réduites. La Figure 3 compare les résultats de toutes les séquences avec $p = 12$. Dans les graphes, chaque point correspond à un sujet et les axes indiquent les résultats de classification (*accuracy*) de chaque méthode.

Grâce au grand nombre de sujets disponibles dans la base de données, nous avons pu réaliser des tests statistiques pour

étudier s'il y avait une méthode qui donnait de meilleurs résultats en moyenne. Pour cela, nous avons effectué une régression linéaire pour chaque paire de méthodes des Figures 2 et 3 en forçant l'ordonnée à l'origine à zéro. Si une des méthodes est en moyenne supérieure à l'autre, le modèle estimé doit avoir un coefficient directeur (\hat{m}) différent de 1 ; si la méthode représentée à l'ordonnée est supérieure, \hat{m} sera plus grand que 1, sinon, il est plus petit. Si nous ne pouvons pas rejeter l'hypothèse que le coefficient directeur est égal à 1, alors les méthodes sont dites équivalentes. Nous avons fait des tests statistiques du type F pour chaque paire de méthodes. En fixant le seuil de rejection de l'hypothèse nulle (\mathcal{H}_0 : "coefficient angulaire égal à un") à 1%, nous avons obtenu les résultats montrés dans le Tableau 1.

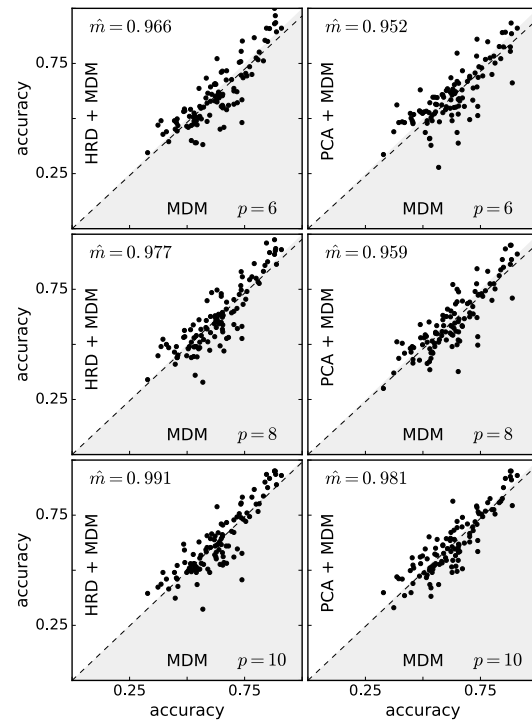


FIGURE 2 – Comparaison des résultats de classification avec réduction de dimension face à des séquences sans réduction. Nous considérons plusieurs valeurs pour la dimension p des matrices de covariances réduites. Chaque point dans les graphes correspond aux résultats de classification avec un sujet de la base de données.

À partir de ces résultats, on conclut que pour une réduction de dimension très forte, comme $p = 6$ et $p = 8$, les performances des séquences de classification sont inférieures à celles utilisant toutes les électrodes, comme indiqué par \hat{m} plus petit que 1. Ceci est probablement lié à une perte d'information qui aurait été importante pour bien séparer les classes. Pour $p = 10$, la méthode HRD donne des résultats équivalents à la classification avec toutes les électrodes, alors que la PCA reste inférieure. Un tel résultat montre qu'utiliser la géomé-

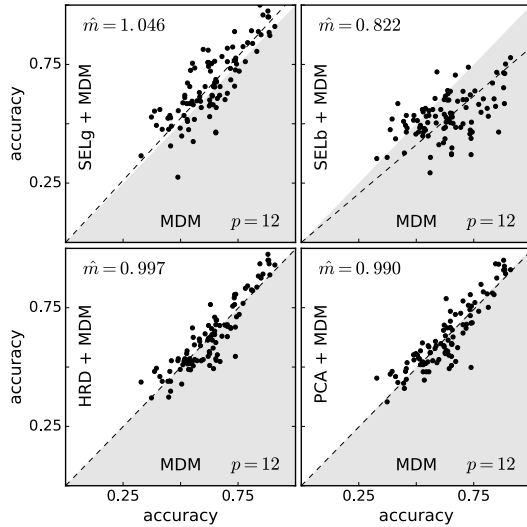


FIGURE 3 – Comparaison des résultats de classification avec réduction de dimension face à des séquences sans réduction. Nous fixons la dimension des covariances réduites à $p = 12$. Chaque point dans les graphes correspond aux résultats de classification avec un sujet de la base de données.

TABLE 1 – Résultats des tests statistiques pour chaque séquence de classification par rapport à la séquence MDM. Le seuil pour réjection de l’hypothèse nulle est fixé à 1%.

	p	\hat{m}	rej. H_0 ?		p	\hat{m}	rej. H_0 ?
HRD+MDM	6	0.966	OUI	PCA+MDM	6	0.952	OUI
	8	0.977	OUI		8	0.959	OUI
	10	0.991	NON		10	0.981	OUI
	12	0.997	NON		12	0.99	NON
SELg+MDM	12	1.046	OUI	SELb+MDM	12	0.822	OUI

trie intrinsèque des matrices de covariance peut être utile pour perdre moins d’information lorsque l’on réduit leur dimension. Pour $p = 12$, les deux méthodes deviennent équivalentes à l’estimation sans réduction de dimension. Cela nous indique qu’il existe un seuil pour p à partir duquel les réductions de dimension gardent suffisamment d’information pour garantir des bonnes performances de classification.

Les résultats avec SELb+MDM et SELg+MDM illustrent les deux extrêmes possibles lorsqu’on réduit les matrices de covariance à un même nombre de dimensions. D’une part, SELb+MDM sélectionne seulement des électrodes qui n’apportent pas beaucoup d’information pour classifier les essais et sont toutes très proches les unes des autres. Par conséquent, la performance de cette méthode est inférieure à celle avec toutes les électrodes. D’autre part, avec SELg+MDM on atteint des résultats supérieurs à ceux de MDM, comme indiqué par \hat{m} supérieur à 1 dans le Tableau 1. Ceci s’explique en grande partie par le choix d’électrodes qui apportent le plus d’information physiologique et sont les plus utiles pour séparer les deux classes, alors qu’avec 64 électrodes on garde des signaux qui

ne représentent que du bruit.

Il est important de remarquer que les méthodes HRD+MDM et PCA+MDM sont automatiques et ne prennent en compte aucune information sur la base de données pour réduire les dimensions des matrices de covariance. Ceci n’est pas le cas avec SELg+MDM, où la sélection des électrodes est faite à partir d’informations a priori sur la base de données.

4 Conclusion

Nous avons démontré la possibilité de réduire la dimension des matrices de covariance pour classifier des signaux EEG en conservant de bonnes performances. De plus, nous avons constaté qu’en sélectionnant les électrodes contenant le plus d’informations physiologiques, les performances deviennent même supérieures à celles des algorithmes sans réduction de dimension. Pour la suite, nous envisageons de proposer des extensions de la PCA et HRD capables de prendre aussi en compte les informations physiologiques des signaux EEG. Le but est d’avoir des méthodes automatiques de réduction de dimension sans avoir à choisir à la main les électrodes les plus importantes.

5 Remerciements

Ce travail a été partiellement financé par le projet ERC CHES (2012-ERC-AdG-320684). Les auteurs remercient Victor Maurandi pour l’aide avec la rédaction du papier.

Références

- [1] Congedo, M., Barachant A., Bhatia R., Riemannian Geometry for EEG-based Brain Computer Interfaces; a primer and a review. 2017; 65(9) : 2211-2220.
- [2] Harandi, M, Salzmann, M., Hartley, R., Dimensionality Reduction on SPD Manifolds : The Emergence of Geometry-Aware Methods, preprint arXiv :1605.06182.
- [3] Absil, P.-A., Mahony, R., Sepulchre, R. Optimization Algorithms on Matrix Manifolds, Princeton University Press, Princeton, United States (2007).
- [4] Bhatia, R. Positive Definite Matrices, Princeton University Press, Princeton, United States (2007).
- [5] Townsend, J., Koep, N., Weichwald, S., Pymanopt : A Python Toolbox for Manifold Optimization using Automatic Differentiation, preprint arXiv :1603.03236
- [6] Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer N., Wolpaw, J. R., BCI2000 : A general-purpose brain-computer interface (BCI) system, IEEE Transactions on Biomedical Engineering. 2004;51(6) :1034-1043.