

# Réduction de dimension pour la Séparation Aveugle de Sources

Florent BOUCHARD<sup>1</sup>, Pedro RODRIGUES<sup>1</sup>, Jérôme MALICK<sup>2</sup>, Marco CONGEDO<sup>1</sup>

<sup>1</sup>Gipsa-lab, CNRS, Univ. Grenoble Alpes, Grenoble INP  
11 rue des Mathématiques, 38400 Saint-Martin d'Hères, France

<sup>2</sup>Laboratoire Jean Kuntzmann, CNRS, Univ. Grenoble Alpes  
700 Avenue Centrale, 38400 Saint-Martin d'Hères, France

florent.bouchard@gipsa-lab.fr

**Résumé** – Dans cet article, nous considérons le problème de la réduction de dimension dans le cadre de la séparation aveugle de sources. Nous proposons (1) un nouveau modèle basé sur la diagonalisation par blocs conjointe qui permet de retrouver simultanément les sources et leur sous-espace et (2), une approche par optimisation Riemannienne pour le résoudre, originale dans ce contexte. Notre modèle produit des sources indépendantes du bruit et moins sensibles à l'estimation a priori du nombre de sources. L'efficacité de notre modèle par rapport à l'état de l'art est illustré sur la séparation de sources d'un enregistrement électroencéphalographique.

**Abstract** – In this article, we consider the dimension reduction problem in the context of blind source separation. We propose (1) a new model based on approximate joint block diagonalization, which allows to find the sources along with their subspace simultaneously, and (2), a Riemannian optimization approach to solve it, original in this context. Our model recovers sources that are independent from the noise and less sensitive to the a priori estimation of the number of sources. The efficiency of our model as compared to the state of the art is illustrated through the source separation of an electroencephalographic recording.

## 1 Introduction

La séparation aveugle de sources (*blind source separation*, BSS) [1] s'avère être un outil majeur du traitement du signal et de l'analyse de données multivariées, avec des applications dans de nombreux domaines d'ingénierie comme les communications, le traitement d'images et l'analyse de signaux audio et biomédicaux. Nous référons le lecteur à [1] pour une revue complète de la théorie et des applications. Dans cet article, nous considérons la BSS linéaire instantanée qui repose sur le modèle de mixage [1]

$$\mathbf{x}(t) = A\mathbf{s}(t), \quad (1)$$

où  $t \in \{1, \dots, T\}$ ,  $\mathbf{x}(t) \in \mathbb{R}^n$  et  $\mathbf{s}(t) \in \mathbb{R}^p$  correspondent respectivement aux observations et aux sources (avec  $p \leq n$ ), et  $A \in \mathbb{R}^{n \times p}$  est la matrice de mixage, qui est de rang plein. L'objectif est de trouver une estimation  $\hat{\mathbf{s}}(t)$  des sources  $\mathbf{s}(t)$  en faisant uniquement l'hypothèse de leur *indépendance statistique* [1]. La BSS se réduit à chercher une matrice  $B \in \mathbb{R}^{p \times n}$  de rang plein telle que  $\hat{\mathbf{s}}(t) = B\mathbf{x}(t)$ .

Il est possible de résoudre la BSS en utilisant la diagonalisation conjointe (*approximate joint diagonalization*, AJD) de matrices contenant des statistiques des observations [1]. Étant donné un ensemble  $\{C_k\}$  de  $K$  matrices symétriques, aussi supposées positives définies (SPD) ici, on cherche  $B \in \mathbb{R}^{p \times n}$  telle que l'ensemble  $\{BC_kB^T\}$  contient des matrices aussi diagonales que possible. Le diagonalisateur conjoint  $B$  est obtenu par la minimisation d'un critère de diagonalité de l'ensemble  $\{BC_kB^T\}$ .

Bien souvent, seulement  $p < n$  sources ont une contribution significative dans les données et peuvent être correctement retrouvées. Dans ce cas, il est nécessaire de procéder à une réduction de la dimension de  $\mathbf{x}(t)$ . En effet, lorsqu'on estime un trop grand nombre de sources, certaines d'entre elles sont composées de bruit ou de sources de puissance trop faible pour être identifiées. Ces sources de bruit, dont on n'assume pas l'indépendance statistique, perturbent les méthodes de BSS, dégradant l'estimation des vraies sources.

Pour traiter ce problème de réduction de dimension, on peut distinguer deux familles de méthodes dans la littérature. La première solution est d'estimer au préalable le sous-espace des sources par un filtre  $W \in \mathbb{R}^{p \times n}$  à partir des statistiques des observations  $\mathbf{x}(t)$  puis d'effectuer la BSS sur  $W\mathbf{x}(t)$ . Le choix classique pour  $W$  est de prendre la solution de l'*analyse en composantes principales* [2] qui permet de sélectionner le sous-espace qui maximise l'énergie conservée. Récemment, la recherche s'est tournée vers l'utilisation de sous-espaces qui permettent de conserver au maximum la structure géométrique des statistiques des observations [3, 4]. Cependant, comme souligné dans [5], les erreurs d'estimation du sous-espace ne peuvent pas être corrigées au cours de la BSS qui suit la sélection de  $W$  pour ce type de méthodes, ce qui entraîne une accumulation des erreurs au cours de cette procédure en deux étapes.

Pour remédier à ce problème, [5] propose de chercher  $B$  dans  $\mathbb{R}^{p \times n}$  directement afin que le sous-espace puisse être modifié pendant l'étape de BSS. Néanmoins, la méthode proposée

dans cet article contient deux inconvénients majeurs. Le premier problème est que la matrice de démixage est supposée orthogonale, ce qui induit des erreurs irréversibles dans l'estimation de  $B$ . Les outils d'optimisation développés dans [6] permettent de chercher une matrice de rang plein dans  $\mathbb{R}^{p \times n}$ , ce qui corrige ce problème. Le second, qui est plus handicapant et souvent observé en pratique (résultats non montrés), est qu'il est possible de rater des sources et d'obtenir plusieurs fois la même source dans  $\hat{s}(t)$ . Pour comprendre d'où vient ce problème, considérons  $\mathbf{x}(t)$  contenant  $p = 2$  vraies sources telles que  $s_1(t) = \mathbf{b}_1 \mathbf{x}(t)$  et  $s_2(t) = \mathbf{b}_2 \mathbf{x}(t)$ . Prenons  $\mathbf{v}$  orthogonal à  $\mathbf{b}_1$  et  $\mathbf{b}_2$  et définissons  $\tilde{\mathbf{b}} = \alpha \mathbf{b}_1 + \beta \mathbf{v}$ . Ainsi  $\tilde{s}(t) = \tilde{\mathbf{b}} \mathbf{x}(t)$  correspond à une version bruitée de  $s_1(t)$ , la matrice  $B$  formée de  $\mathbf{b}_1$  et de  $\tilde{\mathbf{b}}$  est de rang plein par construction et on a

$$BC_k B^T = \begin{pmatrix} \mathbf{b}_1 C_k \mathbf{b}_1^T & \mathbf{b}_1 C_k (\alpha \mathbf{b}_1 + \beta \mathbf{v})^T \\ (\alpha \mathbf{b}_1 + \beta \mathbf{v}) C_k \mathbf{b}_1^T & \tilde{\mathbf{b}} C_k \tilde{\mathbf{b}}^T \end{pmatrix}. \quad (2)$$

Il est possible de tomber dans un minimum local du critère avec  $(\mathbf{v}, \alpha, \beta)$  tels que  $\mathbf{b}_1 C_k (\alpha \mathbf{b}_1 + \beta \mathbf{v})^T \approx 0$ . Dans ce cas, on ne retrouve donc pas  $s_2(t)$  mais seulement  $s_1(t)$  et  $\tilde{s}(t)$ .

Dans ce papier, nous proposons un nouveau modèle, basé sur la diagonalisation par blocs conjointe (*approximate joint block diagonalization*, AJBD), qui permet au sous-espace estimé des sources d'évoluer au cours de la BSS tout en évitant les possibles solutions dégénérées. L'autre avantage majeur de notre méthode est que, contrairement aux approches précédentes, nous tenons compte du bruit contenu dans les  $n - p$  dimensions restantes et essayons de l'éliminer des sources estimées. La section 2 contient la description de la méthode que nous proposons. Dans la section 3, nous testons cette méthode sur des données électroencéphalographiques (EEG) et nous la comparons avec une méthode de l'état de l'art. Enfin, des conclusions et perspectives sont données section 4.

## 2 Méthode

### 2.1 Modèle proposé

Plutôt que chercher une matrice de rang plein  $B$  dans  $\mathbb{R}^{p \times n}$  telle que les matrices  $BC_k B^T$  soient diagonales, nous proposons de chercher une matrice inversible  $B$  dans  $\mathbb{R}^{n \times n}$  de sorte que les matrices  $BC_k B^T$  soient diagonales par blocs de la forme

$$BC_k B^T = \begin{pmatrix} \Sigma_k & 0_{p, n-p} \\ 0_{n-p, p} & N_k \end{pmatrix}, \quad (3)$$

où  $\Sigma_k$  est dans  $\mathcal{D}_p^{++}$  (variété des matrices diagonales à éléments strictement positifs),  $N_k$  est dans  $\mathcal{S}_{n-p}^{++}$  (variété des matrices SPD) et  $0_{m,q}$  est la matrice nulle de dimension  $m \times q$ . Dans ce cas, la matrice  $B_p \in \mathbb{R}^{p \times n}$  composée des  $p$  premières lignes de  $B$  est telle que les matrices  $B_p C_k B_p^T$  sont diagonales, ainsi  $\hat{s}(t) = B_p \mathbf{x}(t)$  correspond aux sources. Les  $n - p$  lignes restantes sont des composantes de bruit dont les statistiques sont contenues dans les matrices  $N_k$ .

Comme on ne cherche pas à diagonaliser les matrices  $N_k$ , le bruit ne vient pas perturber les algorithmes de BSS (à condition

d'avoir une estimation appropriée de la dimension  $p$ ). De plus, la diagonalisation par blocs de (3) permet d'obtenir l'indépendance statistique entre les sources estimées et le bruit. Enfin, comme la recherche se fait sur l'ensemble des matrices inversible dans  $\mathbb{R}^{n \times n}$ , le sous-espace des sources estimées engendré par  $B_p$  peut évoluer au cours de la BSS et les erreurs d'estimation au moment de l'initialisation peuvent être corrigées.

Notons que le modèle de diagonalisation par blocs de (3) est un cas particulier de l'analyse en sous-espaces indépendants (*independent subspace analysis*, ISA), introduit dans [7]. L'extension de ce qui suit au cas général de l'ISA est immédiat, cependant nous ne la faisons pas pour ne pas alourdir le formalisme.

La résolution du problème d'AJBD résultant de (3) est obtenue par minimisation d'un critère de bloc-diagonalité de l'ensemble de matrices  $\{BC_k B^T\}$ . Plusieurs critères sont possibles ; en particulier on peut généraliser les critères de [8] sous la forme

$$f(B) = \sum_k w_k d(BC_k B^T, \Lambda_k(B)), \quad (4)$$

où  $w_k$  sont des poids positifs,  $d(\cdot, \cdot)$  est une divergence sur  $\mathcal{S}_p^{++}$  et  $\Lambda_k(B)$  est une matrice de la forme

$$\begin{pmatrix} \Lambda_p & 0_{p, n-p} \\ 0_{n-p, p} & S_{n-p} \end{pmatrix}, \quad (5)$$

avec  $\Lambda_p \in \mathcal{D}_p^{++}$  et  $S_{n-p} \in \mathcal{S}_{n-p}^{++}$ . L'ensemble des matrices de la forme (5) est dénoté  $\mathcal{D}_{n,p}^{++}$ . Tout comme  $\mathcal{D}_n^{++}$ , l'espace  $\mathcal{D}_{n,p}^{++}$  possède les mêmes propriétés géométriques que  $\mathcal{S}_n^{++}$  (même métrique, même géodésique, etc.). Dans cet article nous nous limiterons à la distance de Frobenius, la distance naturelle des moindres carrés (voir (8)). De plus, en définissant  $\text{bdiag}(\cdot)$  comme l'opérateur qui annule tous les éléments sauf les  $p$  premiers éléments diagonaux et les éléments du dernier bloc diagonal de dimension  $(n - p) \times (n - p)$ , on a également la propriété

$$\begin{aligned} \text{tr}(\text{bdiag}(P)Q) &= \text{tr}(\text{bdiag}(P) \text{bdiag}(Q)) \\ &= \text{tr}(P \text{bdiag}(Q)), \end{aligned} \quad (6)$$

où  $P$  et  $Q$  sont dans  $\mathbb{R}^{n \times n}$  et  $\text{tr}(\cdot)$  dénote la trace. De ce fait, l'ensemble du travail présenté dans [8] peut être généralisé pour l'AJBD et le choix naturel pour  $\Lambda_k(B)$  est la matrice diagonale par blocs la plus proche de  $BC_k B^T$  selon la divergence  $d(\cdot, \cdot)$

$$\Lambda_k(B) = \underset{\Lambda \in \mathcal{D}_{n,p}^{++}}{\text{argmin}} d(BC_k B^T, \Lambda). \quad (7)$$

### 2.2 Moindres carrés

Un des critères historique de l'AJD est basée sur la distance de Frobenius (aussi appelée méthode des moindres carrés) [9]. Il est possible d'utiliser cette distance dans (4) pour définir un critère de l'AJBD. La distance de Frobenius est définie par

$$d_F^2(C, \Lambda) = \|C - \Lambda\|_F^2, \quad (8)$$

où  $\|\cdot\|_F$  est la norme de Frobenius. La matrice  $\Lambda$  diagonale par blocs la plus proche de  $C \in \mathcal{S}_n^{++}$  selon cette distance est

$$\Lambda = \text{bdiag}(C). \quad (9)$$

Pour démontrer ce résultat, il suffit de calculer le gradient de  $\Lambda \mapsto d_F^2(C, \Lambda)$ , qui vaut  $2(\Lambda - \text{bdiag}(C))$ , et de l'annuler. Le critère à minimiser est donc

$$f_F(B) = \sum_k w_k \|BC_k B^T - \text{bdiag}(BC_k B^T)\|_F^2. \quad (10)$$

Pour optimiser (10), nous adaptons à notre situation la stratégie développée dans [10]. À chaque itération, on considère le sous-problème d'optimisation avec la variable auxiliaire  $H$  et le critère

$$\tilde{f}_F(H) = \sum_k w_k \|BC_k B^T - H \text{bdiag}(BC_k B^T) H^T\|_F^2. \quad (11)$$

Partant de  $H_0 = I_n$ , nous faisons une itération dans une direction de descente de (11) pour obtenir  $H$  puis nous mettons  $B$  à jour avec  $B \leftarrow H^{-1}B$  et nous réitérons cette procédure jusqu'à convergence (lorsque  $H$  reste proche de l'identité). L'idée derrière cette stratégie est qu'on cherche à rapprocher  $\text{bdiag}(BC_k B^T)$  de  $BC_k B^T$  avec  $H$  puis on inverse  $H$  pour rapprocher  $BC_k B^T$  de  $\text{bdiag}(BC_k B^T)$ . Nous renvoyons à [11, Algo 1] pour plus de détails sur cette méthode d'optimisation.

### 3 Expérience numérique

Pour tester la méthode développée ici, nous effectuons la BSS sur les données EEG d'un patient épileptique enregistré avec 19 électrodes placées selon le système international 10-20. L'enregistrement est échantillonné à 128Hz et filtré entre 1 et 32Hz. Nous utilisons uniquement 7 secondes pour nos analyses. La figure 1 montre les 5 dernières secondes de la fenêtre d'analyse pour 9 des 19 électrodes disponibles. Ici, nous nous intéressons à la source qui génère les trois pointes-ondes lentes visibles sur le scalp (zones grisées sur la figure 1).

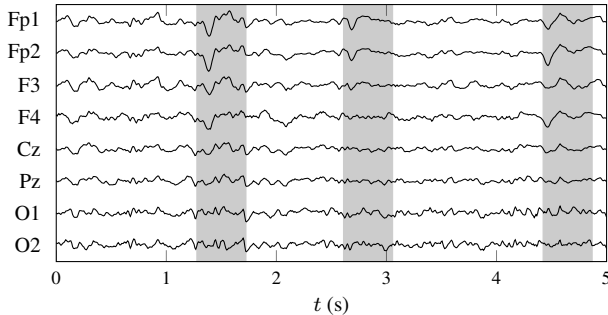


FIG. 1 – 5 dernières secondes de l'enregistrement EEG d'un patient épileptique pour 9 des 19 électrodes. On souhaite retrouver la source qui génère les pointes-ondes lentes présentes dans les zones grisées (principalement visibles sur les électrodes frontales).

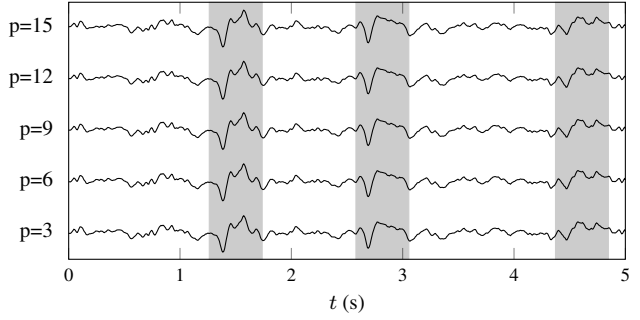


FIG. 2 – Sources des pointes-ondes lentes (zones grisées) estimées par notre méthode AJBD pour différentes valeurs de  $p$ .

Pour la BSS, nous nous servons de la procédure de [12]. En résumé, nous effectuons l'analyse sur les cospectres de Fourier calculés avec des fenêtres glissantes de une seconde avec un chevauchement de 75%, puis normalisées et pondérées. Nous blanchissons les données avec l'inverse de la racine carrée de la moyenne arithmétique des cospectres en réordonnant les lignes par ordre décroissant de puissance. En ce qui concerne la réduction de dimension, nous considérons  $p \in \{3, 6, 9, 12, 15\}$ . Nous comparons notre méthode, qu'on dénote AJBD, avec l'AJD classique (critère basé sur Frobenius, même méthode d'optimisation) des cospectres réduits où on garde seulement les  $p$  composantes les plus puissantes. Cette méthode, qui est l'état de l'art actuellement, est dénotée DR+AJD (DR pour *dimension reduction*). Les résultats ont été obtenus sous Matlab avec la toolbox *manopt* [13]. Systématiquement, la source qui nous intéresse est toujours la troisième source retrouvée par les deux méthodes.

La figure 2 contient les sources estimées correspondant aux pointes-ondes lentes obtenues avec AJBD pour les différentes valeurs de  $p$ . On observe que dans ce cas, les sources extraites sont très similaires pour toutes les valeurs de  $p$ . Cela est confirmé par la figure 4 où l'on peut voir que la corrélation entre les sources obtenues pour l'ensemble des valeurs de  $p$  et la source obtenue avec  $p = 15$  est toujours très proche de un. Par conséquent, notre méthode est très robuste par rapport au choix de  $p$  sur ces données et pour cette source.

Sur la figure 3, où les sources extraites avec DR+AJD sont représentées, on peut remarquer des variations entre les sources obtenues pour les différentes valeurs de  $p$ , ce qui est également

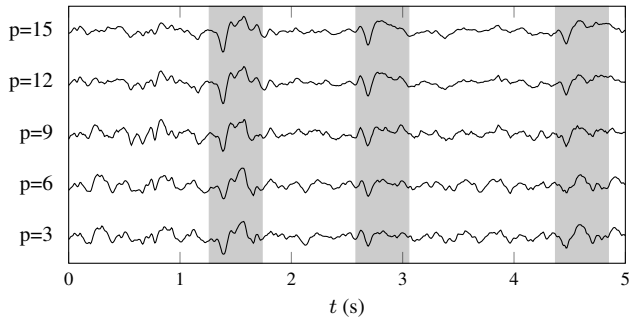


FIG. 3 – Sources des pointes-ondes lentes (zones grisées) estimées par la méthode DR+AJD pour différentes valeurs de  $p$ .

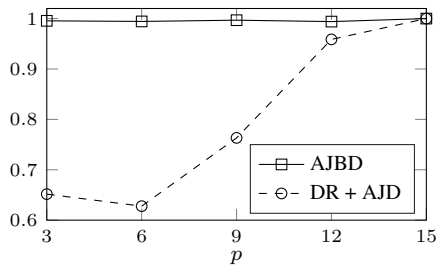


FIG. 4 – Corrélation de Pearson sur les 7 secondes d’enregistrement analysé entre les sources obtenues pour les différentes valeurs de  $p$  et la source obtenue avec  $p = 15$  pour chaque méthode. Notons que la corrélation entre les sources obtenues avec AJBD et DR+AJD pour  $p = 15$  est de 0.94.

confirmé par la figure 4. Pour l’ensemble des résultats, on arrive toujours à distinguer les trois pointes-ondes lentes mais, pour  $p \leq 9$ , elles semblent être mélangées avec du bruit visible en dehors des périodes d’activation de la source et qui s’intensifie lorsque  $p$  diminue. Donc dans ce cas, le choix de  $p$  est important et influence fortement la qualité de la source retrouvée. Visuellement, il semble que la meilleure des sources obtenues avec DR+AJD est celle correspondant à  $p = 15$ .

Cela confirme également que le sous-espace des sources estimées par AJBD évolue au cours du processus d’optimisation puisque l’initialisation des deux méthodes est la même, prouvant l’efficacité du modèle proposé sur ces données. Lorsqu’on compare la source retrouvée avec AJBD à celle extraite avec DR+AJD pour  $p = 15$ , il apparaît que la troisième pointe-onde lente est mieux estimée avec DR+AJD. Ces deux sources sont tout de même très similaires puisque leur corrélation de Pearson vaut 0.94 sur les 7 secondes d’enregistrement analysé.

## 4 Conclusions et perspectives

Dans cet article, nous avons proposé un modèle pour la réduction de dimension adapté à la BSS qui possède deux avantages majeurs. Le premier avantage est que le sous-espace des sources estimées n’est pas fixe et peut évoluer au cours de la séparation. Le second avantage est que nous tenons compte du bruit contenu dans les données et que nous essayons de l’éliminer des sources. Le désavantage est que nous devons chercher une matrice de démixage de dimension  $n \times n$ , où  $n$  est la dimension des observations, ce qui peut s’avérer problématique pour des données de grande dimension.

Un inconvénient de toute méthode de ce type est le choix arbitraire du nombre de sources  $p$ . En pratique, ce choix repose sur une démarche empirique, ce qui introduit un biais de l’expérimentateur qui va choisir une valeur qui va lui permettre de mettre en avant ses résultats. La méthode proposée dans cet article ne permet pas de trouver la valeur optimale de  $p$  mais, comme illustré avec les données EEG, les sources estimées dépendent peu du choix de  $p$ , ce qui limite grandement ce biais.

Cette approche inédite s’est montrée très performante sur l’exemple de données EEG que nous avons considéré par sa

flexibilité et sa robustesse quant au choix de la dimension  $p$  de l’espace des sources à extraire. Il nous faut à présent considérer d’autres divergences, comme celle de Kullback-Leibler ou la distance Riemannienne [8, 11], et effectuer une évaluation systématique sur de grosses bases de données. Notons finalement que ce travail montre une fois de plus l’avantage et la flexibilité des méthodes de BSS utilisant la diagonalisation conjointe.

## Remerciements

Ce travail a été supporté par le LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01), financé par le programme français “Investissement d’avenir”, et par l’ERC CHESS 2012-ERC-AdG-320684.

## Références

- [1] P. Comon and C. Jutten. *Handbook of Blind Source Separation : Independent Component Analysis and Applications*. Academic Press, 1st edition, 2010.
- [2] I. Jolliffe. *Principal component analysis*. Springer Series in Statistics, Springer, 2nd edition, 2002.
- [3] M. Harandi, M. Salzmann, and R. Hartley. Dimensionality reduction on SPD manifolds : The emergence of geometry-aware methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.
- [4] M. Congedo, P. Rodrigues, F. Bouchard, A. Barachant, and C. Jutten. A closed-form unsupervised geometry-aware dimensionality reduction method in the Riemannian manifold of SPD matrices. In *EMBC 2017*. IEEE, 2017.
- [5] F. J. Theis, T. P. Cason, and P.-A. Absil. Soft dimension reduction for ICA by joint diagonalization on the Stiefel manifold. In *Independent Component Analysis and Signal Separation*, pages 354–361. Springer, 2009.
- [6] F. Bouchard, L. Korczowski, J. Malick, and M. Congedo. Approximate joint diagonalization within the Riemannian geometry framework. In *24th European Signal Processing Conference (EUSIPCO-2016)*, pages 210–214, 2016.
- [7] P. Comon. Supervised classification, a probabilistic approach. In *ESANN-European Symposium on Artificial Neural Networks*, pages 111–128, 1995.
- [8] K. Alyani, M. Congedo, and M. Moakher. Diagonality measures of Hermitian positive-definite matrices with application to the approximate joint diagonalization problem. *Linear Algebra and its Applications*, 2016.
- [9] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non Gaussian signals. *IEE Proceedings-F*, 140(6) :362–370, dec 1993.
- [10] P. Tichavský and A. Yeredor. Fast approximate joint diagonalization incorporating weight matrices. *Signal Processing, IEEE Transactions on*, 57(3) :878–891, 2009.
- [11] F. Bouchard, J. Malick, and M. Congedo. Approximate joint diagonalization according to the natural Riemannian distance. In *13th International Conference on Latent Variable Analysis and Signal Separation*, pages 290–299. Springer, 2017.
- [12] M. Congedo, C. Gouy-Pailler, and C. Jutten. On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics. *Clinical Neurophysiology*, 119(12) :2677–2686, 2008.
- [13] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre. Manopt, a Matlab toolbox for optimization on manifolds. *Journal of Machine Learning Research*, 15 :1455–1459, 2014.