

# CNN-SPDNet : un réseau de neurones riemannien profond pour l'interprétation spatio-fréquentielle en EEG BCI

Gatien DARLEY<sup>1</sup> Stéphane BONNET<sup>1</sup>

<sup>1</sup>Université Grenoble Alpes, CEA, Leti, 38000 Grenoble, France

**Résumé** – Pour décoder les signaux d'électroencéphalographie (EEG), les réseaux riemanniens profonds tels que le SPDNet sont particulièrement performants car mêlant deux techniques : la géométrie riemannienne et l'apprentissage profond. Cet article vise à améliorer ce réseau avec l'ajout en amont d'une couche de convolution (CNN) pour apprendre de manière *end-to-end* des bandes de fréquences propres à chaque sujet. La contribution principale réside dans l'interprétation des poids appris par le réseau CNN-SPDNet : à la fois les filtres fréquentiels et les filtres — et patterns — spatiaux. Des résultats convaincants sont obtenus sur un jeu de compétition BCI avec un gain moyen de 5.4% pour la méthode proposée par rapport à un SPDNet opérant sur une bande de fréquences fixe. L'interprétation neurophysiologique des poids appris complète avantageusement ce type d'approche.

**Abstract** – For electroencephalography (EEG) signals decoding, deep riemannian networks, such as the SPDNet, are particularly effective as they combine two techniques: riemannian geometry and deep learning. This paper aims to enhance this network by introducing an upstream convolutional layer (CNN) to learn subject-specific frequency bands in an *end-to-end* manner. The main contribution lies in the interpretation of the learned weights of the CNN-SPDNet: both the frequency filters of the CNN and, more importantly, the spatial filters — and patterns — learned by the SPDNet. Convincing results are shown on a BCI competition dataset with a mean improvement of 5.4% with the proposed method compared to an SPDNet operating on a fixed frequency band. The neurophysiological interpretation of learned weights is a useful complement to this type of approach.

## 1 Introduction

Un domaine majeur en traitement de signaux d'électroencéphalographie (EEG) est l'interface cerveau-machine, ou Brain Computer Interface (BCI) en anglais, qui vise à traduire les signaux EEG en commandes à l'aide d'outils d'apprentissage automatique. Dans cet article, nous nous concentrons sur un BCI fondé sur l'imagerie motrice (IM) : il s'agit de l'imagination mentale de mouvements des membres sans aucune activation musculaire. Une telle tâche est suffisante pour observer l'activation cérébrale dans la zone du cortex sensorimoteur correspondante comme si un mouvement réel s'était produit. L'idée de tels BCI moteurs est donc d'associer une action dans le monde réel avec l'imagination de mouvements de différentes parties du corps [1], [8].

La particularité des signaux EEG est leur dépendance spatio-fréquentielle. Pour un EEG d'humain sain, la gamme de fréquences observées se situe environ entre 7 et 35 Hz en IM. Ces fréquences sont subdivisées en plusieurs groupes : alpha (8-12 Hz) et bêta (12-30 Hz) sont les deux principaux rythmes cérébraux pour l'imagerie motrice.

L'aspect spatial est capturé à travers de nombreuses électrodes placées sur le cuir chevelu. L'analyse EEG dépend fortement des sujets et des sessions d'enregistrement, on parle alors de variabilité inter-sujet et inter-session.

En EEG il devient de plus en plus courant de travailler avec les matrices de covariance comme descripteurs des signaux car celles-ci permettent de capturer efficacement les relations spatiales entre les différentes électrodes [1]. Ces matrices sont symétriques définies-positives (SPD), et présentent une structure de données non euclidienne car elles forment une variété riemannienne. Naturellement, des réseaux de neurones sur variétés ont alors vu le jour pour prendre en entrée ce type de

données structurées.

Cet article se focalise sur l'étude du réseau SPDNet [4]. Tout l'enjeu de ce réseau de neurones profond est de préserver la structure SPD à travers les couches pour l'apprentissage non linéaire. Si ce réseau a parfois été utilisé dans la littérature en EEG ([2], [10]), à la connaissance des auteurs, aucune étude n'interprète les poids appris. Or ils contiennent de l'information spatiale ou fréquentielle ce qui permet de vérifier que le réseau apprend effectivement et qu'il n'apprend pas sur du bruit. Surtout cela apporte de l'information d'un point de vue neurophysiologique.

Enfin, appliqué à de l'EEG, il est possible d'améliorer le SPDNet par l'apprentissage *end-to-end* de filtres fréquentiels en amont du réseau SPDNet [10]. Cela rend le réseau plus robuste et ne nécessite pas de connaître a priori la bande passante des signaux EEG qui varie entre les sujets.

Nos contributions sont donc les suivantes : (1) Développer et interpréter un réseau de neurones spécifiquement adapté pour de l'EEG basé sur le SPDNet via l'apprentissage *end-to-end* de bandes de fréquences. (2) Proposer une interprétation spatio-fréquentielle des poids appris par un SPDNet.

## 2 Méthode

### 2.1 Filtres et patterns spatiaux

Un aspect clef de l'interprétation des algorithmes de traitement de signaux EEG concerne l'analyse de filtres temporels et spatiaux et de patterns spatiaux. Cette section vise à expliquer mathématiquement ceux-ci.

Soit  $N_C$  le nombre de voies (i.e. d'électrodes) et  $N_T$  le nombre d'échantillons temporels respectivement. Étant donné un signal multivarié  $\mathbf{X} \in \mathbb{R}^{N_C \times N_T}$ , un **filtre spatial**  $\mathbf{w} \in \mathbb{R}^{N_C \times 1}$

transforme la donnée EEG de la manière suivante :

$$\mathbf{z} = \mathbf{w}^T \mathbf{X}$$

avec  $\mathbf{z} \in \mathbb{R}^{1 \times N_T}$  le signal filtré spatialement. Pour  $M$  filtres spatiaux on peut les regrouper dans une matrice de filtres spatiaux  $\mathbf{W} \in \mathbb{R}^{N_C \times M}$  de telle manière que

$$\mathbf{Z} = \mathbf{W}^T \mathbf{X}$$

où la matrice  $\mathbf{Z}$  contient les signaux filtrés pour chacun des  $M$  filtres et chaque colonne de  $\mathbf{W}$  est un filtre spatial. Ces filtres spatiaux assignent des poids aux électrodes selon un critère d'optimisation [5]. La covariance des signaux filtrés  $\mathbf{Z}$  est

$$\Sigma_z = \mathbf{W}^T \Sigma_x \mathbf{W} \quad (1)$$

où  $\Sigma_x$  est la covariance de  $\mathbf{X}$ .

On peut ensuite introduire la notion de **patterns spatiaux** en utilisant l'espace des caractéristiques utilisées pour la classification et la matrice des filtres. Les patterns spatiaux font référence à la distribution topographique du signal à travers le réseau d'électrodes et peuvent être extraits à partir des filtres spatiaux. En effet, puisque  $\mathbf{Z} = \mathbf{W}^T \mathbf{X}$ , on obtient, sous condition de (pseudo)-inversibilité :

$$\mathbf{X} = \mathbf{W}^{-T} \mathbf{Z}$$

où les colonnes de  $\mathbf{W}^{-T}$  sont les patterns spatiaux. Ainsi, chaque pattern spatial est une composante de l'activité cérébrale. La matrice  $\mathbf{W}^{-T}$  peut être vue comme une matrice de mélange pour les sources  $\mathbf{Z}$ . Donc l'activité cérébrale observée est une combinaison linéaire des patterns extraits.

## 2.2 SPDNet

Inventé en 2017 par Z.Huang et L.Van Gool [4], ce réseau de neurones à la particularité de prendre en entrée des matrices SPD et de rester sur la variété. Ainsi, chaque couche reçoit une matrice SPD en entrée et produit en sortie une matrice SPD (excepté la dernière couche dont le but est de sortir de la variété). L'idée derrière ce réseau est de transformer les matrices d'entrée pour réduire leur dimension et les rendre plus discriminatives en vue de les classifier. C'est le rôle de la couche BiMap décrite dans la section 2.2.1. À travers cet article, nous expliquons que pour des données EEG, interpréter les poids de cette couche a un sens neurophysiologique.

### 2.2.1 Description couche par couche

L'architecture implémentée du SPDNet est décrite Fig.1. En EEG, notamment en BCI, nous prenons comme descripteur de chaque essai la matrice de covariance spatiale du signal. Soit  $\mathbf{X} \in \mathbb{R}^{N_C \times N_T}$  l'essai d'entrée, celle-ci est estimée par l'estimateur SCM :  $\mathbf{C}_1 = \frac{1}{N_T-1} \mathbf{X} \mathbf{X}^T = \Sigma_x$ .

**BiMap ou filtrage spatial** La première couche, BiMap, permet de transformer la matrice de covariance du signal d'entrée en la covariance des signaux filtrés spatialement :  $\mathbf{C}_2 = \mathbf{W}_s^T \mathbf{C}_1 \mathbf{W}_s$ , où  $\mathbf{W}_s \in \mathbb{R}^{N_C \times M}$ , en réduisant la dimension. Ainsi, la couche BiMap s'interprète comme un filtrage spatial dans le cadre de l'EEG. Notre contribution est l'interprétation des poids appris  $\mathbf{W}_s$  de cette couche comme des

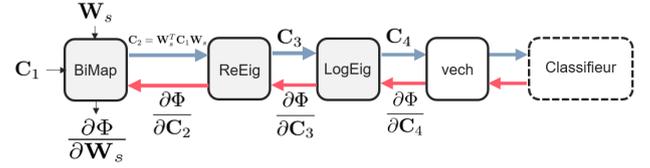


FIGURE 1 : Architecture SPDNet utilisée pour la classification. La propagation avant est représentée par les flèches bleues. Les opérations réalisées par chaque couche sont décrites en section 2.2.1. La rétropropagation du gradient est en rouge.  $\Phi$  est la fonction de coût. Les poids  $\mathbf{W}_s$  sont les paramètres appris par le réseau.

filtres spatiaux en faisant le lien avec (1). À la connaissance des auteurs, bien que ce réseau ait été souvent utilisé, cette interprétation n'a pas été faite auparavant. Cela permet pourtant d'expliquer les résultats d'apprentissage mais surtout cela apporte de l'information sur l'activité cérébrale. Avec cette méthode, des filtres spatiaux optimaux sont appris de manière *end-to-end*, à la différence d'autres méthodes non basées sur de l'IA [5].

**ReEig** Connaissant la décomposition en valeurs propres de  $\mathbf{C}_2 = \mathbf{U}_2 \Lambda_2 \mathbf{U}_2^T$ , cette couche implémente  $\mathbf{C}_3 = \mathbf{U}_2 \max(\Lambda_2, \epsilon \mathbf{I}_N) \mathbf{U}_2^T$  et apporte de la non-linéarité. Elle a pour but de rectifier les matrices SPD en ajustant leurs petites valeurs propres positives. Le nombre de blocs BiMap-ReEig est choisi expérimentalement. Pour notre étude nous implémentons un seul bloc car nous nous intéressons davantage à l'interprétation des poids et celle-ci est délicate pour les couches plus profondes à cause des non-linéarités.

**LogEig** Enfin, une couche pour sortir de la variété est nécessaire. En effet, le classifieur en sortie du SPDNet prend en entrée un vecteur et non une matrice. La couche LogEig implémente une opération de cartographie logarithmique à l'identité dans le cadre d'utilisation de la métrique AIRM i.e.  $\mathbf{C}_4 = \text{Log}_I(\mathbf{C}_3) = \log(\mathbf{C}_3)$ . Effectivement, cette opération permet de réduire la variété des matrices SPD à un espace euclidien [1]. Finalement, la sortie de la couche LogEig est remodelée en un vecteur — c'est le rôle de "vech" — qui nourrit le classifieur supervisé [1].

Le classifieur est une couche entièrement connectée avec une fonction *softmax*. La fonction de coût choisie est une fonction d'entropie croisée.

### 2.2.2 Rétropropagation

Il y a deux points importants pour généraliser la rétropropagation pour le SPDNet [4]. D'abord l'optimisation sur variété de Stiefel des poids de la couche BiMap  $\mathbf{W}_s$ . Ensuite, le calcul des gradients des fonctions matricielles.

## 2.3 SPDNet appliqué à l'EEG : le CNN-SPDNet

Dans le cadre de l'étude de signaux EEG, ceux-ci doivent être filtrés dans une bande de fréquences adaptée à l'imagerie motrice. Cependant, fixer une bande passante a priori n'est pas optimal puisque chaque sujet possède des caractéristiques fréquentielles qui lui sont propres. D'autant plus, cela nécessite

TABLE 1 : Résultats en termes de précision (en %) pour chaque sujet avec les modèles présentés sur la base de données BCI competition IV 2a. La précision moyenne est calculée sur tous les sujets et en conservant uniquement les sujets les plus répondants (\*) [11]. En gras figure le meilleur score pour chaque sujet.

Modèle	S1*	S2	S3*	S4	S5	S6	S7*	S8*	S9*	MOY (±SD)	MOY (±SD) sujets *
EEGNet-8,2	62.1	<b>43.4</b>	68.7	<b>50.7</b>	<b>45.2</b>	<b>44.8</b>	52.1	69.8	<b>79.8</b>	57.4 (±13.1)	66.5 (±11.0)
SPDNet6	62.5	42.4	71.5	47.9	30.1	39.2	48.3	74.6	73.3	54.4 (±16.5)	66.0 (±11.0)
CNN1-SPDNet6	<b>72.9</b>	40.6	<b>86.8</b>	44.1	36.1	43.8	<b>63.6</b>	<b>75.3</b>	75.4	59.8 (±18.8)	<b>74.8</b> (±8.3)

de connaître en amont la bande de fréquences idéale pour l'application étudiée. C'est pourquoi nous développons, en nous appuyant sur les travaux de D. Wilson [10], le CNN-SPDNet, une méthode qui permet d'apprendre les filtres fréquentiels optimaux en *end-to-end*.

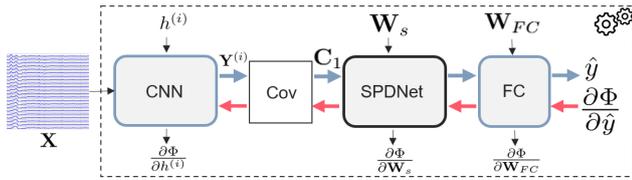


FIGURE 2 : CNN-SPDNet.  $h^{(i)}$ ,  $W_s$  et  $W_{FC}$  sont les poids appris. La couche « cov » calcule la covariance des signaux temporels filtrés.

Pour cela est ajouté en amont du SPDNet une couche de convolution temporelle (ou CNN) Fig.2 qui extrait des filtres FIR adaptés et donc améliore la généralisabilité. Nous réalisons une convolution 2-d avec un seul filtre 1-d pour toutes les voies. Cela réduit le nombre de paramètres par rapport à l'apprentissage d'un filtre par voie. Il est possible d'apprendre plusieurs bandes de fréquences, i.e. plusieurs noyaux de convolution. En notant  $Y$  les signaux filtrés en sortie du CNN,  $I$  le nombre de filtres fréquentiels appris et  $*_2$  la convolution temporelle 2-d, on obtient :  $Y^{(i)} = X *_2 h^{(i)}$ .

Pour cette étude, 1 filtre fréquentiel et 6 filtres spatiaux sont appris, nous utilisons alors la notation CNN1-SPDNet6. J. Paillard *et al.* ou D. Wilson *et al.* ont testé une méthode similaire via le réseau GREEN [7] et EE(G)-SPDNet [10] respectivement mais en mettant l'accent sur l'étude des modes de convolution. Nous cherchons à apporter une interprétation neurophysiologique de ce réseau basé sur le SPDNet.

## 2.4 Méthode de comparaison : EEGNet

Nous comparons la méthode proposée à un réseau de neurones convolutionnel de la littérature, spécifiquement développé pour le traitement de signaux EEG en BCI : l'EEGNet [6]. Ce dernier est constitué de trois couches principales :

- une **convolution temporelle** qui à la manière du CNN-SPDNet extrait des bandes de fréquences.
- une **convolution profonde** qui extrait des filtres spatiaux pour chaque bande de fréquences apprises.
- une couche de **convolution séparable**. Celle-ci est la combinaison de deux couches de convolutions : une profonde et une ponctuelle. Elle résume l'information temporelle contenue dans les cartes de caractéristiques et croise les informations qu'elles contiennent.

## 3 Résultats

### 3.1 Jeu de données

Les données EEG utilisées sont issues du jeu de données publiques BCI Competition IV 2a [9]. Les signaux sont enregistrés avec 22 électrodes, échantillonnés à 250 Hz et filtrés entre 0.5 et 100 Hz. Ce jeu de données comprend 2 sessions. 9 sujets réalisent des tâches d'imagerie motrice consistant en 4 classes de mouvements : mains gauche et droite, pieds et langue ; suivant un protocole spécifique [9]. On nomme epoch ou essai un segment déclenché par un stimulus correspondant à une tâche mentale. Il y en a 72 par classe, donnant 288 essais par sujet. Nous évaluons les performances en intra-session pour chaque sujet en divisant les essais en un jeu d'entraînement et un jeu de test selon 5 plis. Le score final est obtenu en moyennant sur les 5 plis. La durée d'un essai est réduite au segment temporel compris entre 0.5 et 2.5 s après le stimulus. Pour notre étude, nous souhaitons mettre en avant la supériorité d'une méthode apprenant les filtres fréquentiels et montrer la limite des méthodes où ceux-ci sont fixés en amont. Nous choisissons donc de comparer les différentes méthodes sur des signaux EEG filtrés entre 1-40 Hz ce qui est supérieur à la plage de fréquence adaptée à l'imagerie motrice (8-32 Hz). Les noyaux de convolutions sont de taille 64 pour l'EEGNet-8,2 et 50 pour le CNN1-SPDNet6. Nous initialisons ceux-ci avec un filtre FIR passe-bande 1-40 Hz. L'initialisation des filtres spatiaux contraints à la variété de Stiefel est elle aléatoire.

### 3.2 Performances en précision

Les performances en termes de précision sont regroupées par sujet dans le tableau 1. En imagerie motrice certains sujets sont moins répondants au protocole et parfois ne parviennent pas à imaginer les mouvements. Ici, les sujets S2, S4, S5 et S6 peuvent ainsi être qualifiés de « mauvais » sujets [11]. D'abord, on constate que l'ajout du CNN en amont du SPDNet permet effectivement d'améliorer la précision (+5.4% en moyenne). Cela prouve l'utilité d'apprendre des bandes de fréquences. Ensuite, par rapport à l'EEGNet, notre méthode se place au-dessus, +2.4%, ce qui montre aussi l'intérêt de cette méthode riemannienne. Le CNN-SPDNet donne presque toujours les meilleurs résultats pour les bons sujets marqués avec « \* », avec un gain moyen de 8.3%.

**Paramètres appris** Le tableau 2 montre l'avantage des réseaux SPD sur l'EEGNet par leur bien plus faible nombre de paramètres appris : on constate une diminution d'un facteur 10. Cela n'est pas négligeable si l'on souhaite embarquer le modèle ou lorsque l'on s'intéresse à la frugalité.

TABLE 2 : Nombre de paramètres appris par les modèles.

Modèle	EEGNet-8,2	SPDNet6	CNN1-SPDNet6
# paramètres	3 444	<b>220</b>	270

### 3.3 Interprétation des paramètres appris

Dans cette section nous présentons les résultats de notre contribution principale : l'interprétation des filtres spatiaux du réseau SPDNet.

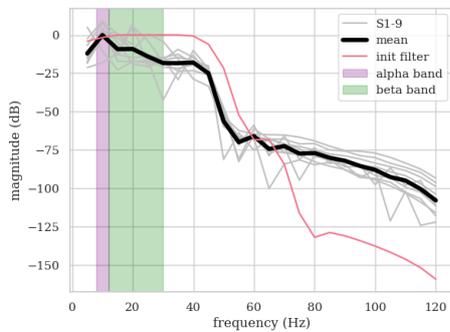


FIGURE 3 : Filtre fréquentiel FIR moyen (noir) obtenu sur les 9 sujets (gris). Les couleurs représentent les bandes de fréquences où l'on observe l'imagerie motrice. Le filtre initial 1-40 Hz (avant entraînement) est représenté en rouge.

L'apprentissage de bandes de fréquences spécifiques se vérifie Fig.3 : en moyenne on passe du filtre rouge au filtre noir bien plus sélectif, notamment dans la bande alpha.

Or on sait que les comportements moteurs, comme l'imagination d'une tâche mentale, entraînent une désynchronisation liée à l'évènement (ERD) dans la bande alpha [8], c'est pourquoi le réseau de neurones extrait principalement ces fréquences. Plus largement, on s'intéresse en IM aux fréquences entre 7-35 Hz, ce que l'on retrouve globalement Fig.3.

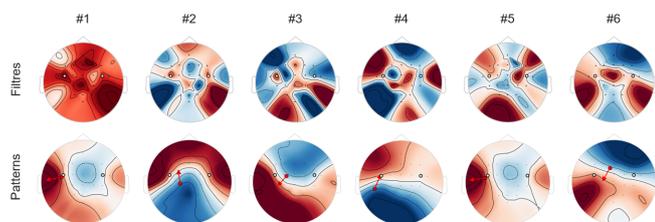


FIGURE 4 : Filtres et patterns spatiaux appris sur le sujet S8. Les électrodes C3 et C4, correspondant aux régions actives lors de l'IM des mains, sont représentées par un rond blanc. Les flèches rouges sont les approximations des patterns par une unique source électrique dipolaire virtuelle. Ils ne correspondent pas aux dipôles cérébraux physiologiques.

Intéressons nous à l'aspect spatial Fig.4. On constate que les filtres appris sur S8, choisi pour exemple, ont été optimisés au cours de la rétropropagation. En effet, les zones du cortex moteur pour le mouvement des membres étudiés ressortent. Par exemple, cela est bien visible au niveau de l'électrode située sur l'aire motrice de la main droite sur le filtre 3 ou des pieds sur le filtre 6. Cependant, ce sont les patterns qui rendent compte d'informations neurophysiologiques [3]. Or ils peuvent être expliqués par des sources virtuelles dipolaires génératrices

proches du cortex moteur : on retrouve des patterns de base de l'activité cérébrale pour cette imagerie motrice. Cette méthode permet d'avoir une idée de ce à quoi ressemble l'activité cérébrale, et cela sans s'appuyer explicitement sur des modèles de sources biophysiques. Remarquons qu'augmenter la profondeur du réseau rend l'interprétation plus complexe pour les couches profondes à cause de la non-linéarité.

## 4 Conclusion

Cet article permet de compléter les connaissances sur le SPDNet en EEG en décrivant comment interpréter les poids appris comme des filtres spatiaux. De plus, est mise en avant l'utilité d'apprendre des bandes de fréquences par CNN via l'implémentation du CNN-SPDNet. On pourrait imaginer utiliser l'information apprise, notamment les filtres fréquentiels, pour d'autres algorithmes (MDRM, CSP par exemple [1]).

## Références

- [1] A. BARACHANT *et al.* : Multiclass brain-computer interface classification by Riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4), 2012.
- [2] I. CARRARA *et al.* : Geometric Neural Network based on Phase Space for BCI decoding. *Journal of Neural Engineering*, octobre 2024.
- [3] S. HAUFE *et al.* : On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, 87:96–110, 2013.
- [4] Z. HUANG et L. VAN GOOL : A Riemannian network for SPD matrix learning. *In Association for the Advancement of Artificial Intelligence (AAAI)*, 2017.
- [5] Z. J. KOLES *et al.* : Spatial patterns underlying population, differences in the background EEG. *Brain Topography*, 2, 1990.
- [6] V. J. LAWHERN *et al.* : EEGNet : a compact convolutional neural network for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 15(5), 2018.
- [7] J. PAILLARD *et al.* : GREEN : A lightweight architecture using learnable wavelets and riemannian geometry for biomarker exploration with EEG signals. *Patterns*, 6(3): 101182, 2025.
- [8] G. PFURTSCHELLER *et al.* : Event-related eeg/meg synchronization and desynchronization : basic principles. *Clinical Neurophysiology*, 110(11):1842–1857, 1999.
- [9] M. TANGERMANN *et al.* : Review of the bci competition iv. *Frontiers in Neuroscience*, pages 55–55, 2012.
- [10] D. WILSON *et al.* : Deep Riemannian networks for end-to-end EEG decoding. *arXiv*, 2024.
- [11] P. ZANINI *et al.* : Transfer learning : A Riemannian geometry framework with applications to brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, 65(5):1107–1116, 2018.