Débruitage de signaux électroencéphalographiques par analyse de corrélation canonique

ESIR 2 spécialité Ingénierie Biomédicale - Séance 2/3

Laboratoire LTSI - UMR INSERM U1099 - Université de Rennes 1

1 L'analyse de corrélation canonique

Soient deux vecteurs aléatoires $\boldsymbol{x}$ et $\boldsymbol{y}$ centrés, i.e. d'espérance nulle. La méthode de CCA consiste à identifier le couple $(\boldsymbol{w}_x,\boldsymbol{w}_y)$ de vecteurs colonne de norme unité (on précisera la norme employée ci-dessous) tel que les projections u et v définies respectivement par les équations suivantes :

$$u = \boldsymbol{w}_x^\mathsf{T}\boldsymbol{x} \tag{1}$$

$$v = \boldsymbol{w}_y^\mathsf{T}\boldsymbol{y} \tag{2}$$

soient le plus corrélées possible. Montrons comment l'identification du couple $(\boldsymbol{w}_x,\boldsymbol{w}_y)$ par maximisation du coefficient de corrélation peut être être réécrite sous la forme d'un simple problème de diagonalisation matricielle. Comme rappelé dans la section "Rappels statistiques" de l'énoncé du TP1, le coefficient de corrélation entre u et v est défini par :

$$\rho\,(u,v) = \frac{\mathsf{E}[uv]}{\sqrt{\mathsf{E}[u^2]}\sqrt{\mathsf{E}[v^2]}} \tag{3}$$

En intégrant (1) et (2) dans cette équation, on obtient alors :

$$\rho\,(u,v) = \frac{\mathsf{E}[\boldsymbol{w}_x^\mathsf{T}\boldsymbol{x}\boldsymbol{w}_y^\mathsf{T}\boldsymbol{y}]}{\mathsf{E}[(\boldsymbol{w}_x^\mathsf{T}\boldsymbol{x})^2]\mathsf{E}[(\boldsymbol{w}_y^\mathsf{T}\boldsymbol{y})^2]} \tag{4}$$

Comme u et v sont des variables aléatoires scalaires, on a alors $uv = uv^\mathsf{T}$ et donc $\boldsymbol{w}_y^\mathsf{T}\boldsymbol{y} = (\boldsymbol{w}_y^\mathsf{T}\boldsymbol{y})^\mathsf{T} = \boldsymbol{y}^\mathsf{T}\boldsymbol{w}_y$. D'où on obtient :

$$\begin{aligned}\mathsf{E}[\boldsymbol{w}_x^\mathsf{T}\boldsymbol{x}\boldsymbol{w}_y^\mathsf{T}\boldsymbol{y}] &= \mathsf{E}[\boldsymbol{w}_x^\mathsf{T}\boldsymbol{x}\boldsymbol{y}^\mathsf{T}\boldsymbol{w}_y] \\ &= \boldsymbol{w}_x^\mathsf{T}\mathsf{E}[\boldsymbol{x}\boldsymbol{y}^\mathsf{T}]\boldsymbol{w}_y \\ &= \boldsymbol{w}_x^\mathsf{T}\boldsymbol{C}_{xy}\boldsymbol{w}_y\end{aligned} \tag{5}$$

où $\boldsymbol{C}_{xy} = \mathsf{E}[\boldsymbol{x}\boldsymbol{y}^\mathsf{T}]$ désigne la matrice de covariance de $\boldsymbol{x}$ et $\boldsymbol{y}$ qui rappelons-le sont supposés centrés. De manière analogue, on obtient :

$$\begin{aligned}\mathsf{E}[(\boldsymbol{w}_x^\mathsf{T}\boldsymbol{x})^2] &= \boldsymbol{w}_x^\mathsf{T}\boldsymbol{C}_{xx}\boldsymbol{w}_x, \\ \mathsf{E}[\left(\boldsymbol{w}_y^\mathsf{T}\boldsymbol{y}\right)^2] &= \boldsymbol{w}_y^\mathsf{T}\boldsymbol{C}_{yy}\boldsymbol{w}_y.\end{aligned}$$

Par conséquent l'équation (4) se réécrit de la manière suivante :

$$\rho(u,v) = \frac{\mathsf{E}[uv]}{\sqrt{\mathsf{E}[u^2]}\sqrt{\mathsf{E}[v^2]}} = \frac{\boldsymbol{w}_x^\mathsf{T}C_{xy}\boldsymbol{w}_y}{\sqrt{\boldsymbol{w}_x^\mathsf{T}C_{xx}\boldsymbol{w}_x\boldsymbol{w}_y^\mathsf{T}C_{yy}\boldsymbol{w}_y}} \tag{6}$$

La CCA vise donc à identifier le couple $(\boldsymbol{w}_x,\boldsymbol{w}_y)$ par maximisation de la fonction ψ suivante et ce par rapport à $\boldsymbol{v}_x$ et $\boldsymbol{v}_y$:

$$\psi(\boldsymbol{v}_x,\boldsymbol{v}_y) = \frac{\boldsymbol{v}_x^\mathsf{T}C_{xy}\boldsymbol{v}_y}{\sqrt{\boldsymbol{v}_x^\mathsf{T}C_{xx}\boldsymbol{v}_x\boldsymbol{v}_y^\mathsf{T}C_{yy}\boldsymbol{v}_y}} \tag{7}$$

Considérons comme normes respectives de $\boldsymbol{v}_x$ et $\boldsymbol{v}_y$ les valeurs $\sqrt{\boldsymbol{v}_x^\intercal C_{xx}\boldsymbol{v}_x}$ et $\sqrt{\boldsymbol{v}_y^\intercal C_{yy}\boldsymbol{v}_y}$. Sous l'hypothèse de norme unité des vecteurs $\boldsymbol{v}_x$ et $\boldsymbol{v}_y$, la maximisation de ψ équivaut à maximiser le numérateur de ψ sous la contrainte $\boldsymbol{v}_x^\intercal C_{xx}\boldsymbol{v}_x = \boldsymbol{v}_y^\intercal C_{yy}\boldsymbol{v}_y = 1$. Une telle maximisation sous contrainte peut être effectuée en maximisant le Lagrangien suivant en fonction de $\boldsymbol{v}_x$ et $\boldsymbol{v}_y$:

$$\mathcal{L}(\boldsymbol{v}_x, \boldsymbol{v}_y) = \boldsymbol{v}_x^\intercal C_{xy}\boldsymbol{v}_y - \frac{\lambda_1}{2}(\boldsymbol{v}_x^\intercal C_{xx}\boldsymbol{v}_x - 1) - \frac{\lambda_2}{2}(\boldsymbol{v}_y^\intercal C_{yy}\boldsymbol{v}_y - 1) \tag{8}$$

Nous nous ramenons ainsi à un problème de maximisation sans contrainte. Cherchons alors les points stationnaires du Lagrangien. Il nous faut pour cela calculer les dérivées du Lagrangien par rapport à $\boldsymbol{v}_x$ et $\boldsymbol{v}_y$. Or on a :

$$\begin{aligned}
\frac{\partial \boldsymbol{v}_x^\intercal C_{xy}\boldsymbol{v}_y}{\partial \boldsymbol{v}_x} &= C_{xy}\boldsymbol{v}_y \\
\frac{\partial \boldsymbol{v}_x^\intercal C_{xy}\boldsymbol{v}_y}{\partial \boldsymbol{v}_y} &= C_{xy}^\intercal\boldsymbol{v}_x \\
\frac{\partial \boldsymbol{v}_x^\intercal C_{xx}\boldsymbol{v}_x}{\partial \boldsymbol{v}_x} &= C_{xx}^\intercal\boldsymbol{v}_x + C_{xx}\boldsymbol{v}_x \\
&= 2C_{xx}\boldsymbol{v}_x \\
\frac{\partial \boldsymbol{v}_y^\intercal C_{yy}\boldsymbol{v}_y}{\partial \boldsymbol{v}_y} &= 2C_{yy}\boldsymbol{v}_y
\end{aligned}$$

D'où :

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \boldsymbol{v}_x} &= C_{xy}\boldsymbol{v}_y - \lambda_1 C_{xx}\boldsymbol{v}_x \\
\frac{\partial \mathcal{L}}{\partial \boldsymbol{v}_y} &= C_{xy}^\intercal\boldsymbol{v}_x - \lambda_2 C_{yy}\boldsymbol{v}_y
\end{aligned} \tag{9}$$

En cherchant les valeurs respectives $\boldsymbol{w}_x$ et $\boldsymbol{w}_y$ de $\boldsymbol{v}_x$ et $\boldsymbol{v}_y$ pour lesquelles les dérivées du Lagrangien s'annulent (i.e. les points stationnaires du Lagrangien), on obtient :

$$C_{xy}\boldsymbol{w}_y = \lambda_1 C_{xx}\boldsymbol{w}_x \tag{10}$$

$$C_{xy}^\intercal\boldsymbol{w}_x = \lambda_2 C_{yy}\boldsymbol{w}_y \tag{11}$$

Par conséquent, si $\boldsymbol{C}_{xx}$ et $\boldsymbol{C}_{yy}$ sont inversibles, nous obtenons :

$$\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{xy}\boldsymbol{w}_y = \lambda_1\boldsymbol{w}_x \tag{12}$$

$$\boldsymbol{C}_{yy}^{-1}\boldsymbol{C}_{xy}^\intercal\boldsymbol{w}_x = \lambda_2\boldsymbol{w}_y \tag{13}$$

En pré-multipliant respectivement les équations (10) et (11) par $\boldsymbol{w}_x^\intercal$ et $\boldsymbol{w}_y^\intercal$, et en utilisant le fait que $\boldsymbol{w}_x^\intercal\boldsymbol{C}_{xx}\boldsymbol{w}_x = 1$ et $\boldsymbol{w}_y^\intercal\boldsymbol{C}_{yy}\boldsymbol{w}_y = 1$ on obtient :

$$\begin{aligned}
\boldsymbol{w}_x^\intercal\boldsymbol{C}_{xy}\boldsymbol{w}_y &= \lambda_1\boldsymbol{w}_x^\intercal\boldsymbol{C}_{xx}\boldsymbol{w}_x = \lambda_1 \\
\boldsymbol{w}_y^\intercal\boldsymbol{C}_{xy}^\intercal\boldsymbol{w}_x &= \lambda_2\boldsymbol{w}_y^\intercal\boldsymbol{C}_{yy}\boldsymbol{w}_y = \lambda_2
\end{aligned} \tag{14}$$

Enfin, comme $\boldsymbol{w}_x^\intercal\boldsymbol{C}_{xy}\boldsymbol{w}_y = \left(\boldsymbol{w}_y^\intercal\boldsymbol{C}_{xy}^\intercal\boldsymbol{w}_x\right)^\intercal = \boldsymbol{w}_y^\intercal\boldsymbol{C}_{xy}^\intercal\boldsymbol{w}_x$, on déduit des deux équations précédentes que $\lambda_1 = \lambda_2$. Posons $\lambda = \lambda_1 = \lambda_2$. Si on remplace $\boldsymbol{w}_x$ de l'équation (13) par son expression issue de l'équation (12) et si on remplace $\boldsymbol{w}_y$ de l'équation (12) par son expression issue de l'équation (13) on obtient alors :

$$\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{xy}\boldsymbol{C}_{yy}^{-1}\boldsymbol{C}_{xy}^\intercal\boldsymbol{w}_x = \lambda^2\boldsymbol{w}_x, \tag{15}$$

$$\boldsymbol{C}_{yy}^{-1}\boldsymbol{C}_{xy}^\intercal\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{xy}\boldsymbol{w}_y = \lambda^2\boldsymbol{w}_y. \tag{16}$$

Le vecteur $\boldsymbol{w}_x$ apparaît alors comme le vecteur propre de $\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{xy}\boldsymbol{C}_{yy}^{-1}\boldsymbol{C}_{xy}^\intercal$ associé à la valeur propre λ^2. Le vecteur $\boldsymbol{w}_y$ est quant à lui le vecteur propre de $\boldsymbol{C}_{yy}^{-1}\boldsymbol{C}_{xy}^\intercal\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{xy}$ associé à la valeur propre λ^2. Or il apparaît d'après les équations (14) et (6) que, sous la contrainte de norme unité des vecteurs $\boldsymbol{w}_x$ et $\boldsymbol{w}_y$, la valeur propre λ^2 des matrices $\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{xy}\boldsymbol{C}_{yy}^{-1}\boldsymbol{C}_{xy}^{\mathsf{T}}$ et $\boldsymbol{C}_{yy}^{-1}\boldsymbol{C}_{xy}^{\mathsf{T}}\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{xy}$ correspond à $\psi(\boldsymbol{w}_x, \boldsymbol{w}_y)$. De ce fait, afin de maximiser la fonction ψ il faut calculer $\boldsymbol{w}_x$ et $\boldsymbol{w}_y$ comme étant les vecteurs propres respectives associées à la valeur propre λ^2 **maximale** des matrices $\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{xy}\boldsymbol{C}_{yy}^{-1}\boldsymbol{C}_{xy}^{\mathsf{T}}$ et $\boldsymbol{C}_{yy}^{-1}\boldsymbol{C}_{xy}^{\mathsf{T}}\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{xy}$.

Revenons à notre problème de débruitage de l'activité EEG décrit dans l'énoncé du TP1. Il suffit d'utiliser la méthode CCA en prenant $\boldsymbol{x} = \boldsymbol{x}[m]$ et $\boldsymbol{y} = \boldsymbol{x}[m-1]$. Le vecteur $\bar{\boldsymbol{w}}_e$ est alors identifié de manière analogue à celle décrite ci-dessus permettant d'identifier le vecteur $\boldsymbol{w}_x$. La valeur propre λ^2 maximale correspondra alors à la fonction d'autocorrélation $\rho[1]$ du processus aléatoire $\{\bar{\boldsymbol{s}}_e[m]\}$. Nous supposons ici par soucis de simplicité que les processus aléatoires impliqués sont tous stationnaires au sens large à l'ordre deux. Si cette hypothèse peut paraître trop forte dans le contexte applicatif décrit, il est à noter que seule l'hypothèse d'ergodicité est réellement nécessaire à l'application de la méthode CCA dans le contexte présent. Il suffit alors juste de réécrire la méthode de CCA en remplaçant les statistiques impliquées par une moyenne temporelle de ces dernières, qui pourront être estimées sous l'hypothèse d'ergodicité.

2 Génération des données EEG intercritiques bruitées et estimation des statistiques

- Charger les jeux de données "data.mat" à l'adresse "perso.univ-rennes1.fr/laurent.albera/" dans l'onglet "Teaching" à la rubrique "ESIR 2 module de TS3" (indication : utiliser la fonction MATLAB "load").
- Construire la matrice "XX" de données EEG intercritiques bruitées, et ce à partir des données chargées en mémoire, de manière à que X soit une réalisation du processus vectoriel aléatoire $\{\boldsymbol{x}[m]\}$ décrit par l'équation (1) de l'énoncé du TP1. On pourra utiliser la fonction MATLAB "randn" afin de générer une réalisation d'un processus vectoriel aléatoire gaussien centré réduit.
- Construire la matrice "X" comme étant la matrice "XX" privée de sa première colonne. Construire ensuite la matrice Y comme étant la matrice "XX" privée de sa dernière colonne.
- Estimer la matrice d'autocovariance $\boldsymbol{C}_x[1]$ du processus $\{\boldsymbol{x}[m]\}$ à partir des matrices "X" et "Y" construites précédemment. On la notera "Cx1" sous MATLAB. Commencer par centrer les lignes des matrices "X" et "Y" au moyen de la commande MATLAB "mean". Puis estimer $\boldsymbol{C}_x[1]$ à l'aide d'un simple produit matriciel des matrices centrées de "X" et "Y" en normalisant le produit par le nombre de colonnes de "X".
- Estimer la matrice d'autocovariance $\boldsymbol{C}_x[0]$ du processus $\{\boldsymbol{x}[m]\}$ à partir de la matrice centrée de "X". On la notera "Cx0" sous MATLAB.
- Afin d'appliquer la méthode CCA, il faut calculer certaines matrices d'autocovariance telles que $\boldsymbol{C}_x[1]$ et $\boldsymbol{C}_x[0]$ qui joueront respectivement le rôle de $\boldsymbol{C}_{xy}$ et $\boldsymbol{C}_{xx}$ (voir section 1). Quelle hypothèse faite sur le processus $\{\boldsymbol{x}[m]\}$ nous permet ici de considérer $\boldsymbol{C}_{xx} = \boldsymbol{C}_{yy}$ et ne requiert par conséquent le calcul d'aucune autre matrice d'autocovariance?