



HAL
open science

Inférence et décomposition modale de réseaux dynamiques en neurosciences

Gaëtan Frusque

► **To cite this version:**

Gaëtan Frusque. Inférence et décomposition modale de réseaux dynamiques en neurosciences. Traitement du signal et de l'image [eess.SP]. Université de Lyon, 2020. Français. NNT : 2020LYSEN080 . tel-03125330v2

HAL Id: tel-03125330

<https://theses.hal.science/tel-03125330v2>

Submitted on 3 Feb 2021

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.



Numéro National de Thèse : 2020LYSEN080

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée par
l'École Normale Supérieure de Lyon

École Doctorale N° 512
École Doctorale en Informatique et Mathématiques de Lyon

Spécialité de doctorat : Informatique
Discipline : Traitement du signal, apprentissage.

Soutenue publiquement le 07/12/2020, par :
Gaëtan FRUSQUE

**Inférence et Décomposition Modale de
Réseaux Dynamiques en Neurosciences**

Devant le jury composé de :

Achard, Sophie	DR-CNRS	INRIA Grenoble - Rhône-Aples	Rapporteuse
Brie, David	PU	Université de Lorraine	Rapporteur
Tyvaert, Louise	PU-PH	Université de Lorraine	Examinatrice
Dobigeon, Nicolas	PU	INP-ENSEIGHT	Examineur
Gonçalves, Paulo	DR-INRIA	ENS de Lyon	Directeur de thèse
Borgnat, Pierre	DR-CNRS	ENS de Lyon	Co-encadrant
Jung, Julien	PH	Hospices Civils de Lyon	Co-encadrant

À mes parents :
Marie & Daniel

Remerciements

Les travaux de cette thèse sont issus de trois années de recherches au sein de l'équipe Dante du laboratoire de l'informatique du parallélisme de l'ENS de Lyon. Je garderais toujours un excellent souvenir de ces années. C'est pour cela que je tiens à remercier chaleureusement mes encadrants qui m'ont fait confiance :

À Paulo Gonçalves qui m'a permis de mieux présenter et défendre mes recherches. Ces enseignements ont grandement contribué à ma formation doctorale.

À Pierre Borgnat qui a des qualités humaines indéniables, en plus de ton expertise j'ai beaucoup apprécié tes encouragements.

À Julien Jung pour sa disponibilité, sa bonne humeur et son expertise en neurosciences. Ce fut un grand plaisir de travailler avec toi.

Grâce à vous trois, j'ai appris énormément.

Je souhaite aussi remercier tous les membres des équipes DYCOG et CMO du CRNL pour être à l'initiative de ce projet, pour leur générosité et pour m'avoir fourni plusieurs jeux de données.

Je tiens à remercier Sophie Achard et David Brie pour avoir accepté d'évaluer mes travaux de recherche, ainsi que Louise Tyvaert et Nicolas Dobigeons de faire partie de mon jury. Je suis très reconnaissant de l'intérêt que vous avez apporté à ma thèse.

Je veux remercier l'ensemble de mes collègues de l'équipe Dante, c'était un plaisir de travailler et de discuter avec vous. Je tenais à mentionner Esteban Baupista qui a été pour moi un mentor et un ami durant cette thèse.

J'aimerais tous particulièrement remercier mes parents, ma tante Marcie ainsi que tout le reste de ma famille pour leur soutien implacable. Pour ce qui est de ma famille de coeur : la Pifoufe et mes amis proches, vous serez toujours une source d'inspiration et un soutien émotionnel essentiel pour moi, merci.

Résumé

Les graphes dynamiques permettent de comprendre l'évolution de systèmes complexes qui évoluent dans le temps. Ceux-ci peuvent être vus comme une succession de graphes complets partageant les mêmes nœuds, mais dont les poids associés à chaque lien changent. Ces types de graphes ont récemment fait l'objet d'une attention considérable. En effet, les applications actuelles génèrent de plus en plus de données structurées, qu'il est nécessaire d'analyser avec des outils méthodologiques sophistiqués. Cependant, il n'existe pas de consensus sur les manières de les inférer et de les étudier. Dans cette thèse, on propose des méthodes d'analyse de graphes dynamiques spécifiques. Elles peuvent avoir des applications en neurosciences ou dans l'étude des réseaux sociaux comme Twitter et Facebook par exemple. L'enjeu applicatif de cette thèse est l'épilepsie, l'une des maladies neurologiques les plus répandues dans le monde affectant environ 1% de la population.

La première partie concerne l'inférence de graphe dynamique à partir de signaux neurophysiologiques. Cette inférence est généralement réalisée à l'aide de mesures de connectivité fonctionnelle permettant d'évaluer la similarité entre deux signaux. Ces mesures sont appliquées entre chaque paire de signaux et dans le temps, le réseau dynamique obtenu peut être vu comme une représentation imparfaite de l'état cortical du cerveau durant une crise d'épilepsie. De nombreuses mesures de connectivité fonctionnelle existent et leurs utilisations dépendent de l'application. La comparaison de ces mesures est donc d'un grand intérêt pour comprendre les caractéristiques des graphes obtenus. Dans un premier temps, on fait l'hypothèse que les signaux suivent une distribution gaussienne et qu'ils sont stationnaires. On compare alors des mesures de connectivité fonctionnelle impliquant la phase et l'amplitude instantanée des signaux. On s'intéresse en particulier à une mesure nommée Phase-Locking-Value (PLV) qui quantifie la synchronie des phases entre deux signaux. On propose ensuite, afin d'inférer des graphes dynamiques robustes et interprétables, deux nouvelles mesures de PLV conditionnées et régularisées. La première suppose que les signaux sont gaussiens et stationnaires et la seconde basée sur la mesure usuelle de PLV non paramétrique. On montre et on compare les performances de ces deux mesures sur des signaux tests générés par des oscillateurs de Roessler avant d'appliquer cette nouvelle approche sur plusieurs crises d'épilepsies provenant de 5 patients différents.

La seconde partie présente des méthodes de décomposition de graphes dynamiques. L'objectif est de proposer une méthode semi-automatique afin de caractériser les informations les plus importantes du réseau pathologique de plusieurs crises d'un même patient. Dans un premier temps, on considère des crises qui ont des durées et des évolutions temporelles similaires. Dans ce cas les données peuvent être commodément représentées sous forme de tenseur. Une décomposition tensorielle spécifique est alors appliquée pour extraire les sous-graphes pathologiques communs à l'ensemble des crises et caractéristiques de la propagation de l'épilepsie dans le cerveau. Des contraintes particulières sont imposées sur les composantes de cette décomposition afin d'obtenir des sous-graphes parcimonieux s'activant durant une période particulière de la crise d'épilepsie. Dans un second temps, on considère des crises qui ont des durées hétérogènes. Dans ce cas la décomposition tensorielle utilisée dans la situation précédente n'est plus applicable. Plusieurs alternatives à cette décomposition sont donc proposées. Ce sont des méthodes qui en plus d'extraire les sous-graphes caractéristiques communs à toutes les crises, permettent d'observer leurs profils d'activation temporelle spécifiques à chaque crise. Afin d'évaluer les différentes stratégies pour décomposer un graphe dynamique, on considère un modèle de crise d'épilepsie simplifiée. Enfin, chaque méthode est appliquée sur deux patients dont l'un avec des crises de durées homogènes et l'autre avec des crises de durées variables. L'une des stratégies est retenue : il s'agit une forme contrainte de décomposition jointe en matrice non négative.

Enfin, on utilise la méthode retenue pour une application clinique. On considère les enregistre-

ments électroencéphalographiques (EEG) intracrâniens de 27 crises d'épilepsie provenant de 9 patients différents. Les patients souffrent tous d'épilepsies focales résistantes aux médicaments. L'objectif est de proposer une visualisation simplifiée de l'évolution du réseau cortical lors d'une crise d'épilepsie. Les décompositions obtenues sont comparées à l'interprétation visuelle du clinicien. Dans l'ensemble, on constate que les sous-graphes extraits correspondent aux régions du cerveau impliquées dans la crise d'épilepsie. De plus l'évolution de l'activation de ces sous-graphes est cohérente avec l'interprétation visuelle. On pense que la méthode proposée peut compléter l'analyse visuelle des signaux SEEG enregistrés durant des crises en mettant en évidence les ensembles les plus importants des réseaux épileptiques avec leur dynamique d'activation.

Abstract

Dynamic graphs make it possible to understand the evolution of complex systems evolving over time. A dynamical graph can be seen as a succession of complete graphs sharing the same nodes, but with the weights associated with each link changing over time. This type of graph has recently received considerable attention. Indeed, current applications generate more and more structured data, it is necessary to analyze them with sophisticated methodological tools. However, there is no consensus on how to infer and study them. In this thesis, we propose specific methods for dynamical graph analysis. They can have applications in neuroscience or in the study of social networks such as Twitter and Facebook for example. The issue of this thesis is epilepsy, one of the most common neurological diseases in the world affecting around 1% of the population.

The first part concerns the inference of dynamical graph from neurophysiological signals. To assess the similarity between each pairs of signals, in order to make the graph, we use measures of functional connectivity. These measurements are applied between each pair of signals and over time, the resulting dynamic network can be seen as an imperfect representation of the cortical state of the brain during an epileptic seizure. Many functional connectivity measures exist, depending on the application, some are preferred. The comparison of these measurements is therefore of great interest to understand the characteristics of the resulting graphs. First, we assume that the signals are stationary and follow a Gaussian distribution. We then compare functional connectivity measurements involving the instantaneous phase and amplitude of the signals. We are particularly interested in a measure called Phase-Locking-Value (PLV) which quantifies the phase synchrony between two signals. We then propose, in order to infer robust and interpretable dynamic graphs, two new indexes that are conditioned and regularized PLV. The first assumes that the signals are stationary and Gaussian whereas the second is based on the usual nonparametric PLV. We show and compare the performance of these two indexes on signals generated by a coupled Roessler attractor before applying this new approach to several epileptic seizures from 5 different patients.

The second part concerns tools for dynamical graphs decompositions. The objective is to propose a semi-automatic method in order to characterize the most important patterns in the pathological network from several seizures of the same patient. First, we consider seizures that have similar durations and temporal evolutions. In this case the data can be conveniently represented as a tensor. A specific tensor decomposition is then applied to extract the pathological subgraphs, that are common to all the seizures and characterize the propagation of epilepsy in the brain. Secondly, we consider seizures that have heterogeneous durations. In this case the tensor decomposition previously used is no longer applicable. Several alternatives to this decomposition are therefore proposed. These are methods which, in addition to extracting the characteristic subgraphs common to all the seizures, make it possible to observe their temporal activation profiles specific to each seizures. In order to evaluate the different strategies for decomposing a dynamical graph, we consider a simplified model of epileptic seizure. Finally, each method is applied to two patients, one with seizures of homogeneous durations and the other of variable durations. One of the strategies is retained : it is a constrained form of joined non-negative matrix decomposition.

Finally, the selected method is used for a clinical application. We consider iEEG recording of 27 seizures from 9 different patients. The patients all suffer from drug-resistant focal epilepsies. The objective is to provide a simplified visualization of the evolution of the cortical network during an epileptic seizure. The obtained decompositions are compared to the visual interpretation of the clinician. As a whole, we found that activated subgraphs corresponded to brain regions involved during the course of the seizures and their time course were highly consistent with classical visual interpretation.

We believe the proposed method can complement the visual analysis of SEEG signals recorded during seizures by highlighting and characterizing the most significant parts of epileptic networks with their activation dynamics.

Table des matières

Remerciements	I
Résumé (Français/Anglais)	III
Table des figures	XI
Liste des tableaux	XV
Glossaire	XVII
Introduction	1
1 Contexte : Étude des réseaux dynamiques en neurosciences	7
1.1 Mesurer l'activité cérébrale lors d'une crise d'épilepsie	7
1.1.1 Épilepsie	7
1.1.2 Origine des signaux du cerveau	7
1.1.3 Enregistrement électroencéphalographie stéréotaxique (SEEG)	8
1.2 Les connectivités fonctionnelles	10
1.2.1 Mesurer la similarité entre deux signaux	10
1.2.2 Expression mathématique de mesures de CFs	11
1.2.3 Principales mesures de connectivité fonctionnelle	14
1.3 Décomposition et analyse des réseaux de connectivité fonctionnelle	18
1.3.1 Analyse statique des CFs	18
1.3.2 Le problème de la multimodalité	18
1.3.3 Analyse dynamique et multimodale des CFs	19
1.4 Conclusion	20
Appendices	21
1.A Prétraitement	21
1.A.1 Montage bipolaire :	21
1.A.2 Élimination d'électrodes :	21
1.A.3 Problème des fréquences à 50 Hz :	22
1.A.4 Encadrement de la crise d'épilepsie :	22
1.A.5 Filtrage	23
2 Connectivités fonctionnelles sous hypothèse de données gaussiennes	25
2.1 Formules de moments d'ordres supérieurs des relations phase-amplitude dans le cas gaussien	25
2.1.1 Fonctions spéciales	25
2.1.2 Signal analytique bivarié gaussien	27
2.1.3 Formule des moments phases-amplitudes	29
2.2 Démonstration de la formule des moments phases-amplitudes	32
2.2.1 Changement de variable	32
2.2.2 Loi marginale conditionnée au produit des amplitudes	34
2.2.3 La loi de Von-Mises hyperbolique	35

2.2.4	Réduction au calcul d'une intégrale à une seule variable	37
2.3	Extension aux signaux multivariés gaussiens	37
2.3.1	La corrélation partielle	37
2.3.2	Conditionnement de la formule des moments phases-amplitudes	40
2.4	Conclusion	41
Appendices		43
2.A	Démonstrations	43
2.A.1	Démonstration de l'Équation (2.58)	43
2.A.2	Démonstration de l'Équation (2.60)	43
3	Estimation de la Phase-Locking-Value partielle et régularisée	45
3.1	Phase-Locking-Value conditionnelle	46
3.1.1	Le lasso graphique	46
3.1.2	Phase-Locking-Value partielle paramétrique	47
3.1.3	Phase-Locking-Value partielle non paramétrique	49
3.1.4	Principale différence entre les pPLV paramétriques et non paramétriques	51
3.2	Application aux mesures de connectivités fonctionnelles	51
3.2.1	Régularisation pertinente dans le calcul dynamique de CFs	51
3.2.2	Résumé des différentes mesures de CFs introduites	52
3.2.3	Implémentation	52
3.3	Comparaison sur un modèle des estimateurs de pPLV régularisés	55
3.3.1	Modèle de signaux avec couplage de phase	55
3.3.2	Intérêt des régularisations proposées	56
3.3.3	Application des méthodes sur le modèle	58
3.4	Application sur des données sEEG	68
3.5	Conclusion	70
Appendices		72
3.A	Détails sur l'implémentation de la version dynamique du lasso Graphique	72
3.A.1	ADMM	72
3.A.2	Etape (I), Mise à jour de la matrice Θ	73
3.A.3	Etape (II).(a), Mise à jour de la matrice \mathbf{B}	74
3.A.4	Etape (II).(b), Mise à jour des matrices \mathbf{C} et \mathbf{D}	76
3.A.5	Critère d'arrêt	77
3.A.6	Mise à jour du paramètre de l'ADMM	78
4	Contexte : Décomposition matricielle et tensorielle pour l'extraction d'informations structurées	79
4.1	Apprentissage par décomposition matricielle	79
4.1.1	La décomposition matricielle	79
4.1.2	Motivations pour le choix des régularisations et contraintes	80
4.1.3	Exemples usuels de décomposition matricielle	82
4.2	Le concept de tenseur	86
4.2.1	Le tenseur	86
4.2.2	Matricer	86
4.2.3	Les produits tensoriels	87
4.2.4	Vectoriser	89
4.3	Les décompositions tensorielles usuelles	90
4.3.1	La décomposition Canonique Polyadique (CP)	90
4.3.2	La décomposition en valeurs singulières d'ordre supérieure (HOSVD)	91
4.3.3	Parafac2	92
4.4	Conclusion	94

5	Partitionnement de réseaux dynamiques via une décomposition tensorielle parcimonieuse	95
5.1	Partitionner un réseau dynamique	96
5.1.1	Le réseau dynamique de CFs	96
5.1.2	Partitionner un tenseur avec k-means	96
5.1.3	La réduction de dimension	98
5.1.4	Décomposition tensorielle proposée	98
5.2	Une modification de HOOI pour résoudre le problème de la HOSMR sans contrainte	99
5.2.1	Lien entre la HOSMR et la HOSVD	100
5.2.2	Modification de l'algorithme HOOI	100
5.3	Implémentation de la HOSMR	103
5.3.1	Implémentation d'une SVD parcimonieuse	103
5.3.2	Algorithme de la HOSMR	107
5.4	Évaluations numériques sur un modèle	109
5.4.1	Un modèle de graphe dynamique pour les CFs	109
5.4.2	Description de l'expérience	110
5.4.3	Résultats et discussions	111
5.5	Application sur des données réelles	113
5.5.1	Les données sEEG	113
5.5.2	Application	113
5.6	Conclusion	113
	Appendices	115
5.A	Comparaison de performances entre les algorithmes HOOI et mHOOI	115
5.B	Projection efficace sur la contrainte elastic-net	117
5.C	Le score ARI	117
6	Décomposition simultanée de réseaux dynamiques multi-crisés avec des durées hétérogènes	119
6.1	Extensions de la HOsSVD appliquées aux matrices de CFs avec des durées hétérogènes	119
6.1.1	Modèle Parafac2 sous contraintes	119
6.1.2	Décomposition jointe en matrices non-négatives	121
6.1.3	Décomposition multivues en matrices non-négatives	122
6.2	Comparaison de différentes méthodes sur un modèle	125
6.2.1	Un modèle de graphe multiplex avec durée hétérogène pour les FCs	125
6.2.2	Description de l'expérience	126
6.2.3	Résultats et discussions	128
6.3	Application sur des données sEEG réelles.	130
6.3.1	Patient avec des crises de durées homogènes	130
6.3.2	Patient avec des crises hétérogènes	134
6.4	Conclusion	138
	Appendices	140
6.A	Implémentation des algorithmes présentés dans ce chapitre	140
6.A.1	Implémentation de l'algorithme snnP2	140
6.A.2	Implémentation de l'algorithme joinNMF	141
6.A.3	Implémentation de l'algorithme graphNMF (et multNMF)	145
6.B	Lien entre la régularisation multivues et l'analyse canonique des corrélations	147
7	Extraction semi-automatique de réseaux dynamiques décrivant les crises d'épilepsie d'un patient	149
7.1	Méthode d'extraction de réseaux caractérisant une crise d'épilepsie	149
7.1.1	Description de la méthode	149
7.1.2	La méthode joinNMF pour extraire les sous graphes fonctionnels	150

7.1.3	Détails sur l'utilisation de la méthode joinNMF	150
7.2	Application de la procédure BTND sur 9 patients	151
7.2.1	Détails sur les données sEEG utilisées	151
7.2.2	Comparaison du réseau dynamique estimé par analyse visuelle conventionnelle et par la méthode BTND	154
7.2.3	Résultats sur 9 patients	155
7.3	Conclusion	156

Table des figures

1	Graphe dynamique inféré à partir de signaux temporels, indiquant l'évolution des similarités entre les signaux dans le temps. Chaque graphe est calculé sur une portion du signal multivarié. Le poids des liens provient d'une mesure de similarité calculée entre chaque paire de signaux.	2
1.1	(a) - Schéma d'un neurone, (b) - Le neurone, avec son axone et ses dendrites, modélisé comme un dipôle	8
1.2	(a) exemple d'implémentation des tiges stéréotaxiques dans le cerveau - (b) tiges stéréotaxiques - (c) schéma du référentiel de Tailarach ¹ - (d) Électrodes stéréotaxiques visualisées selon le référentiel de Tailarach. Une coupe du cerveau selon les trois plans principaux (axial (x-y), sagittal (x-z), coronal (y-z)) est représentée - (e) Enregistrement d'une crise d'épilepsie par SEEG, on représente en couleur la succession de plusieurs étapes (le début de crise correspond à l'étape rouge, l'étape bleue coïncide avec la propagation de la crise dans les différentes zones mesurées, enfin l'étape grise montre la fin de la crise.)	9
1.3	Deux signaux s_1 et s_2 sont représentés sur le tracé du bas, et les mesures de CF correspondantes (notées $CF(s_1, s_2)$) en fonction du temps sont affichées sur le tracé du haut. Les mesures sont faites sur T échantillons temporels équidistants en utilisant des fenêtres rectangulaires de largeur $2T_f$	11
1.4	(a) - 4 signaux suivant un modèle simple de corrélation, (b) & (c) - structure de graphe pour deux segments centrés autour des échantillons n_t et $n_{t'}$ (la largeur des liens est proportionnelle à la mesure de CF).	12
1.5	(à gauche) Graphe dynamique de CFs caractérisé par une succession de matrices d'adjacence à trois instants, (à droite) matrice de CF \mathbf{X} contenant les éléments supérieurs diagonaux des matrices d'adjacence	13
1.6	Mesure dynamique des CFs pour chaque paire de signaux de l'exemple fourni dans la Figure 1.4, les zones foncées correspondent à des mesures de connectivités élevées . . .	13
1.7	(a) signal $s(n)$, (b) amplitude instantanée de $s(n)$ en rouge, (c) phase instantanée de $s(n)$ en rouge	15
1.8	Phase instantanée de deux signaux SEEG réels, on marque d'un trait rouge tous les instants où le changement de signe de la phase ϕ_2 implique un changement de signe de la phase ϕ_1 dans les 0.05 secondes qui suivent (la durée du signal représenté est de 11 s).	16
1.9	(tracé du bas) Les 2 signaux s_1 et s_2 provenant de l'enregistrement réel SEEG, (tracé du haut) mesures de CF en utilisant la PLV (en rouge) et la corrélation de Pearson (en bleu).	17
1.A.1	Schéma d'un montage bipolaire pour deux tiges, les signaux formant notre jeu de données sont notés $s_i(n)$ et le potentiel des électrodes $e_j(n)$	21
1.A.2	(a) Score pour 50 signaux d'un enregistrement SEEG réels - (b) Forme des signaux 44, 45 et 46 comparée à celle des signaux exploitables.	22
1.A.3	(a) Somme des densités spectrales de puissance de l'ensemble des signaux d'un enregistrement SEEG - (b) Application d'un filtre notch pour éliminer la fréquence à 50 Hz.	22
2.1	(a) Fonction de Bessel modifiée de première espèce, $I_\kappa(x)$, pour $\kappa \in \{1, 2, 3, 4\}$, (b) Fonction de Bessel modifiée de deuxième espèce, $K_\kappa(x)$, pour $\kappa \in \{1, 2, 3, 4\}$	26
2.2	Intégrale elliptique de première et seconde espèce, respectivement notées $\mathcal{K}(x)$ et $\mathcal{E}(x)$	27

2.3	Représentation de l'espérance obtenue pour les sept premières situations du Tableau (2.1) quand $k_{ii} = k_{ij} = 1$ et ρ varie entre 0 et 1. La figure (a) représente les espérances 1 et 2, la figure (b) les espérances 3 et 4, la figure (c) les espérances 5 et 6, la figure (d) l'espérance 7.	31
2.4	Le rectangle rouge représente l'ensemble des valeurs que peut prendre la variable ϕ_c en fonction de φ	34
2.5	(a) Distribution de Von Mises, pour différentes valeurs de β , $\mu = 0$ et $v = 1$. (b) Distribution de Von Mises hyperbolique, pour différentes valeurs de α , $\bar{\mu} = 0$ et $v = 1$	37
2.6	Graphe associé au premier processus	39
2.7	Graphe associé au second processus	40
3.1	(a) Les trois états de couplage en fonction de t . (b) Matrice de CFs idéale (la couleur noire indique la présence d'une synchronie entre les signaux i et j à l'instant t).	56
3.2	Exemples de matrice de CFs obtenues par les méthodes (a) $pPLV^{nonpar}(0, 0)$ - (b) $pPLV^{param}(0, 0)$ - (c) $pPLV^{nonpar}(\lambda, 0)$ - (d) $pPLV^{param}(\lambda, 0)$ - (e) $pPLV^{nonpar}(\lambda, \gamma)$ - (f) $pPLV^{param}(\lambda, \gamma)$	58
3.3	(a), (b), Exemples de signaux provenant du modèle de Roessler pour $t = 1$ et $i = 2$ et 3. - Matrices de CFs obtenues respectivement à partir des méthodes (c) $pPLV^{nonpar}(0, 0)$, (d) $pPLV^{nonpar}(\lambda, \gamma)$, (e) $pPLV^{param}(\lambda, \gamma)$, (f) $\overline{pPLV}^{param}(\lambda, \gamma)$, (g) $p\rho(\lambda, \gamma)$ et (h) $\overline{p\rho}(\lambda, \gamma)$	59
3.4	(a), (b), Exemples de signaux provenant du modèle de Roessler et de sa transformation avec un retard constant pour $t = 1$ et $i = 2$ et 3. - Matrices de CFs obtenues respectivement à partir des méthodes (c) $pPLV^{nonpar}(0, 0)$, (d) $pPLV^{nonpar}(\lambda, \gamma)$, (e) $pPLV^{param}(\lambda, \gamma)$, (f) $\overline{pPLV}^{param}(\lambda, \gamma)$, (g) $p\rho(\lambda, \gamma)$ et (h) $\overline{p\rho}(\lambda, \gamma)$	61
3.5	(a), (b), Exemples de signaux provenant du modèle de Roessler et de sa transformation avec une amplitude constante pour $t = 1$ et $i = 2$ et 3. - Matrices de CFs obtenues respectivement à partir des méthodes (c) $pPLV^{nonpar}(0, 0)$, (d) $pPLV^{nonpar}(\lambda, \gamma)$, (e) $pPLV^{param}(\lambda, \gamma)$, (f) $\overline{pPLV}^{param}(\lambda, \gamma)$, (g) $p\rho(\lambda, \gamma)$ et (h) $\overline{p\rho}(\lambda, \gamma)$	63
3.6	(a), (b), Exemples de signaux provenant du modèle de Roessler et de sa transformation avec une modulation d'amplitude pour $t = 1$ et $i = 2$ et 3. - Matrices de CFs obtenues respectivement à partir des méthodes (c) $pPLV^{nonpar}(0, 0)$, (d) $pPLV^{nonpar}(\lambda, \gamma)$, (e) $pPLV^{param}(\lambda, \gamma)$, (f) $\overline{pPLV}^{param}(\lambda, \gamma)$, (g) $p\rho(\lambda, \gamma)$ et (h) $\overline{p\rho}(\lambda, \gamma)$	65
3.7	(a), (b), Exemples de signaux provenant du modèle de Roessler et de sa transformation avec une sur-modulation pour $t = 1$ et $i = 2$ et 3. - Matrices de CFs obtenues respectivement à partir des méthodes (c) $pPLV^{nonpar}(0, 0)$, (d) $pPLV^{nonpar}(\lambda, \gamma)$, (e) $pPLV^{param}(\lambda, \gamma)$, (f) $\overline{pPLV}^{param}(\lambda, \gamma)$, (g) $p\rho(\lambda, \gamma)$ et (h) $\overline{p\rho}(\lambda, \gamma)$	67
3.8	Proportion des CFs les plus actives communes à 2 crises, 3 crises, (4 crises) ou spécifique à une crise pour les patients 1 à 5 respectivement figure (a)-(e).	69
3.9	Matrices de CFs d'un enregistrement iEEG pendant une crise, en utilisant : (a) la PLV, (b) la pPLV non paramétrique sans régularisation $pPLV(0, 0)$, (c) la pPLV non paramétrique avec régularisation $pPLV(\lambda, \gamma)$	70
3.A.1	Exemple de sous-gradient (en rouge) représenté par des droites de pente p pour la fonction $f(x) = x $ en $x = 0$	75
4.1	Illustration des 4 premières composantes obtenues via la décomposition en valeur singulière (SVD).	83
4.2	Illustration de la décomposition en matrices non-négatives (NMF).	84
4.3	Illustration de la décomposition matricielle liée à k -means.	86
4.4	Illustration de la décomposition canonique polyadique (CP).	90
4.5	Illustration de la décomposition en valeurs singulières d'ordre supérieure (HOSVD).	91
4.6	Illustration de Parafac2.	93

5.1	(a) 4 signaux sEEG simulés, (b) Matrice de CFs $\mathbf{X} \in \mathbb{R}^{L \times T}$ entre les 6 paires de signaux calculées par PLV (voir la Définition 9), (c) Sous-graphe obtenu après application de k-means sur la matrice de CFs avec $N = 3$ partitions. Sur la représentation, les nœuds correspondent aux électrodes 1 à 4, et les liens correspondent à la solution du partitionnement. Les deux graphes rouges sont associés à des liens qui s'activent de manière synchrone dans le temps, le graphe bleu contient les liens restants et non synchronisés.	97
5.2	(a) Matrice de CFs de données sEEG réelles, (b) exemple d'un modèle synthétique pour une crise avec comme paramètres d'incertitudes $\mathbf{b} = [1, 0.2, 0.1, 0.1]$	110
5.3	Réalisation du tenseur matricé $\mathbf{X}_{(L)}$ correspondant à 4 crises (ou matrice de CFs) avec pour paramètres d'incertitudes $\mathbf{b} = [1, 0.2, 0.05, 0]$	110
5.4	ARI score pour chaque méthode, en fonction de 7 modèles associés à 7 vecteurs b_i pour les configurations (1) à (7) du Tableau 5.2.	112
5.5	ARI score pour chaque méthode, en fonction de 7 modèles associés à 7 vecteurs b_i pour les configurations (8) à (15) du Tableau 5.2.	112
5.6	(haut) Profils d'activation de la HOSMR (variable \mathbf{V} de l'Algorithme 9). (bas) Partition des CFs correspondant aux 4 étapes d'activation de la crise du patient considéré).	114
5.A.	Score $tvar(t)$ pour les 100 premières itérations en utilisant, comme entrée des algorithmes HOOI et mHOOI, le tenseur \mathcal{X} avec le vecteur d'incertitudes \mathbf{b} correspondant. Un zoom est représenté sur les dernières itérations et un autre pour les premières itérations montrant les premières valeurs du score $ldvar(t)$	116
6.1	Représentation de la concaténation de quatre matrices de CFs $\mathbf{X}\{s\} \in \mathbb{R}^{L \times T(s)}$, $\forall s \in \{1, \dots, 4\}$ correspondant au choix particulier $[\alpha = 1, \beta = 0, \sigma = 0.05, \delta = 0]$	126
6.2	(a) représentation des trois facteurs de CFs organisés selon 3 classes, (b)-(c) un exemple de matrice \mathbf{F} contenant les facteurs de CFs estimés respectivement pour $K = 3$ ou $K = 4$	127
6.3	Score $\rho^{(*)}$ atteint par les 5 méthodes proposées pour 4 conditions expérimentales avec $\sigma = 0.05$ et (a) $\alpha = 0$, (b) $\alpha = 1$	128
6.4	Score $\rho^{(*)}$ atteint par les 5 méthodes proposées pour 4 conditions expérimentales avec $\sigma = 0.5$ et (a) $\alpha = 0$, (b) $\alpha = 1$	129
6.5	Score $\rho^{(*)}$ atteint par les 5 méthodes proposées pour 4 conditions expérimentales avec $\sigma = 1.5$ et (a) $\alpha = 0$, (b) $\alpha = 1$	129
6.6	(a) Les cinq profils d'activation de la crise générique, obtenus grâce à la méthode HOSMR pour le patient avec des crises de durées homogènes. (b) - (f) Sous-graphes obtenus à partir des facteurs de CFs 1 - 5 fournis par la méthode HOSMR.	132
6.7	Les cinq profils d'activation spécifiques à chacune des 4 crises obtenus grâce à la méthode snnP2 pour le patient avec des crises de durées homogènes. Le trait rouge indique le début de la crise.	132
6.8	(a) - (e) Sous-graphes obtenus à partir des facteurs de CFs 1 - 5 fournis par la méthode snnP2.	132
6.9	Les cinq profils d'activation spécifiques à chacune des 4 crises obtenus grâce à la méthode multNMF pour le patient avec des crises de durées homogènes. Le trait rouge indique le début de la crise.	133
6.10	Sous-graphes obtenus à partir des facteurs de CFs fournis par la méthode multNMF pour le patient avec des crises homogènes. Les graphes sont rangés par crises et en groupes qui semblent apporter une information clinique similaire (quand c'est possible).	133
6.11	(a) Les cinq profils d'activation de la crise générique, obtenus grâce à la méthode joinNMF pour le patient avec des crises de durées homogènes.	134
6.12	(a) - (e) Sous-graphes obtenus à partir des facteurs de CFs 1 - 5 fournis par la méthode joinNMF.	134
6.13	(a) Les cinq profils d'activation de la crise générique obtenus grâce à la méthode HOSMR pour le patient avec des crises de durées hétérogènes. (b) - (f) Sous-graphes obtenus à partir des facteurs de CFs 1 - 5 fournis par la méthode HOSMR.	136

6.14	Les cinq profils d'activation de la crise générique, obtenus grâce à la méthode <code>snnP2</code> pour le patient avec des crises de durées hétérogènes.	136
6.15	Graphes binaires de CFs obtenus pour chaque profil d'activation selon la méthode <code>snnP2</code> . Les graphes de CFs (a) - (e) sont respectivement associés aux profils d'activation 1 - 5.	136
6.16	Les cinq profils d'activation spécifiques à chacune des 4 crises obtenus grâce à la méthode <code>multNMF</code> pour le patient avec des crises de durées hétérogènes. Le trait rouge indique le début de la crise.	137
6.17	Sous-graphes obtenus à partir des facteurs de CFs fournis par la méthode <code>multNMF</code> pour le patient avec des crises hétérogènes. Les graphes sont rangés par crise et en facteurs qui semblent apporter une information clinique similaire (quand c'est possible).	137
6.18	(a) Les cinq profils d'activation de la crise générique, obtenus grâce à la méthode <code>joinNMF</code> pour le patient avec des crises de durées hétérogènes.	138
6.19	(a) - (e) Sous-graphes obtenus à partir des facteurs de CFs 1 - 5 fournis par la méthode <code>joinNMF</code>	138
7.1	Aperçu de la méthode BTND : (a) On découpe chaque enregistrement en segments temporels, avec potentiellement du recouvrement. (b) Pour chaque segment, on calcule pour chaque paire d'électrodes le niveau de connectivité grâce à une mesure de CFs. (c) On réorganise les mesures FC dans un vecteur. Ces dernières, empilées dans le temps, forment pour chaque crise une matrice CF. (d) la liste des matrice de CFs est décomposée en sous-graphes avec leurs profils d'activation temporelle.	150
7.2	En haut, sont représentées les principales structures cérébrales ciblées par les électrodes intracrâniennes pour le patient 1. Ensuite, on montre l'enregistrement de deux crises (crise 1 et 2) pour un sous-ensemble d'électrodes sélectionnées. Sous chaque enregistrement, on fournit les profils d'activation spécifiques de tous les sous-graphes obtenus. les lignes rouges et vertes correspondent respectivement aux instants du début et de fin de crise.	158
7.3	Les 6 sous-graphes révélés par la méthode BTND pour le patient 1.	159
7.4	En haut, sont représentées les principales structures cérébrales ciblées par les électrodes intracrâniennes pour le patient 2. Ensuite, on montre l'enregistrement de deux crises (crise 1 et 2) pour un sous-ensemble d'électrodes sélectionnées. Sous chaque enregistrement, on fournit les profils d'activation spécifiques de tous les sous-graphes obtenus. les lignes rouges et vertes correspondent respectivement aux instants du début et de fin de crise.	160
7.5	Les 7 sous-graphes révélés par la méthode BTND pour le patient 2. La ligne pointillée sépare les électrodes situées à l'hémisphère droit et à l'hémisphère gauche du cerveau.	161

Liste des tableaux

1.1	Récapitulatifs des travaux présentés dans la Section 1.3	20
2.1	Réécriture de la formule générale de l'Équation (2.24) pour des valeurs particulières des paramètres (m, p, q)	31
2.2	Réécriture de la formule générale conditionnée de l'Équation (2.86) pour des valeurs particulières des paramètres (m, p, q)	42
3.1	Récapitulation des différentes étapes pour obtenir 5 mesures conditionnelles de CFs, les notations sont précisées dans les définitions ou propositions associées à la première colonne.	53
3.2	Performance des méthodes $pPLV^{nonpar}$ et $pPLV^{param}$ sans régularisation $(0, 0)$, avec seulement la parcimonie $(\lambda, 0)$ et avec la régularisation parcimonieuse et temporelle (λ, γ) sur le modèle de Roessler.	57
3.3	Performances de la méthode $pPLV^{nonpar}(0, 0)$ (cas non régularisé) et des 5 mesures de CFs étudiées avec régularisations (λ, γ) sur le modèle de Roessler sans transformation.	59
3.4	Performances de la méthode $pPLV^{nonpar}(0, 0)$ (cas non régularisé) et des 5 mesures de CFs étudiées avec régularisations (λ, γ) sur le modèle de Roessler avec retard.	61
3.5	Performance de la méthode $pPLV^{nonpar}(0, 0)$ (cas non régularisé) et des 5 mesures de CFs étudiées avec régularisations (λ, γ) sur le modèle de Roessler avec amplitudes constantes.	63
3.6	Performance de la méthode $pPLV^{nonpar}(0, 0)$ (cas non régularisé) et des 5 mesures de CFs étudiées avec régularisations (λ, γ) sur le modèle de Roessler avec modulation d'amplitude.	65
3.7	Performance de la méthode $pPLV^{nonpar}(0, 0)$ (cas non régularisé) et des 5 mesures de CFs étudiées avec régularisations (λ, γ) sur le modèle de Roessler avec sur modulation.	67
5.1	Liste des méthodes comparées afin de réduire le tenseur \mathcal{X} en une matrice de facteur de CFs \mathbf{F} . Les implémentations utilisées pour les expériences peuvent être trouvées en utilisant les liens suivants : 1) Algorithmes utilisés dans l'article [77] FCTensDec ; (2) N-way toolbox version 3.30[9] ; (3) MATLAB Tensor Toolbox Version 2.6 [17] ; (4) SPAMS toolbox version 2.6 [140]	111
5.2	Une encoche est indiquée lorsque l'incertitude b_i est active dans la génération du modèle synthétique.	111
5.C.1	Table de contingence entre les partitions C^{exp} et C^{true} avec $P = 4$	117
6.1	Liste des méthodes comparées afin de convertir les matrices de CFs $\mathbf{X}\{s\}$, $\forall s \in \{1, \dots, S\}$ en une matrice de facteur de CFs \mathbf{F} . Le temps de calcul correspond à la durée moyenne nécessaire à chaque méthode pour calculer les facteurs de CFs de toutes les expériences proposées.	127
7.1	Détails cliniques de chaque patient. légende ; HS : Sclérose hippocampique / LT : temporal gauche / RT : temporal droit/ hypoM : hypometabolisme / ATL : Lobectomie temporal antérieure / NA : pas enregistré / Le résultat chirurgical est exprimé en classe Engel.	153
7.2	Liste des principales structures cérébrales ciblées par les contacts sEEG. Légende ; L : gauche, R : droite, NA : pas enregistré, ANT HIPPOC : hippocampe antérieur, POST HIPPOC : hippocampe postérieur, OFC : Orbito Frontal Cortex, ANT CING : Anterior Cingulate, LAT FRONTAL : Lateral Frontal Cortex.	154
7.3	Comparaison qualitative entre l'ensemble des structures activées, déterminées par analyse visuelle et la méthode BTND pour les patients 1 et 2	162

- 7.4 Comparaison qualitative entre l'ensemble des structures activées, déterminées par analyse visuelle et la méthode BTND pour le patient 3, 4, et 5 (voir légende dans la Tableau 7.3) 163
- 7.5 Comparaison qualitative entre l'ensemble des structures activées, déterminées par analyse visuelle et la méthode BTND pour le patient 6, 7, 8 et 9. (voir légende dans le Tableau 7.3) 164

Glossaire

Indices

i , (ou j) : Signal d'une électrode
 n , (ou m) : Échantillon temporel
 l : Mesure de connectivité fonctionnelle
 t : Segment temporel
 k : Facteur suite à une réduction des données
 s : Crise d'épilepsie
 p : Classe suite à un partitionnement des données

Signaux

$s(n)$: Signal temporel
 $z(n)$: Signal analytique
 $\epsilon(n)$: Signal conditionné
 $a(n)$: Amplitude instantanée
 $\phi(n)$: Phase instantanée
 $\psi(n)$: Phase relative
 $v(n)$: Multiplication d'amplitudes
 $w(n)$: Proportion logarithmique des amplitudes

Matrices et tenseurs

\mathcal{X} : Mesure de connectivités fonctionnelles en fonction du temps et des crises
 $\mathbf{X}\{s\}$: Liste de matrice de connectivités fonctionnelles en fonction du temps
 \mathbf{K} : Matrice de covariance
 Θ : Matrice de précision
 \mathbf{F} : Matrice de facteurs de connectivités fonctionnelles (chap 4-7)
 \mathbf{V} : Matrice de profils d'activation temporelle (chap 4-7)
 \mathbf{A} : Matrice indicatrice normalisée pour partitionner les données (chap 4-7)
 Δ : Laplacien indiquant le niveau de similarité entre chaque paire de crise (chap 4-7)

Mesures de connectivités fonctionnelles

CF : Connectivité fonctionnelle
 ρ : Corrélation
 $p\rho$: Corrélation partielle
 PLV : Phase-Locking-Value
 $pPLV$: PLV conditionnée

AEC : Corrélation des amplitudes

Fonctions

$D_{s_1 s_2}(f)$: Densité spectrale d'énergie croisée entre les signaux s_1 et s_2

$\mathcal{H}[\bullet]$: Transformée de Hilbert

$I_\kappa(\bullet)$: Fonction de Bessel modifiée de première espèce

$K_\kappa(\bullet)$: Fonction de Bessel modifiée de deuxième espèce

$\mathcal{K}(\bullet)$: Intégrale elliptique de première espèce

$\mathcal{E}(\bullet)$: Intégrale elliptique de deuxième espèce

${}_2F_1(\bullet)$: Fonction hypergéométrique de Gauss

$p(\bullet)$: Densité de probabilité

$\mathcal{L}(\bullet)$: Lagrangien

$\mathcal{L}_\kappa(\bullet)$: Lagrangien augmenté

$\nabla f(\bullet)$: Gradient de la fonction f

$\partial f(\bullet)$: Sous-gradient de la fonction f

$SD_\lambda(\bullet)$: Fonction de seuillage doux

$ST_\lambda(\bullet)$: Fonction de seuillage doux complexe

$NSD_\lambda(\bullet)$: Fonction de seuillage doux non négatif

Introduction

Motivations

Les données ont un rôle essentiel dans la société moderne. Elles sont l'une des ressources fondamentales des progrès de notre civilisation pour comprendre le monde et prendre de meilleures décisions. De nos jours, l'analyse de données est un domaine particulièrement important pour étudier les systèmes complexes, c'est-à-dire les ensembles constitués d'un grand nombre d'entités en interaction et dont on ne peut prévoir leurs évolutions par le calcul. Comme exemple de système complexe, on peut considérer la bourse. Les briques de base de ce système sont des courtiers qui effectuent des transactions, achètent ou vendent des actions. Localement les lois qui régissent ce système sont donc simples. Cependant, le comportement global de la bourse est imprévisible, chaotique et cela nous rend dépendant aux krachs boursiers qui impactent parfois durablement la vie de milliard de personnes. De nombreux exemples de systèmes complexes nous entourent au quotidien, on peut citer la dynamique d'une cellule, le cerveau, les réseaux sociaux, une fourmilière, la météo, etc...

Une approche récente pour analyser un système complexe est de modéliser les connections entre les différentes entités qui le constitue. C'est-à-dire de connaître le réseau caractérisant les principales interactions qui ont un rôle sur l'évolution du système. Ces modèles peuvent être identifiés uniquement grâce à de nombreuses données collectées sur le système. Ainsi, grâce aux progrès récents qui ont été réalisés pour collecter et amasser des données, la science des réseaux est en train de devenir un enjeu essentiel au XXI^{ème} siècle.

De nombreuses attentions ont été portées récemment sur l'inférence et l'analyse des réseaux statiques. On peut par exemple citer l'analyse spectrale des graphes ou le traitement des signaux sur graphe. Cependant, l'élaboration d'outils méthodologiques adaptés à l'analyse des réseaux qui évoluent dans le temps, aussi nommés réseaux dynamiques, est une thématique de recherche encore récente. Pourtant, ce domaine est essentiel pour analyser de nombreux phénomènes. Par exemple, l'étude de la dynamique des réseaux sociaux comme Twitter ou Facebook est actuellement un enjeu majeur pour comprendre et limiter la propagation des "fake news".

La figure 1 montre l'exemple d'un réseau dynamique que l'on peut construire à partir de signaux collectés. Chaque signal indique l'activité d'une entité constitutive du système complexe. On précise qu'il s'agit ici d'un enregistrement neuro-physiologique, mesurant l'activité de régions locales du cerveau. A différents instants, des segments temporels sont extraits pour mesurer la similarité entre chaque paire de signaux. L'évolution de ce réseau permet de comprendre comment les différentes entités interagissent entre elles dans le temps. Ce genre de réseau dynamique peut être particulièrement difficile et chronophage à analyser. En effet, le nombre de liens à étudier est quadratique avec le nombre de signaux. De plus, il est nécessaire d'interpréter leurs évolutions en fonction du temps. Malheureusement, les outils élaborés pour les réseaux uniques se généralisent mal dans ce contexte. C'est ce constat qui a motivé le sujet de cette thèse, c'est à dire de trouver des outils de décomposition et de représentation pertinente d'un graphe dynamique.

Un enjeu applicatif particulièrement intéressant se situe dans la compréhension d'une maladie neurologique : l'épilepsie. Dans l'objectif d'extraire le foyer de l'épilepsie responsable de la crise, des électrodes intracrâniennes sont placées dans des régions d'intérêts du cerveau pour enregistrer leurs activités au cours du temps. La compréhension de la dynamique du réseau pathologique provoquée par une crise offre des informations précieuses pour s'assurer du succès de la chirurgie. Il faut alors, à partir des signaux enregistrés, inférer et analyser un réseau dynamique qui représente l'évolution du cerveau lors d'une crise.

La littérature sur l'analyse des graphes statiques, c'est-à-dire un graphe représentant l'intégralité de l'enregistrement est vaste et a permis d'offrir de nombreuses avancées en neurosciences. Il existe déjà des méthodes permettant de diagnostiquer si un patient souffre de la maladie d'Alzheimer, ou de modéliser l'influence de l'activité sportive sur le réseau cortical. Cependant, ces méthodes ne peuvent s'appliquer à l'épilepsie qui est un processus dynamique par nature. Parmi les stratégies permettant d'analyser un graphe dynamique issu de données psychophysiques, les méthodes de décompositions modales sont avantageuses car elles permettent de montrer des relations complexes entre des groupes de connexions caractéristiques dans le temps. De plus, ces méthodes permettent d'intégrer des modalités supplémentaires comme les différentes crises d'un même patient afin d'obtenir des résultats plus robustes. Le problème de ce type de décompositions est qu'elles peuvent produire des résultats peu intelligibles. Cela montre la nécessité de proposer des décompositions modales qui permettent d'extraire des composantes interprétables cliniquement.

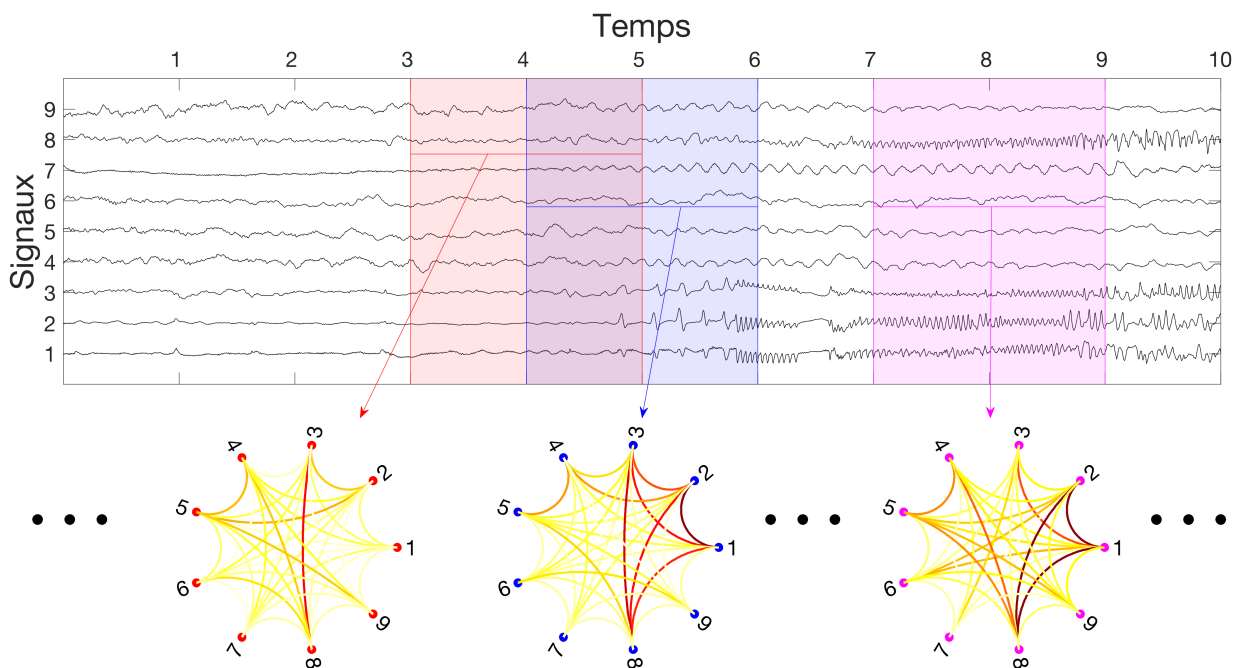


FIGURE 1 – Graphe dynamique inféré à partir de signaux temporels, indiquant l'évolution des similarités entre les signaux dans le temps. Chaque graphe est calculé sur une portion du signal multivarié. Le poids des liens provient d'une mesure de similarité calculée entre chaque paire de signaux.

Solutions proposées

Afin de produire des outils innovants pour visualiser et analyser les graphes dynamiques complexes, on propose deux stratégies différentes :

- L'inférence de réseau dynamique robuste à partir de mesures de connectivités fonctionnelles, qui souligne uniquement les connections essentielles pour caractériser la crise d'épilepsie que l'on cherche à étudier. Cette stratégie est étudiée dans les chapitres II et surtout III.
- La décomposition du réseau dynamique en facteurs indiquant des groupes de connections particulièrement importants pour caractériser la crise d'épilepsie. Ces facteurs sont associés à des profils d'activation temporelle afin de comprendre leurs dynamiques et comment ils s'organisent ou se structurent dans le temps. Cette stratégie est étudiée dans les chapitres V, VI et VII.

L'organisation de cette thèse est la suivante :

Chapitre 1 : Chapitre d'introduction en neurosciences computationnelles, montrant les méthodes d'inférence et d'analyse de réseaux obtenus grâce à des données neurophysiologiques. Dans un premier

temps, on présente les données étudiées qui sont des enregistrements sEEG des crises d'épilepsie provenant de patients souffrant d'épilepsie focale, puis, la notion essentielle de connectivité fonctionnelle (CF) est introduite. Ces mesures permettent d'inférer des réseaux dynamiques qui sont des représentations imparfaites de l'évolution corticale du cerveau durant une crise d'épilepsie. Enfin on propose un état de l'art original sur les méthodes d'analyses et de décompositions des réseaux dynamiques en neurosciences.

Chapitre 2 : Ce chapitre se concentre sur la notion de CF qui est essentielle pour inférer les graphes dynamiques caractérisant la crise d'épilepsie. Il existe de nombreuses mesures de CFs utilisées selon l'application. On peut citer la corrélation, la cohérence, la Phase-Locking-Value (PLV) ou la corrélation des amplitudes par exemple. On propose dans ce chapitre le cadre de signaux supposés localement gaussiens et stationnaires pour comparer différentes mesures de CFs. On propose une expression qui semble être originale des moments d'ordres supérieurs entre la phase et l'amplitude de deux signaux gaussiens. Ces expressions permettent de décrire le comportement dans le cas gaussien de nombreuses mesures de CFs impliquant la phase ou l'amplitude des signaux. Dans la dernière partie de ce chapitre on propose une extension de l'expression proposée aux cas multivariés pour obtenir les moments d'ordres supérieurs entre la phase et l'amplitude de deux signaux gaussiens conditionnés aux autres signaux du jeu de données.

Chapitre 3 : Ce chapitre se concentre sur une mesure de CFs particulière, la PLV. Afin d'obtenir des réseaux dynamiques de CFs robustes et facilement interprétables, on propose deux nouvelles mesures conditionnelles et régularisées de la PLV. La première se base sur la mesure usuelle de PLV non paramétrique ne faisant pas d'hypothèse de normalité. La seconde sur la mesure de PLV conditionnelle proposée dans le chapitre précédent, supposant que le signal suit une distribution gaussienne multivariée. On conditionne les mesures de PLV afin de prendre en compte les autres signaux enregistrés. Les régularisations sont choisies pour obtenir des réseaux dynamiques parcimonieux et lisses dans le temps. Afin de comparer ces deux mesures de PLV conditionnelles régularisée, et de montrer leurs performances, on exécute ces méthodes sur des signaux tests. Enfin, on applique cette nouvelle mesure de CFs pour estimer le graphe dynamique à partir d'enregistrements réels de crises d'épilepsie.

Chapitre 4 : Ce chapitre permet d'apporter le contexte nécessaire pour comprendre les méthodes de décomposition de graphes dynamiques employées par la suite. On propose l'idée originale d'exploiter plusieurs crises d'un même patient pour plus de robustesse. On les traite conjointement en les agrégeant dans une structure multidimensionnelle : le tenseur. Les méthodes usuelles de décompositions matricielles et tensorielles sont alors présentées.

Chapitre 5 : On propose dans ce chapitre une méthode afin de partitionner un réseau dynamique en un ensemble de sous-graphes contenant une information neuro-physiologique pertinente. Afin d'obtenir des composantes robustes, on propose de considérer plusieurs crises d'un même patient. Malgré la prise en compte de plusieurs crises, on réalise toujours un partitionnement sur des données bruitées et de grandes dimensions. Pour faire face à ces limites, on développe une méthode de réduction dimensionnelle qui extrait les caractéristiques pertinentes des CFs communes à toutes les crises sous forme de facteurs. Étant donné que les données correspondent à des CFs évoluant dans le temps et pour différentes crises, elles peuvent être commodément représentées comme un tenseur de 3-modes. La réduction dimensionnelle proposée est une décomposition tensorielle originale et spécifique au problème qui impose des contraintes structurelles à la solution. Il s'agit de rendre simultanément les facteurs de CFs et les profils d'activation parcimonieux. Alors, le partitionnement est réalisé sur les facteurs de CFs de dimension réduite à la place du jeu de données initiales. Pour évaluer la capacité de notre stratégie à partitionner le réseau dynamique, on la compare à d'autres approches de réduction de dimension (y compris d'autres décompositions de tenseurs). Enfin, cette décomposition de graphes dynamiques est appliquée sur des enregistrements sEEG réels.

Chapitre 6 : Ce chapitre étend la décomposition tensorielle proposée dans le chapitre 5 dans le cas où les crises d'épilepsie ont des durées hétérogènes. On propose alors dans ce chapitre plusieurs

extensions, une basée sur l'algorithme Parafac2, et deux versions différentes de la décomposition en matrices non négatives jointes. On compare ces différentes méthodes entre elles sur des graphes dynamiques de CFs simulées. Enfin on applique les différentes approches proposées sur les crises de deux patients différents. Le premier ayant des crises de durées similaires, tandis que le second possède des crises de durées variables.

Chapitre 7 : Chapitre d'application clinique. On propose, à partir de la méthode la plus avantageuse présentée dans le chapitre 6, une stratégie semi-automatique pour caractériser le réseau pathologique de CFs de patients épileptiques. La stratégie proposée est appliquée à 27 crises provenant de 9 patients différents qui souffrent d'épilepsie focale résistante aux médicaments. La décomposition obtenue est confrontée dans chaque cas à l'interprétation du clinicien.

Publications

Journal

1. Gaëtan Frusque, Julien Jung, Pierre Borgnat, Paulo Gonçalves, (2020). "**Multiplex network inference with sparse tensor decomposition for functional connectivity**", IEEE Transactions on Signal and Information Processing over Networks, 6, 316-328.
→ Les travaux de cet article sont présentés dans le chapitre 5.
2. Gaëtan Frusque, Pierre Borgnat, Paulo Gonçalves, Julien Jung (2020). "**Semi-automatic extraction of functional dynamic networks describing patient's epileptic seizures**", Frontiers in Neurology (submitted 03/07/2020).
→ Les travaux de cet article sont présentés dans le chapitre 7.

Conférence

1. Gaëtan Frusque, Julien Jung, Pierre Borgnat, Paulo Gonçalves, (2020). "**Regularized partial phase synchrony index applied to dynamical functional connectivity estimation**", ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5955-5959).
→ Les travaux de cet article sont présentés dans le chapitre 3.
2. Gaëtan Frusque, Pierre Borgnat, Paulo Gonçalves, (2020). "**Temporarily activated patterns for multi-trial functional connectivity data**", ASPAI'2020 (accepted).
→ Les travaux de cet article sont présentés dans le chapitre 6.
3. Gaëtan Frusque, Julien Jung, Pierre Borgnat, Paulo Gonçalves, (2019). "**Sparse tensor dimensionality reduction with application to clustering of functional connectivity**", In Wavelets and Sparsity XVIII (Vol. 11138, p. 111380N). International Society for Optics and Photonics.
→ Les travaux de cet article sont présentés dans le chapitre 5.
4. Gaëtan Frusque, Julien Jung, Pierre Borgnat, Paulo Gonçalves, (2019). "**Réduction de dimension tensorielle parcimonieuse : Application au clustering de connectivité fonctionnelle**", GRETSI 2019-XXVIIème Colloque francophone de traitement du signal et des images. 2019.
→ Les travaux de cet article sont présentés dans le chapitre 5.

Notations

Dans ce manuscrit, on fixe les notations suivantes : les vecteurs, les matrices et les tenseurs sont respectivement notés en gras minuscule ($\mathbf{x}, \mathbf{y}, \dots$), en gras majuscule ($\mathbf{X}, \mathbf{Y}, \dots$) et en gras calligraphique ($\mathcal{X}, \mathcal{Y}, \dots$). Un indice est noté en minuscule ($i, l, t\dots$), sa borne supérieure en majuscule ($I, L, T\dots$). Tandis que les vecteurs $\mathbf{x}_{:t}$ et \mathbf{x}_l correspondent respectivement à la colonne t et à la ligne l de la matrice \mathbf{X} , les matrices $\mathbf{X}_{:,k}$, $\mathbf{X}_{:t}$ et $\mathbf{X}_{l,:}$ sont des tranches du tenseur \mathcal{X} . La notation $[\mathbf{X}, \mathbf{Y}]$ exprime la concaténation horizontale des matrices \mathbf{X} et \mathbf{Y} . La matrice identité de dimension $N \times N$ est notée \mathbf{I}_N . $D(\mathbf{x})$ est une matrice diagonale dont les éléments diagonaux correspondent au vecteur \mathbf{x} . La matrice transposée, la matrice adjointe et la pseudo-inverse de \mathbf{X} sont notées respectivement \mathbf{X}^t , \mathbf{X}^* et \mathbf{X}^\dagger . L'opération qui calcule la trace d'une matrice est notée $\text{Tr}(\mathbf{X})$, celle qui calcule son déterminant est notée $\det(\mathbf{X})$. Finalement la norme de Frobenius se note $\|\bullet\|_F$, la norme 1 terme à terme se note $\|\bullet\|_1$, on précise que \bullet correspond à un vecteur, une matrice ou un tenseur en fonction du contexte.

Afin de différencier les notions de connectivités fonctionnelles et de signaux sEEG, ces derniers sont notés $\mathbf{s}(n)$ pour indiquer le signal I -dimensionnel fonction de l'échantillon temporel n . Le nombre imaginaire est noté $\mathbf{i} = \sqrt{-1}$ et l'espérance mathématique est notée $\mathbb{E}\{\bullet\}$

Chapitre 1

Contexte : Étude des réseaux dynamiques en neurosciences

1.1 Mesurer l'activité cérébrale lors d'une crise d'épilepsie

1.1.1 Épilepsie

L'épilepsie est l'un des troubles neurologiques les plus répandus dans le monde, affectant environ 1% de la population. L'épilepsie recouvre plusieurs maladies qui ont toutes en commun la répétition de crises épileptiques spontanées. Une crise d'épilepsie est une interruption de l'activité normale du cerveau. Elle se traduit par des manifestations variées allant de troubles de la motricité, de la sensibilité ou du comportement, jusqu'à des secousses convulsives ou une perte de conscience (selon qu'il s'agit d'une crise partielle ou généralisée).

Environ 40% des patients ont des crises résistantes aux médicaments, dans ce cas une intervention chirurgicale peut être proposée afin d'extraire la zone du cerveau qui est responsable des crises, cette zone est nommée le foyer épileptique. Le patient subit une batterie de mesures, notamment EEG, magnétoencéphalographie (MEG) et électroencéphalographie stéréotaxique (SEEG) afin de déterminer la position du foyer et la faisabilité de l'opération.

Les sciences computationnelles appliquées à l'épilepsie ont principalement pour objectif d'aider le clinicien à interpréter les enregistrements réalisés. On peut séparer trois types d'applications majeures. Le premier concerne la localisation du foyer épileptique [28]. Le second s'intéresse à la prédiction de crise, c'est-à-dire à trouver des motifs caractéristiques d'un début de crise [88]. Enfin, le dernier consiste à trouver une visualisation pertinente de la crise (par l'intermédiaire de graphes, de partitionnements ou de décompositions modales par exemple) [115]. C'est dans ce contexte que s'appliquent les méthodes développées dans cette thèse. Dans une moindre mesure, les méthodes développées peuvent aussi aider à localiser le foyer épileptique.

1.1.2 Origine des signaux du cerveau

Le cerveau est le régulateur des autres systèmes d'organes du corps. Il est composé de deux hémisphères et de plusieurs régions corticales associées à des fonctions cérébrales différentes (comme le traitement de la parole ou de l'activité musculaire par exemple). Le composant principal du cerveau permettant de transporter et traiter l'information est la cellule neuronale, ou neurone. Pour un cerveau adulte on recense en moyenne cent milliards de neurones. La Figure 1.1a montre le schéma simplifié d'une cellule neuronale. Elle est composée de dendrites qui reçoivent des stimuli de plusieurs milliers d'autres neurones, d'un corps et d'un noyau cellulaire permettant à la cellule de fonctionner. Enfin, un axone transmet un stimulus à des milliers de dendrites appartenant à d'autres neurones par l'intermédiaire de synapses.

Les stimuli sont ici majoritairement des potentiels électriques qui codent et transmettent l'information. Quand le potentiel à la base d'un axone est suffisamment élevé, dû à la somme des potentiels provenant des dendrites, une impulsion électrique est générée et transmise le long de l'axone. Que ce soit le long d'un axone ou d'une dendrite, le déplacement du courant électrique crée une polarité

positive du côté où le courant part, et une polarité négative là où il arrive. Ce phénomène est à la base du courant primaire à l'origine du champ magnétique du cerveau. Comme les neurones sont entourés d'un milieu conducteur, un courant secondaire opposé au courant primaire est généré dans ce milieu. C'est ce courant secondaire qui est mesuré en électroencéphalographie (EEG) [152]. La modélisation en dipôle permettant de comprendre la génération des courants primaires et secondaires est schématisée dans la Figure 1.1b [28], [26].

Le courant individuel d'un neurone est trop faible pour être mesuré, on mesure en réalité principalement l'action d'un groupe de neurones avec un arrangement parallèle. Cet arrangement permet à l'ensemble des courants individuels de s'additionner pour former une activité d'une amplitude suffisante pour être enregistrée [28].

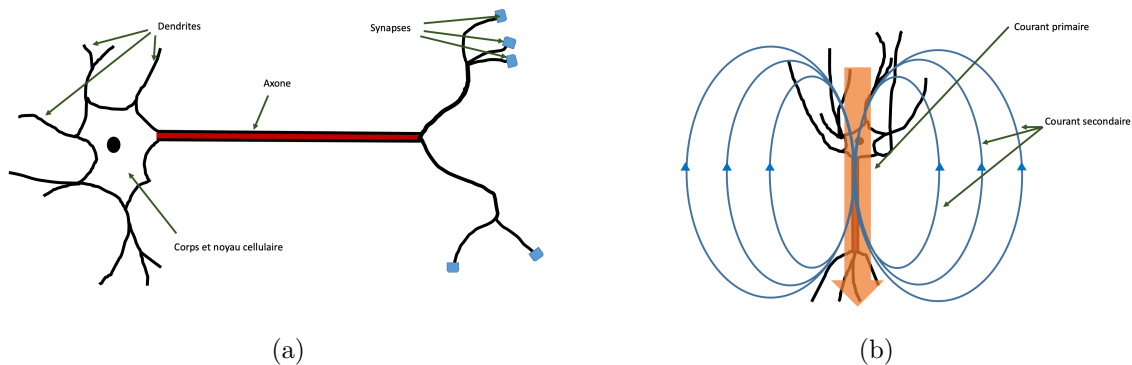


FIGURE 1.1 – (a) - Schéma d'un neurone, (b) - Le neurone, avec son axone et ses dendrites, modélisé comme un dipôle

1.1.3 Enregistrement électroencéphalographie stéréotaxique (SEEG)

La majorité des données traitées dans cette thèse proviennent d'enregistrement SEEG durant une crise d'épilepsie [89], [102]. Cette technique de mesure n'est pas aussi courante que l'EEG. En effet, pour les humains, elle est principalement utilisée dans le contexte d'épilepsie résistante aux médicaments afin d'identifier le foyer de l'épilepsie et de comprendre comment elle se propage dans le cerveau.

Contrairement à l'EEG qui mesure globalement l'activité cérébrale par des électrodes placées sur le cuir chevelu, la SEEG réalise des mesures locales qui enregistrent l'activité d'un faible nombre de neurones. Cette mesure est donc invasive, le patient reste plusieurs jours (environ 15 jours) à l'hôpital avec des électrodes implantées dans le cerveau. Les électrodes utilisées sont réparties sur des tiges, avec environ 5 à 10 électrodes par tige. L'espace entre deux électrodes consécutives est de 3.5 mm. La Figure 1.2a montre un exemple d'implémentation des tiges dans le cerveau et la Figure 1.2b montre une photo d'une tige stéréotaxique. Les mesures sont réalisées en continu, échantillonnées à 256 hz. L'objectif de cette démarche est d'enregistrer un maximum de crises d'épilepsie.

La position des électrodes est décidée au préalable par les cliniciens, elle cible des zones d'intérêt du cerveau qui semblent être le foyer de l'épilepsie, ces suppositions sont faites à partir d'autres mesures (EEG, MEG...) réalisées en amont. Pour la visualisation, on peut représenter les électrodes dans le référentiel de Tailarach, il s'agit d'un référentiel permettant de repérer la position de n'importe quel point dans le cerveau. Celui-ci est basé sur un ensemble de déformations permettant de réaligner les cerveaux de tous les individus. La Figure 1.2c représente un schéma du référentiel de Tailarach [53]. Comme montré dans la Figure 1.2d, les contacts stéréotaxiques peuvent être visualisés sur trois plans, le plan axial ($x-y$), le plan sagittal ($x-z$), le plan coronal ($y-z$).

Finalement, on présente dans la Figure 1.2e un exemple d'enregistrement SEEG pour une crise d'épilepsie. L'amplitude typique des signaux SEEG est de l'ordre du microvolt. Un découpage de la crise, réalisé par un clinicien, montre qu'elle peut être vue comme la succession de plusieurs étapes [115]. L'étape verte représente un enregistrement précédent la crise. Le début de crise correspond à l'étape rouge, on peut noter que positionner le marqueur de début de crise n'est pas une tâche triviale. L'étape bleue coïncide avec une étape de propagation de la crise dans les différentes zones mesurées, enfin l'étape grise montre la fin de la crise et un retour à l'état normal.

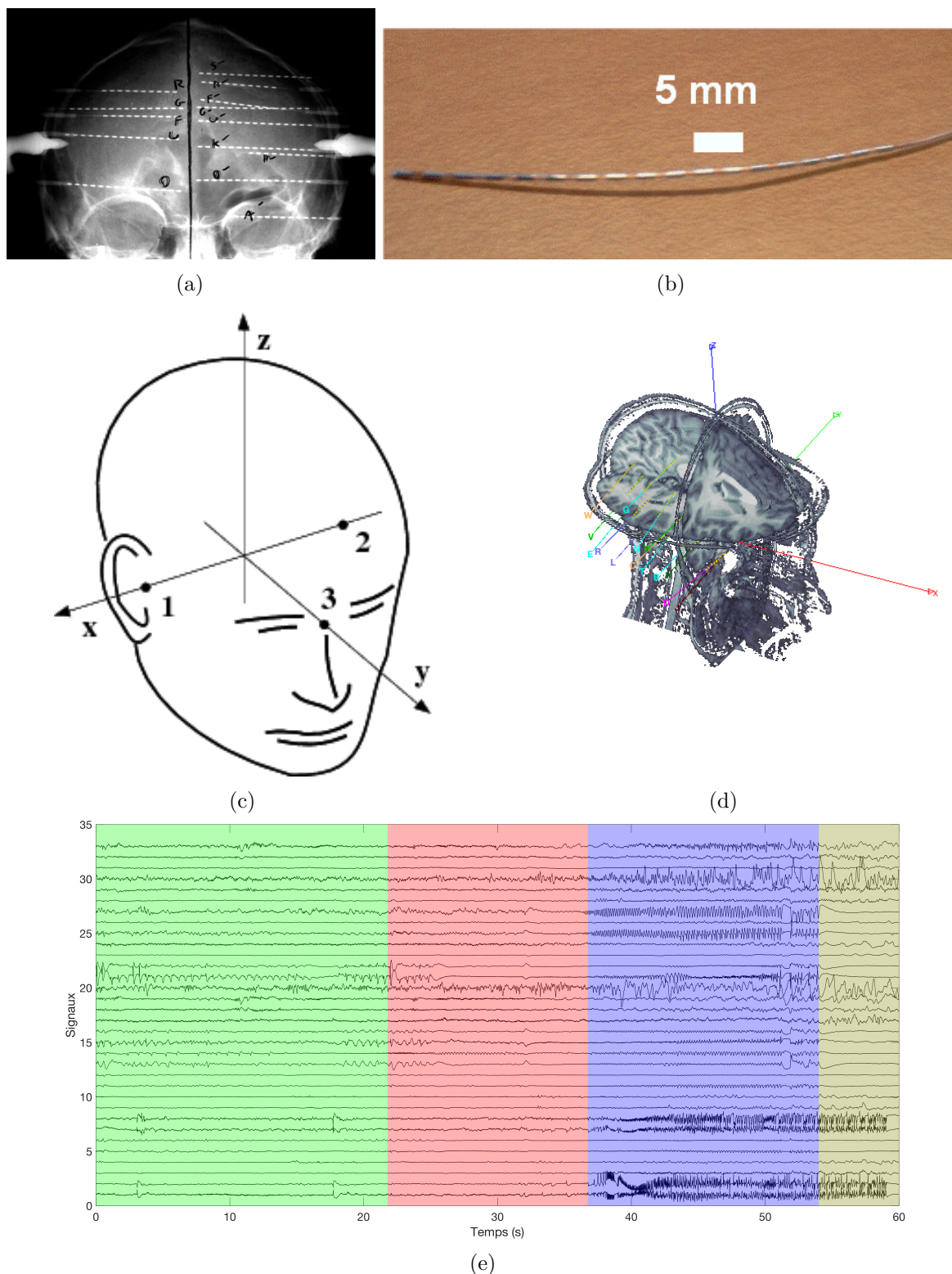


FIGURE 1.2 – (a) exemple d'implémentation des tiges stéréotaxiques dans le cerveau - (b) tiges stéréotaxiques - (c) schéma du référentiel de Tailarach¹ - (d) Électrodes stéréotaxiques visualisées selon le référentiel de Tailarach. Une coupe du cerveau selon les trois plans principaux (axial (x-y), sagittal (x-z), coronal (y-z)) est représentée - (e) Enregistrement d'une crise d'épilepsie par SEEG, on représente en couleur la succession de plusieurs étapes (le début de crise correspond à l'étape rouge, l'étape bleue coïncide avec la propagation de la crise dans les différentes zones mesurées, enfin l'étape grise montre la fin de la crise.)

1. https://mne.tools/0.11/manual/source_localization/forward.html

1.2 Les connectivités fonctionnelles

La plupart des crises d'épilepsie démarrent à cause d'une activité anormale due à une lésion dans une région localisée du cerveau, le foyer épileptique. La propagation de cette activité ictale dans différentes régions du cerveau est observée dans la plupart des cas. Il est alors possible de modéliser le comportement d'une crise, permettant de comprendre son action pathogène, en étudiant la dynamique des connexions entre les différentes régions du cerveau.

Pour établir ces connexions à partir de l'enregistrement SEEG mesurant l'activité locale de zones préalablement choisies par les cliniciens, on utilise des mesures de connectivité fonctionnelle (CF) [192]. Une mesure de CF estime la similarité entre deux signaux durant un intervalle de temps donné. Il s'agit alors d'une mesure indirecte de l'activité synchrone des deux régions monitorées par les électrodes considérées.

Plus de détails sur les mesures de CF sont fournis en Section 1.2.1 avec leurs expressions mathématiques en Section 1.2.2. Comme la notion de similarité est vague, et dépend surtout du contexte étudié, nous proposons dans la Section 1.2.3 une liste non exhaustive des principales mesures de CF appliquées aux signaux du cerveau. Avant toutes mesures, il est important de faire un prétraitement des signaux SEEG, pour corriger les signaux des erreurs et bruits connus. Les différents prétraitements réalisés sont détaillés dans l'Annexe 1.A.

1.2.1 Mesurer la similarité entre deux signaux

En observant l'activité des signaux de l'enregistrement d'une crise d'épilepsie (Figure 1.2e), on observe pour chaque étape des coïncidences d'activité très similaires (cela s'observe particulièrement bien pour l'étape bleue), qui caractérisent l'épilepsie. Ce sont ces similarités entre paires de signaux, qui révèlent une activité synchrone, et donc un lien entre les différentes régions cérébrales ciblées par chaque électrode. L'évolution de ces synchronisations, que l'on remarque par exemple entre les étapes bleues et rouges de la Figure 1.2e, marque le début et la propagation de la crise. Il est donc crucial d'estimer avec des mesures fines la similarité entre les signaux.

1) La mesure de connectivité fonctionnelle :

La connectivité fonctionnelle ne doit pas être confondue avec la connectivité structurelle [87] (mesurant la quantité de lien anatomique entre deux régions du cerveau), et la connectivité effective [75] (mesurant des interactions causales entre deux régions relatives à des tâches cognitives particulières). On la définit ainsi :

Definition 1 : Connectivité fonctionnelle (CF)

Mesure la dépendance statistique d'une paire de signaux du cerveau [5]. Une mesure de CF varie généralement de 0, quand les signaux sont totalement indépendants, à 1 lorsque la dépendance est maximale.

Si l'enregistrement est réalisé sur plus de 2 signaux, alors les mesures de CF sont réalisées pour chaque paire de signaux. Ainsi, si l'on considère I signaux, alors le nombre L de CF est :

$$L = \binom{I}{2} = \frac{I(I-1)}{2}, \quad (1.1)$$

2) Calcul dynamique des CFs :

On considère l'enregistrement SEEG de I signaux numériques de N échantillons temporels ($\mathbf{s}_i \in \mathbb{R}^N \forall i \in \{1, \dots, I\}$). La mesure dynamique de CF [192] pour chaque paire de signaux, consiste à calculer des CFs sur une fenêtre (rectangulaire) glissante. Plus précisément, il s'agit de sélectionner T échantillons temporels équidistants, notés $n_t \forall t \in \{1, \dots, T\}$. Comme la CF est une mesure statistique, on considère une fenêtre rectangulaire de largeur $2T_f$ autour de chaque échantillon n_t afin de sélectionner des segments sur lesquels on suppose le signal stationnaire. Après calcul des CFs pour chaque segment, on obtient $L \times T$ mesures qui représentent l'évolution dynamique des CFs durant la crise.

On illustre ce concept dans la Figure 1.3 où sont représentés deux signaux s_1 et s_2 (tracé du bas), et leurs mesures de CF en fonction du temps (tracé du haut). On remarque que la mesure de CF augmente durant les périodes de temps où les deux signaux semblent synchronisés. Pour les données SEEG, le pas de calcul est d'une seconde et la largeur de la fenêtre choisie est de deux secondes. Ce choix se fait sur la base du compromis suivant : réaliser le calcul de la CFs sur suffisamment d'échantillons (512 ici), échantillonner avec suffisamment de détails l'évolution de la crise. Comme les événements ictaux durent souvent plusieurs secondes le découpage de la crise ainsi réalisé paraît satisfaisant.

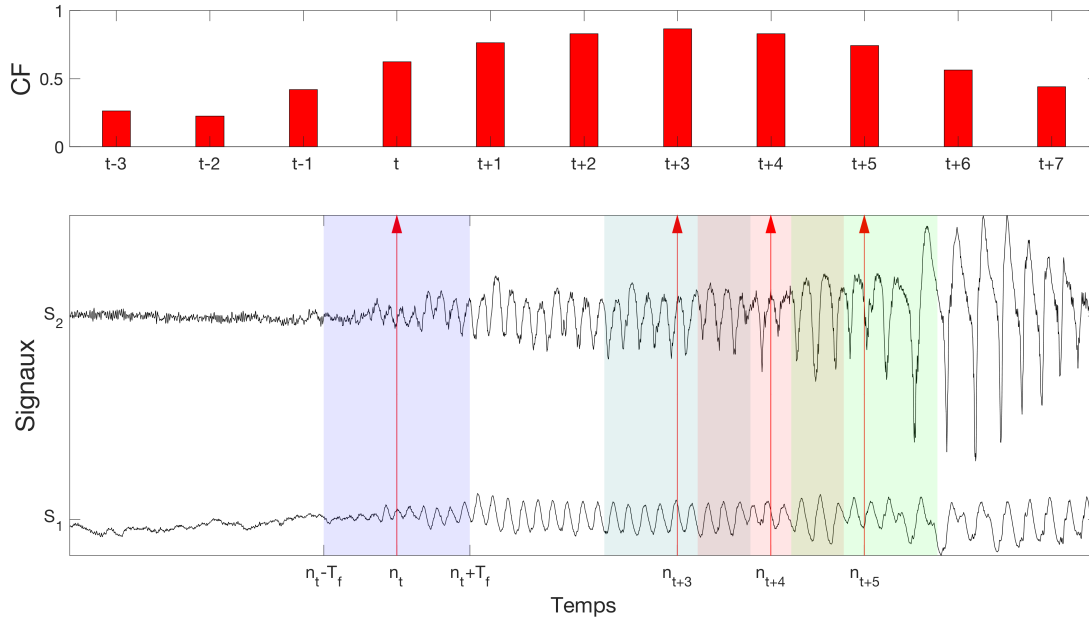


FIGURE 1.3 – Deux signaux s_1 et s_2 sont représentés sur le tracé du bas, et les mesures de CF correspondantes (notées $CF(s_1, s_2)$) en fonction du temps sont affichées sur le tracé du haut. Les mesures sont faites sur T échantillons temporels équidistants en utilisant des fenêtres rectangulaires de largeur $2T_f$.

1.2.2 Expression mathématique de mesures de CFs

1) Une structure de réseaux dynamiques :

Comme une mesure de CF est calculée sur chaque paire de signaux, elle peut être utilisée pour construire le graphe de dépendances entre chaque paire d'électrodes. Cette propriété est illustrée dans la Figure 1.4. 4 signaux sont représentés que l'on découpe en trois étapes, d'abord les signaux 1, 2 et 3 sont corrélés, puis tous les signaux sont indépendants, enfin une dernière étape avec seulement les signaux 1 et 4 de corrélés. Les structures de graphe obtenues pour deux segments localisés aux échantillons n_t et $n_{t'}$ sont représentées dans les Figures 1.4b et 1.4c, où la largeur des liens est proportionnelle à la mesure de CF. La mesure dynamique de CFs peut être associée à une succession de graphes dont les nœuds restent similaires, mais le poids des liens varie en fonction du temps. Chaque graphe peut être considéré comme une estimation bruitée de l'état cortical du cerveau à un instant donné. La succession de ces graphes forme un réseau dynamique[23] caractérisant l'évolution du cerveau.

Une première représentation mathématique est alors possible grâce à l'utilisation de la théorie des graphes [41] :

Definition 2 : Graphe de CFs

Un graphe de CF est noté \mathcal{G} . Il est défini par un triplet $(\mathcal{V}, \mathcal{E}, \mathbf{W}^{CF})$, où \mathcal{V} correspond aux nœuds du graphe, \mathcal{E} l'ensemble des liens de ce graphe. Ici \mathcal{V} correspond aux électrodes. Comme le calcul des CFs est réalisé pour chaque paire d'électrodes, le graphe est complet, c'est-à-dire que tous les liens de

l'ensemble \mathcal{E} sont actifs. Les poids associés à chaque lien sont encodés dans une matrice d'adjacence \mathbf{W}^{CF} telle que $\mathbf{W}_{ij}^{CF} = \mathbf{W}_{ji}^{CF} = CF_{ij}$ correspondent à la mesure de CF entre les signaux i et j .

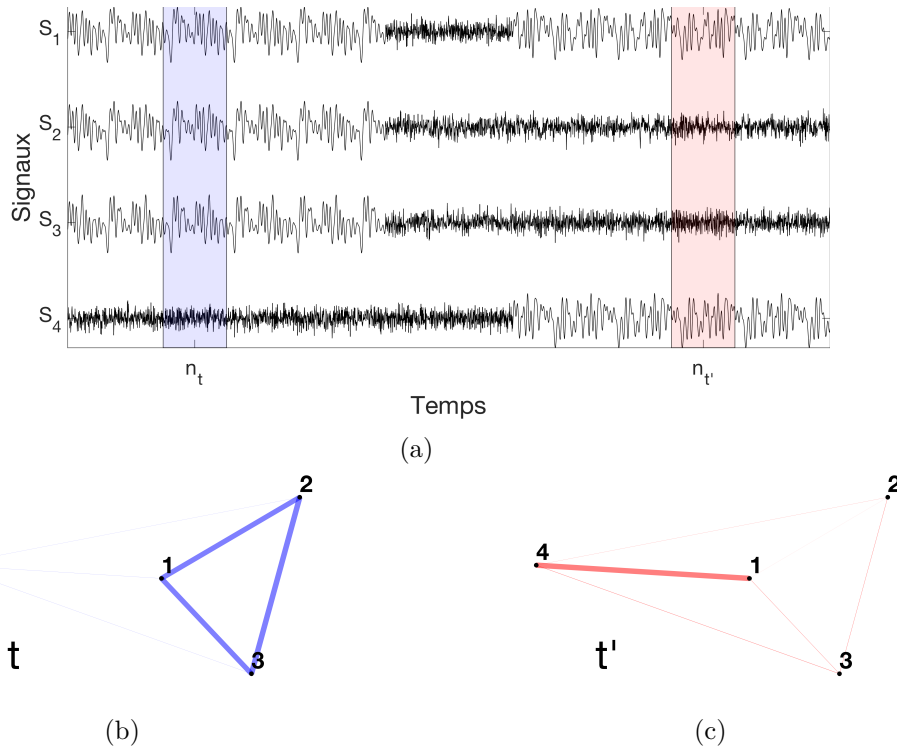


FIGURE 1.4 – (a) - 4 signaux suivant un modèle simple de corrélation, (b) & (c) - structure de graphe pour deux segments centrés autour des échantillons n_t et $n_{t'}$ (la largeur des liens est proportionnelle à la mesure de CF).

Comme ce graphe évolue avec les segments de temps considérés, on obtient une structure de graphe dynamique :

Definition 3 : Graphe/réseau dynamique (de CFs)

Le graphe dynamique de CF est une succession de graphes de CF à différents instants t , noté $\mathcal{G}(t)$, $\forall t \in \{1, \dots, T\}$. Chaque graphe est défini par un triplet $(\mathcal{V}(t), \mathcal{E}(t), \mathbf{W}^{CF}(t))$. $\mathcal{V}(t) = \mathcal{V} \forall t$ car la position et le nombre d'électrodes ne varie pas au cours du temps, de même $\mathcal{E}(t) = \mathcal{E} \forall t$. La seule variation se situe au niveau du poids des liens, c'est-à-dire les matrices d'adjacence $\mathbf{W}^{CF}(t)$ encodées à partir des mesures dynamiques de CF.

Ainsi une première représentation mathématique des mesures dynamiques de CFs, mettant en avant la structure de réseau dynamique sous-jacente, est une série de matrices d'adjacences $\mathbf{W}^{CF}(t) \forall t \in \{1, \dots, T\}$. Notons que des substituts à la matrice d'adjacence peuvent être utilisés, comme la matrice Laplacienne ou de covariance, selon les propriétés algébriques désirées pour traiter les données de CFs.

2) La matrice de CFs :

Le graphe dynamique, caractérisé par une succession de matrices d'adjacence, peut aussi être traduit par une matrice de CF qui est défini ainsi :

Definition 4 : Matrice de CFs

Notée $\mathbf{X} \in \mathbb{R}^{L \times T}$, la matrice de CF contient dans sa colonne t les éléments supérieurs diagonaux de la matrice d'adjacence $\mathbf{W}^{CF}(t)$ (on rappelle que L correspond au nombre de paires de signaux).

Le schéma de la Figure 1.5 illustre le lien entre les deux représentation du réseau dynamique de CF.

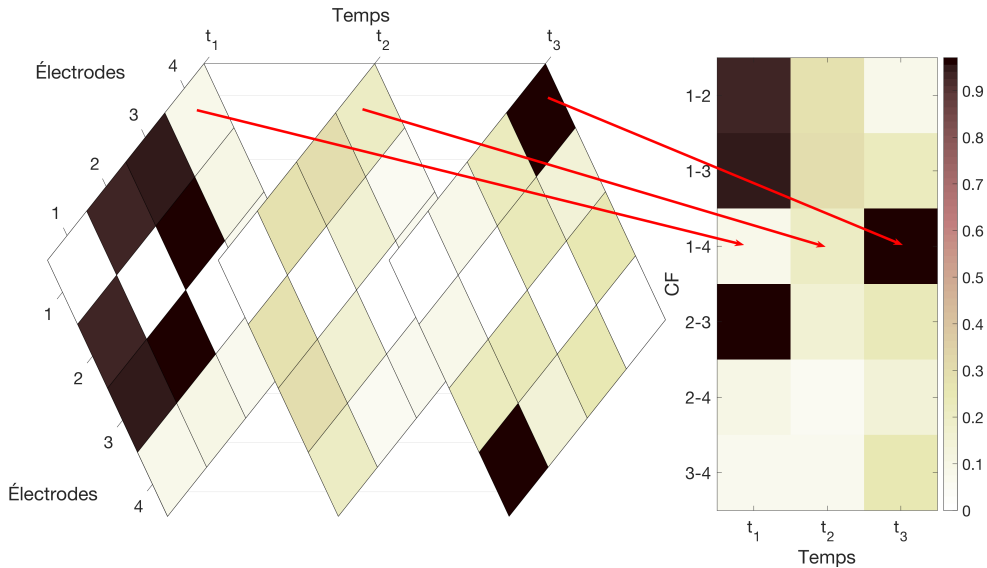


FIGURE 1.5 – (à gauche) Graphe dynamique de CFs caractérisé par une succession de matrices d’adjacence à trois instants, (à droite) matrice de CF \mathbf{X} contenant les éléments supérieurs diagonaux des matrices d’adjacence

Un exemple de matrice de CFs pour les signaux de la Figure 1.4 est proposé dans la Figure 1.6. les zones claires correspondent à des connectivités faibles, et les zones foncées à des connectivités fortes. On remarque que l’on retrouve un découpage en trois étapes, ainsi que les relations de dépendance des signaux initiaux qui étaient visibles en examinant le tracé temporel des signaux.

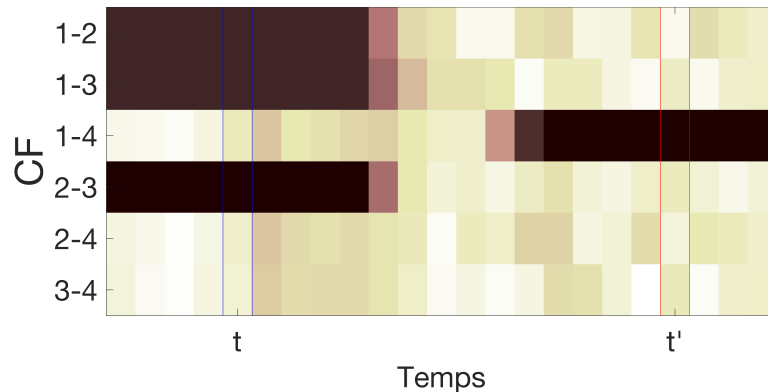


FIGURE 1.6 – Mesure dynamique des CFs pour chaque paire de signaux de l’exemple fourni dans la Figure 1.4, les zones foncées correspondent à des mesures de connectivités élevées

3) L’intérêt de différentes représentations :

Chaque représentation peut être plus ou moins intéressante selon l’objectif souhaité. Pour pouvoir bien caractériser les différences entre toutes les représentations, il est important de préciser les notions de modes et de dimensions :

Definition 5 : Mode ou modalité

Les modes sont les différentes variables caractéristiques d’un jeu de données.

Exemple : La matrice de CFs $\mathbf{X} \in \mathbb{R}^{L \times T}$ possède deux modes, le premier est associé à la mesure de CFs, le second au temps.

Definition 6 : Dimension

La dimension est le nombre de composantes dans chacune des modalités.

Exemple : Le jeu de données $\mathcal{X} \in \mathbb{R}^{60 \times 100 \times 8}$ correspond à un cube de données, il contient donc trois modes. Les dimensions de chaque mode sont respectivement 60, 100 et 8.

Ainsi, la succession de matrices d'adjacence est composée de 3 modes qui sont respectivement Électrodes \times Électrodes \times Temps, tandis que la matrice de FC n'en a que deux qui sont CF \times Temps. On remarque que cette dernière représentation permet de réduire le nombre de modes, mais aussi le nombre d'éléments qu'elle contient. En fonction de l'objectif et des outils d'analyse utilisés, une représentation peut s'avérer plus pertinente qu'une autre. En effet, la première représentation sera plus efficace pour trouver des communautés d'électrodes avec des comportements similaires tandis que la seconde représentation sera préférable pour regrouper des CFs qui s'activent de façon synchrone.

1.2.3 Principales mesures de connectivité fonctionnelle

Il existe plusieurs mesures de CF utilisées dans l'ensemble des applications [192]. En fonction du jeu de données (fMRI, MEG, EEG etc..) et de la tâche cognitive, différentes mesures de CF seront privilégiées [52]. Cependant, on peut noter l'utilisation fréquente des mesures présentées ci-dessous.

1) La corrélation de Pearson :

La corrélation de Pearson [34, 168] est la mesure la plus célèbre pour montrer la dépendance statistique entre deux signaux s_1 et s_2 stationnaires. En considérant les signaux centrés cette mesure s'écrit :

$$\rho = \frac{\mathbb{E}\{s_1 s_2\}}{\sqrt{\mathbb{E}\{s_1^2\}\mathbb{E}\{s_2^2\}}}, \quad (1.2)$$

Plusieurs extensions de cette mesure ont été proposées comme mesure de CF, on peut noter : l'utilisation de la corrélation croisée afin de prendre en considération des retards éventuels entre les deux signaux [78]. Ou encore, l'utilisation de versions non-linéaires de la corrélation de Pearson si on ne souhaite pas réduire les interactions entre les signaux à leurs seules composantes linéaires [198]. On peut aussi déterminer la corrélation entre deux signaux après avoir éliminé les dépendances conditionnelles dues aux autres signaux. Ainsi, on obtient une nouvelle mesure appelée corrélation partielle [201] (que l'on étudiera plus en détail en Section 2.3).

2) Module de la cohérence spectrale :

Le module de la cohérence spectrale [153] est l'équivalent de la corrélation dans le domaine fréquentiel. Elle est mesurée en calculant dans un premier temps, la densité spectrale d'énergie croisée entre les signaux s_1 et s_2 , noté $D_{s_1 s_2}(f)$ et qui s'écrit :

$$D_{s_1 s_2}(f) = \sum_{\tau=-\infty}^{+\infty} \mathbb{E}\{s_1 s_2^{(\tau)}\} e^{-i2\pi f \tau}, \quad (1.3)$$

avec $s_2^{(\tau)}(n) = s_2(n - \tau) \forall n \in \{1, \dots, N\}$, le signal s_2 retardé de τ échantillons. La variable i désigne le nombre imaginaire $i = \sqrt{-1}$. Ainsi le module de la cohérence spectrale correspond à :

$$\text{Coh}(f) = \frac{|D_{s_1 s_2}(f)|}{\sqrt{D_{s_1 s_1}(f)D_{s_2 s_2}(f)}}. \quad (1.4)$$

Cette mesure est particulièrement intéressante quand on sait que le phénomène cognitif d'intérêt est associé à des oscillations dans des fréquences, ou dans une bande spectrale étroite particulière. Une mesure similaire très utilisée, nommée partie imaginaire de la cohérence spectrale [151], consiste (comme son nom l'indique) à remplacer dans l'Équation (1.4) le module de $D_{s_1 s_2}(f)$ par sa partie imaginaire. Il a été démontré que cette mesure permettait de limiter le problème de la conduction volumique [151] (les électrodes peuvent mesurer l'activité de sources spatialement éloignées, introduisant de fausses connexions).

3) La phase-locking-value :

Comme la phase-locking-value (PLV) [124, 13] sera la mesure de CF utilisée principalement dans ce manuscrit, une attention particulière lui est accordée. Afin de comprendre cette mesure de CF, il est important de rappeler que tout signal réel $s(n)$ peut se décomposer de la façon suivante :

$$s(n) = a(n) \cos(\phi(n)), \quad (1.5)$$

où $a(n)$ et $\phi(n)$ correspondent respectivement à l'amplitude et à la phase instantanée du signal $s(n)$ [162]. Il existe une infinité de combinaisons possibles de fonctions $a(n)$ et $\phi(n)$ respectant l'Équation (1.5). Cependant, pour que $a(n)$ soit une amplitude, on impose que cette fonction soit non négative ($a(n) \geq 0$), et que $a(n)$ contienne des fréquences plus basses que la fonction $\cos(\phi(n))$. Pour résoudre ce problème, on peut utiliser la représentation analytique du signal $s(n)$. Notons que $s(n)$ est une fonction réelle qui, par symétrie hermitienne de sa transformée de Fourier, implique une redondance du contenu fréquentiel. L'idée de la représentation analytique est donc de considérer la fonction $s(n)$, après avoir éliminé les fréquences négatives. Il s'agit alors d'une fonction complexe notée $z(n)$ qui peut être écrite sous forme polaire :

$$z(n) = a(n)e^{i\phi(n)}. \quad (1.6)$$

On obtient ainsi l'amplitude et la phase instantanée de $s(n)$ par application du théorème de Bedrosian [32]. Alors, la représentation analytique peut aussi être écrite :

$$z(n) = s(n) + i\mathcal{H}[s](n), \quad (1.7)$$

$$= a(n) \cos(\phi(n)) + i a(n) \sin(\phi(n)), \quad (1.8)$$

où $\mathcal{H}[\bullet]$ est la transformée de Hilbert [120]. Ainsi, pour calculer la fonction $z(n)$, il suffit de calculer la transformée de Hilbert de $s(n)$. Notons que cette méthode pour calculer la phase et l'amplitude instantanée d'un signal produit à coup sûr une amplitude non-négative, cependant il a été prouvé que le contenu fréquentiel de la fonction $a(n)$ sera borné par celui de $\cos(\phi(n))$ uniquement dans le cas où la phase est une fonction linéaire [32], [162]. Malgré tout, dans un contexte plus général, la décomposition obtenue reste pertinente. Nous illustrons cette décomposition à l'aide de la Figure 1.7. Dans la Figure 1.7a, la représentation d'une fonction cosinus modulée par une enveloppe quadratique. Dans la Figure 1.7b son amplitude est non-négative et de basse fréquence. Enfin, dans la Figure 1.7c sa phase instantanée est représentée, elle est ici linéaire (modulo 2π). Ces fonctions peuvent s'écrire :

$$s(n) = 4[1 - (n - 1)^2] \cos(8\pi n), \quad a(n) = 4[1 - (n - 1)^2], \quad \phi(n) = 8\pi n. \quad (1.9)$$

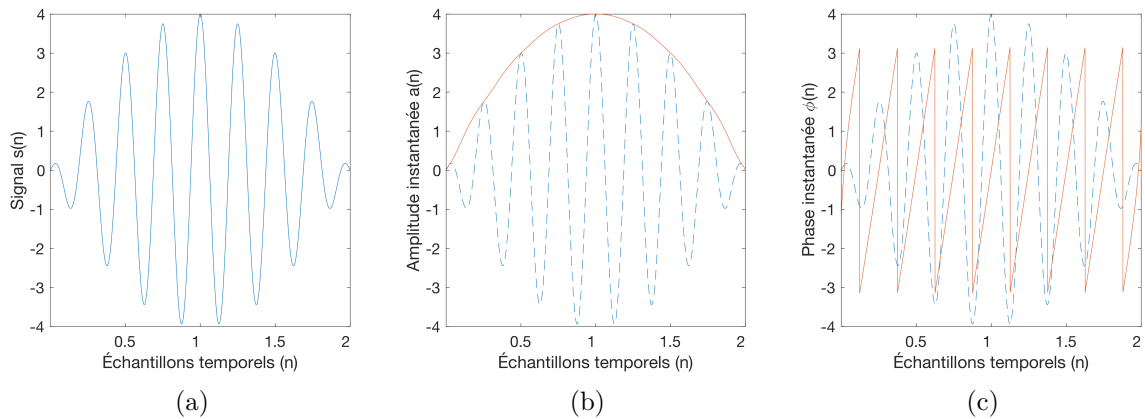


FIGURE 1.7 – (a) signal $s(n)$, (b) amplitude instantanée de $s(n)$ en rouge, (c) phase instantanée de $s(n)$ en rouge

Ainsi la Phase-Locking-Value (PLV) [124, 13] entre deux signaux $s_1 = a_1 \cos(\phi_1)$, $s_2 = a_2 \cos(\phi_2)$, mesure à quel point les phases des deux signaux sont synchrones. Cette mesure quantifie la constance de la différence entre les phases des deux signaux au cours du temps. Comme la phase a un support circulaire, la PLV s'écrit :

$$\text{PLV} = |\mathbb{E}\{e^{i(\phi_1 - \phi_2)}\}|. \quad (1.10)$$

L'intérêt de la PLV, comparé à la corrélation de Pearson, est de ne pas considérer les amplitudes des signaux dans la mesure de synchronie. C'est important pour les signaux SEEG où de fortes disparités d'amplitudes peuvent apparaître en fonction de la distance de l'électrode à l'amas de neurones sources. De plus, le montage bipolaire utilisé pour enregistrer les signaux SEEG (voir l'Annexe 1.A.1) rend la notion d'amplitude difficile à interpréter, étant donné que l'on mesure la différence de potentiel entre deux électrodes adjacentes. On illustre l'avantage d'utiliser la PLV pour mesurer la synchronie de phase en considérant deux signaux SEEG réels avec des phases synchrones. On représente dans la Figure 1.8 la phase instantanée des deux signaux. Afin de mettre en avant le caractère synchrone des deux phases, on marque d'un trait rouge tous les instants où le changement de signe de la phase ϕ_2 implique un changement de signe de la phase ϕ_1 moins de 0.05 secondes après (la durée du signal représenté est de 11 s). Le tracé du bas de la Figure 1.9 représente les deux signaux SEEG dont les phases sont représentées dans le tracé du haut de la Figure 1.8. En haut, onze mesures de la PLV et de la corrélation de Pearson sont calculées. On remarque qu'utiliser la corrélation comme mesure de CF ne permet pas de montrer la synchronie entre les deux signaux (valeurs inférieures à 0.4), contrairement à la mesure de PLV qui monte jusqu'à 0.8.

Comme pour les autres mesures, des extensions de la PLV existent notamment pour répondre au problème de la conduction volumique tel que la mesure Phase-lag-index [180]. Cependant, comme les enregistrements SEEG avec montage bipolaire ne sont pas trop soumis à ce problème, nous préférons l'utilisation de la PLV, le Phase-lag-index devenant dans ce cas trop sélectif.

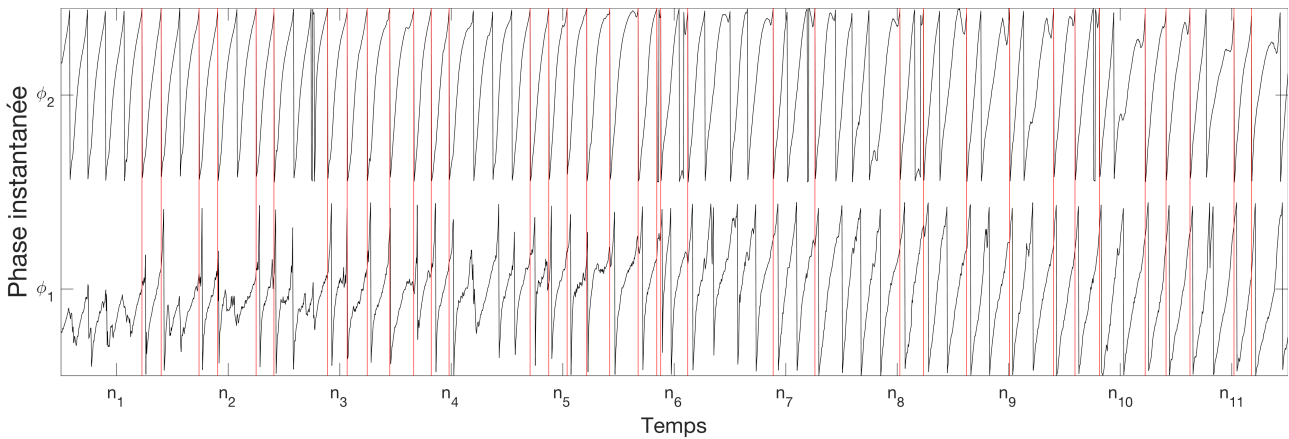


FIGURE 1.8 – Phase instantanée de deux signaux SEEG réels, on marque d'un trait rouge tous les instants où le changement de signe de la phase ϕ_2 implique un changement de signe de la phase ϕ_1 dans les 0.05 secondes qui suivent (la durée du signal représenté est de 11 s).

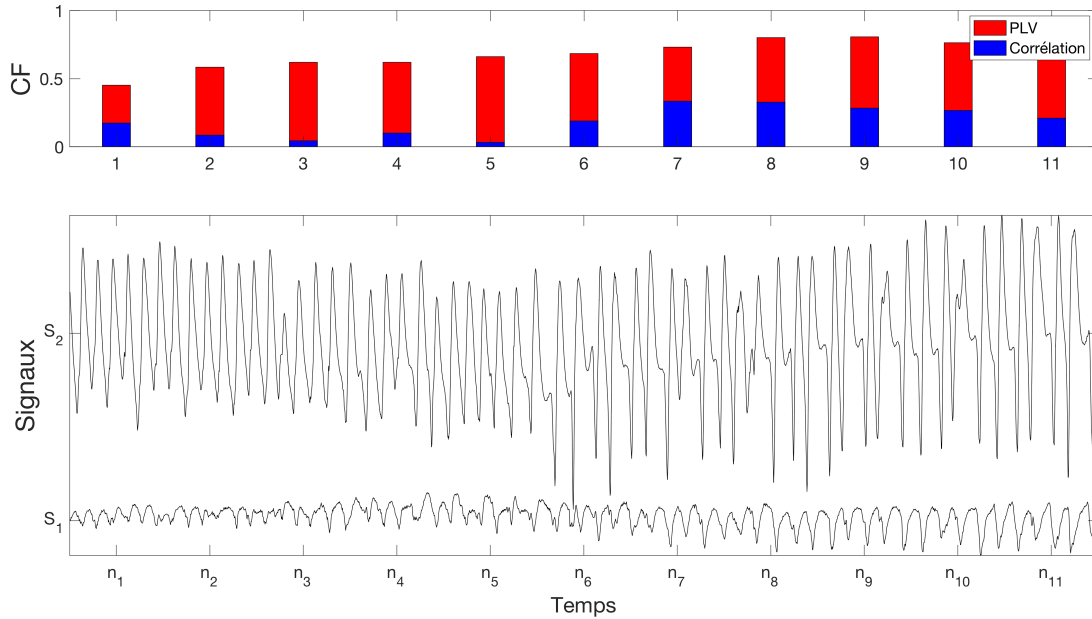


FIGURE 1.9 – (tracé du bas) Les 2 signaux s_1 et s_2 provenant de l’enregistrement réel SEEG, (tracé du haut) mesures de CF en utilisant la PLV (en rouge) et la corrélation de Pearson (en bleu).

4) Corrélation des amplitudes (AEC) :

Une autre mesure, de plus en plus utilisée [167], [160], [156], [159], appelée corrélation des amplitudes (ou des enveloppes) et notée AEC, considère au contraire la corrélation entre les amplitudes des signaux. On a vu que la PLV mesure la synchronie des phases. Cependant, il semble que dans certaines situations, elle soit une mesure trop sélective. Alors, après filtrage du signal dans une bande de fréquences d’intérêt, l’AEC semble être une bonne alternative pour mesurer la similarité entre les signaux sans avancer d’hypothèses physiologiques strictes sur leur comportement. Cette mesure se note :

$$\text{AEC} = \frac{\mathbb{E}\{(a_1 - \mathbb{E}\{a_1\})(a_2 - \mathbb{E}\{a_2\})\}}{\sqrt{\mathbb{E}\{(a_1 - \mathbb{E}\{a_1\})^2\}\mathbb{E}\{(a_2 - \mathbb{E}\{a_2\})^2\}}}, \quad (1.11)$$

5) Autres mesures :

Les mesures proposées ci-dessus correspondent aux approches les plus couramment utilisées. Deux autres grandes familles de CF existent, elles ne sont pas détaillées ici car elles ne réapparaîtront pas dans la suite du manuscrit, cependant elles restent intéressantes à mentionner. La première concerne les mesures de CF utilisant la théorie de l’information [171] où l’on associe à chaque signal des densités de probabilité particulière, permettant de considérer des relations non-linéaires entre les signaux. La seconde concerne les mesures de causalité de Granger, utilisant des modèles autorégressifs afin de déduire des relations causales entre les signaux. Cette mesure de CF est particulièrement intéressante si l’on souhaite inférer des graphes de CF dirigés [40].

Conclusion : On a vu dans cette section une introduction à la mesure dynamique des CFs. Il est important, dans un premier temps, d’effectuer un prétraitement des données pour éviter de fausses connexions. Ensuite, il faut choisir une mesure de CF pertinente selon le problème étudié. Cependant ces mesures sont souvent difficiles à analyser pour les cliniciens, principalement à cause de leurs nombres. Comme le nombre L de mesure de CFs est quadratique au nombre de signaux, si l’on considère 100 signaux on doit analyser environ 5000 CF, cela demande une grande expertise et est très chronophage. C’est pour cela que dans le cadre de ce manuscrit, on propose de simplifier l’étude du réseau dynamique obtenu à partir des mesures de CF, ce problème n’est pas nouveau et a déjà fait l’objet de nombreuses études qui sont présentées dans la partie suivante.

1.3 Décomposition et analyse des réseaux de connectivité fonctionnelle

La création d'outils afin d'analyser les mesures de CFs est un domaine en plein essor qui a déjà fourni de nombreux résultats. On peut décomposer ces études en trois parties : la première concerne les graphes de CFs statiques, c'est-à-dire quand l'ensemble de la série temporelle est réduite en un unique graphe de CF, ces méthodes sont présentées dans la Section 1.3.1. Ensuite, on se concentre sur les extensions des méthodes statiques dans le cadre de mesures évoluant dans le temps en Section 1.3.2. Enfin, en Section 1.3.3, on présente les méthodes récentes de décompositions de graphes de CFs [116] qui permettent d'extraire simultanément des facteurs de CFs et des profils d'activation temporelle. Ces dernières méthodes ont l'avantage de pouvoir s'étendre selon d'autres modalités comme la fréquence, les essais ou les sujets. Les études présentées ne concernent pas uniquement les signaux SEEG de patients épileptiques, car de nombreux outils intéressants ont été proposés pour d'autres pathologies (Schizophrénie, alzheimer, ...) et d'autres méthodes de mesure (MEG, fMIR), elles peuvent cependant être adaptées à notre contexte d'étude.

1.3.1 Analyse statique des CFs

De nombreuses applications pertinentes ont été traitées à partir de l'étude d'un graphe unique de CFs réalisé sur l'ensemble d'un enregistrement. Dans [178], les auteurs montrent que les patients ayant reçu des blessures de guerre (victimes d'une explosion) ont une activation moyenne des CFs différentes d'un groupe standard. De même, dans [7], les auteurs montrent que l'influence d'un médicament (Aprazolam) peut être remarqué par la moyenne des CFs calculées. D'autres études prennent en considération la structure de graphe inférée par les CFs, des métriques sont alors généralement utilisées, comme l'efficacité globale et locale, le degré moyen ou le coefficient de partitionnement moyen [41]. À partir de ces métriques, les auteurs de [168] montrent que des patients atteints de la maladie d'Alzheimer ont des structures de graphe de CFs différentes d'un groupe standard, ils montrent notamment que les réseaux inférés n'ont plus la topologie petit monde (*Small-World*) généralement observée [22]. Ces métriques de graphes ont aussi été utilisées dans [110] pour montrer l'influence de l'activité physique sur la topologie des réseaux de CFs.

Les graphes de CFs ne permettent pas seulement de faire de la classification, ils peuvent aussi être utilisés comme outil pour débruiter les signaux du cerveau. On notera que ce sont plus souvent des graphes de connectivité structurelle qui sont utilisés dans ce cas [98]. Dans [144], les auteurs montrent comment l'étude simultanée du graphe de CF et des signaux (qui vivent sur ces nœuds) est importante pour comprendre l'apprentissage d'une tâche. Comme montré dans [145], ces graphes de CFs peuvent aussi être utilisés comme outil de réduction dimensionnelle des signaux du cerveau.

Cependant, l'étude des graphes de CFs statiques est difficilement applicable aux données de crises d'épilepsie, car l'évolution temporelle des graphes est essentielle pour modéliser ce phénomène.

1.3.2 Le problème de la multimodalité

Dans le cas de l'épilepsie, qui est un phénomène évoluant dans le temps, le découpage de la modalité temporelle est cruciale pour visualiser la propagation pathologique de la crise. Une manière simple de considérer le temps, est d'utiliser des métriques sur chaque graphe de CF dans le but d'observer des variations temporelles de leur topologie. [112] propose par exemple d'utiliser le premier vecteur propre de chaque graphe de CF, le produit scalaire entre ces vecteurs et le vecteur propre moyenné sur les graphes forment une mesure permettant de séparer les états ictaux, pré-ictaux et normaux de données SEEG provenant de patients épileptiques. De manière similaire, pour montrer que la plupart des crises d'un même patient peuvent être découpées en un nombre fini d'états, et que la progression de ces états est cohérente d'une crise à l'autre, [42] détermine le taux de connexion de chaque électrode en fonction du temps, grâce à la mesure de centralité des vecteurs propres des graphes de CF [148]. L'étude distincte de ces graphes de CF ne se fait pas uniquement par l'intermédiaire de méthodes spectrales. Dans [115], les auteurs décomposent chaque graphe de CF en 3 sous-graphes qui maximisent la modularité du réseau [149]. L'étude de la similarité des communautés obtenues au cours du temps montre qu'une crise

d'épilepsie peut être découpée en trois états : initialisation, propagation, terminaison. Les proportions de connections faibles et fortes des sous-graphes semblent être un bon indicateur de l'état de la crise.

Cependant, d'autres modalités peuvent aussi être considérées en plus du temps. Certaines études examinent ainsi les CFs dynamiques obtenues selon les différentes crises d'un même patient ou pour différentes bandes de fréquences. Ainsi, toujours dans le contexte de l'épilepsie afin de montrer des facteurs pré-ictaux spécifiques, les auteurs de [127] calculent les CFs en fonction du temps et de différentes plages de fréquences. Pour diminuer la grande dimensionnalité des données, ils conservent uniquement les bandes de fréquences et les CFs qui semblent le mieux discriminer les états pré-ictaux des états ictaux. Finalement, ces mesures réduites sont comparées à des états de référence pour observer des facteurs pré-ictaux particuliers, qui peuvent survenir parfois plusieurs heures avant une crise.

Notons que ces méthodes ne sont finalement pas si différentes de l'étude des graphes statiques. Cependant en plus de devoir trouver une mesure caractérisant chaque graphe de CF, il faut savoir analyser l'évolution temporelle des scores proposés. Cela devient encore plus complexe lorsque plusieurs modalités sont utilisées comme dans [127]. Face à ces difficultés, des méthodes de décomposition de graphes dynamiques (et multimodales) ont été proposées [116]. Ces méthodes ont l'avantage de décomposer en même temps l'enregistrement en graphes de CFs caractéristiques et en états temporels distincts.

1.3.3 Analyse dynamique et multimodale des CFs

Une première approche pour réaliser la décomposition d'un graphe dynamique se base sur l'utilisation de la matrice de CFs. Ainsi une décomposition matricielle peut être réalisée pour directement extraire des facteurs de CFs avec leurs profils d'activation temporelle. Ainsi une factorisation en matrices non négatives (NMF) est appliquée dans [114] afin de localiser les régions du cerveau qui génèrent une crise d'épilepsie. Une régularisation parcimonieuse sur l'axe temporel est ajoutée dans [44] afin d'avoir des profils d'activation qui discriminent l'évolution des graphes de CFs au cours du développement du cerveau.

Cependant, comme il peut y avoir plusieurs modalités (en plus de la modalité temporelle qui est spécifique), on peut aller au-delà de la représentation matricielle et utiliser des structures plus générales appelées tenseurs [54]. Par exemple, afin de conserver la structure de graphe des CFs, on peut préférer étudier un tenseur en 3 dimensions constitué des matrices d'adjacences (ou laplacienne) à chaque instant au lieu d'étudier la matrice de CF. Dans ce cas, le tenseur sera composé de 3 modes : Electrodes \times Electrodes \times Temps. Dans [193], un tenseur de ce type est décomposé à l'aide d'une décomposition en valeurs singulières d'ordre supérieur (HOSVD) permettant de généraliser les méthodes spectrales de [42], [112], et de traitement des signaux sur graphe de [145], [144] au contexte des graphes dynamiques. Une méthode similaire a été appliquée dans [129], où un enregistrement est réalisé sur des sujets qui regardent une vidéo ou restent inactifs. À partir d'un tenseur dont les modes sont CF \times Temps \times Sujets, ils réussissent à identifier des vecteurs singuliers discriminant les deux tâches. La HOSVD peut être utilisée sur des tenseurs d'ordre supérieur, comme dans [139], où le graphe de CF évolue selon le temps et les sujets, produisant un tenseur dont les modes sont Electrodes \times Electrodes \times Temps \times Sujets. Dans cette étude, la HOSVD est à la fois utilisée pour découper l'axe temporel en plusieurs états et pour associer à chaque état un graphe caractéristique. La HOSVD n'est pas la seule décomposition utilisée dans ce contexte. Dans [158] il s'agit d'une décomposition en composantes principales robustes qui est appliquée afin d'éliminer un bruit de mesure parcimonieux dans les données de CFs. Finalement dans [38], une extension du principe de modularité pour les graphes dynamiques [143] est proposée afin de modéliser la reconfiguration du réseau cortical lors de tâches difficiles. Cette méthode permet donc de trouver des sous-structures de réseaux dynamiques préservant une forte modularité selon l'axe temporel.

On peut noter que l'application de méthodes tensorielles sur les données de CFs reste assez nouvelle, surtout dans le contexte de l'épilepsie. Cependant, ces méthodes ont déjà été de nombreuses fois appliquées directement sur les signaux EEG enregistrés, comme le montre la revue [100] appliquée à l'épilepsie. Dans un contexte plus général, l'utilisation de tenseurs pour traiter directement les signaux du cerveau est répandue [57], [48] et peut permettre à la fois de filtrer les artefacts [30], faire de la localisation de sources [29], de l'inférence de CFs [179] ou de traiter simultanément des données EEG

Référence	Sous-section	Enregistrement	CF utilisée	Méthode d'analyse
[178]	1.3.1	EEG	PLV	Test de Wilkoksum
[7]	1.3.1	EEG	Information mutuelle	Test de Wilkoksum
[168]	1.3.1	EEG	Corrélation non-linéaire	Network metric [41]
[110]	1.3.1	EEG	PLV	Network metric [41]
[144]	1.3.1	fMRI	Cohérence	Annotated graph
[145]	1.3.1	fMRI	Corrélation + autres	Réduction dimensionnelle
[112]	1.3.2	SEEG	Cohérence	Valeurs propres
[42]	1.3.2	ECoG	Cohérence	Centralité [148]
[115]	1.3.2	ECoG	Corss-corrélation	Modularité [149]
[127]	1.3.2	SEEG	PLV	Network metric
[114]	1.3.3	ECoG	Cross-correlation	NMF
[44]	1.3.3	fMRI	cohérence	NMF
[193]	1.3.3	EEG	PLV	HOSVD
[129]	1.3.3	fMRI	Correlation	HOSVD
[139]	1.3.3	EEG	PLV	HOSVD
[158]	1.3.3	EEG	PLV	Robust HOPCA
[38]	1.3.3	fMRI	cohérence	Modularité dynamique [143]

Tableau 1.1 – Récapitulatifs des travaux présentés dans la Section 1.3

et fMRI [31], [57].

Les méthodes d'analyse des CFs sont variées, et soulèvent des problèmes mathématiques et numériques intéressants, surtout lorsque l'on doit prendre en compte plusieurs modalités. Cependant, l'application de décompositions multimodales de graphes dynamiques de CFs dans le contexte de l'épilepsie est un travail qui n'a pas encore été réalisé. De plus, peu d'études utilisent des contraintes de parcimonie sur les facteurs de CFs obtenus, ce qui semble essentiel dans le contexte de l'épilepsie où beaucoup de CF peuvent être passivement impliquées dans l'étape de propagation de la crise, produisant des graphes complexes à étudier.

1.4 Conclusion

Dans cette partie, le contexte de la thèse a été présenté. Dans un premier temps, nous avons montré la provenance des données étudiées en Section 1.1, ensuite les deux étapes clés de l'étude des réseaux dynamiques dans le contexte de l'épilepsie ont été données : la Section 1.2 montre les techniques d'inférence de réseaux dynamiques grâce aux mesures de CFs et la Section 1.3 présente les méthodes d'analyses et de décomposition de ces réseaux. Les contributions de cette thèse se situent dans chacune de ces deux étapes. Les chapitres II et III se concentrent sur l'élaboration de mesures de CFs fines permettant de dévoiler l'évolution dynamique de l'état cortical du cerveau lors d'une crise d'épilepsie. Les chapitres IV à VII présentent quant à eux de nouvelles méthodes tensorielles d'analyse de réseaux dynamiques.

Annexe

1.A Prétraitement

Comme les signaux SEEG proviennent d'électrodes intracrâniennes, les mesures réalisées sont moins bruitées (comparées à un enregistrement EEG), ainsi le nombre de prétraitements nécessaires avant l'inférence de CF reste limité (et restreint à cinq) :

1.A.1 Montage bipolaire :

Rappelons que l'activité du cerveau est enregistrée via des électrodes positionnées sur des tiges implantées en profondeur dans le cerveau. Le montage référentiel consiste à enregistrer la différence de potentiel entre chaque électrode d'une même tige et d'une électrode de référence. Cependant ce montage peut produire des mesures redondantes, de plus l'activité de la référence peut biaiser la mesure. On utilise alors un montage bipolaire, la différence de potentiel est calculée entre deux électrodes adjacentes. Le schéma d'un montage bipolaire pour deux tiges est présenté dans la Figure 1.A.1, les signaux formant notre jeu de données sont notés $s_i(n)$ et les potentiels $e_i(n)$. Cette configuration a l'avantage de privilégier l'activité locale à l'électrode. Notons que l'un des désavantages de cette méthode est de perdre le signal d'une électrode par tige, en effet comme représenté sur la figure, 4 potentiels fournissent 3 signaux. On précise que dans le reste du manuscrit, on emploie toujours le terme d'électrodes pour désigner les signaux obtenus grâce aux mesures SEEG bipolaire.

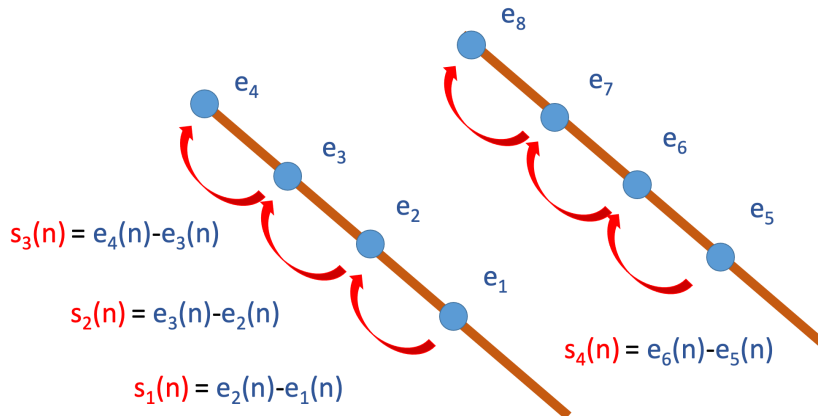


FIGURE 1.A.1 – Schéma d'un montage bipolaire pour deux tiges, les signaux formant notre jeu de données sont notés $s_i(n)$ et le potentiel des électrodes $e_j(n)$

1.A.2 Élimination d'électrodes :

Certaines électrodes peuvent se révéler défectueuses, produisant un signal inexploitable noyé dans le bruit. L'élimination de ces électrodes est une tâche manuelle, suite à la visualisation des signaux. Afin d'être aidé dans cette décision, on utilise un score simple se basant sur la densité spectrale de puissance de chaque signal (noté $D_s(f)$ pour le signal s et pour la fréquence f). En effet, on sait que le contenu

fréquentiel pour les hautes fréquences (soit f supérieure à 0.25 en fréquence normalisée) doit être faible, la densité spectrale de puissance étant généralement une fonction en $1/f$. Le score correspond donc à :

$$score = \int_{f \geq 0.25}^{0.5} D_s(f) df. \quad (1.12)$$

La Figure 1.A.2a représente le résultat de ce score pour 50 électrodes d'un enregistrement SEEG réel. On s'aperçoit que 3 électrodes (la 44, 45 et 46) ont un score particulièrement haut. La Figure 1.A.2b montre les formes de ces 3 signaux, ainsi que de deux autres choisis au hasard afin d'illustrer leur caractère bruité et non exploitable. Ces trois signaux sont donc éliminés.

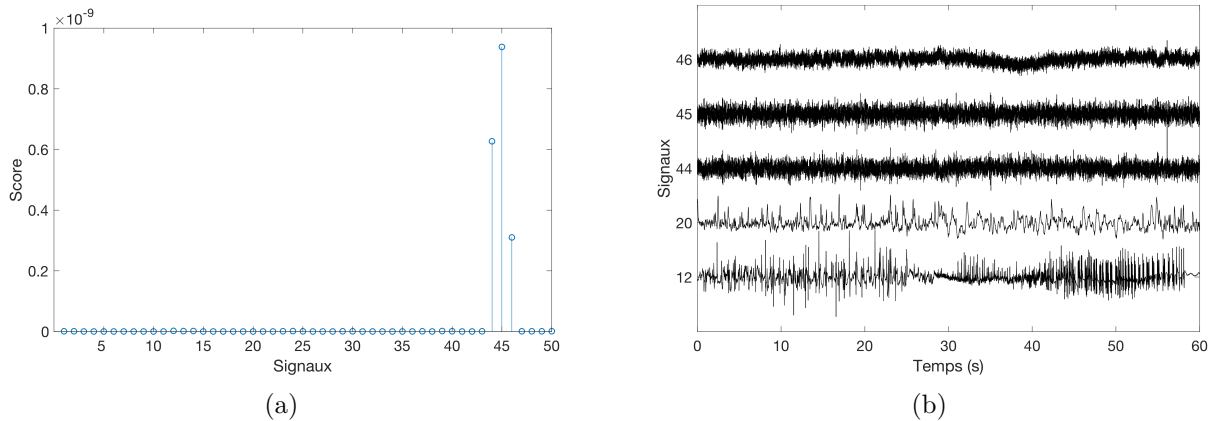


FIGURE 1.A.2 – (a) Score pour 50 signaux d'un enregistrement SEEG réels - (b) Forme des signaux 44, 45 et 46 comparée à celle des signaux exploitables.

1.A.3 Problème des fréquences à 50 Hz :

Les électrodes utilisées sont affectées par la pollution électromagnétique à 50 Hz, il s'agit de la fréquence d'oscillation du courant alternatif en France. La Figure 1.A.3a montre en fonction de la fréquence, la somme des densités spectrales de puissance de l'ensemble des signaux d'un enregistrement SEEG réels. On observe bien un pic à 50 Hz. Un filtre Notch, qui est un filtre coupe-bande sélectif, est utilisé pour éliminer le contenu de la fréquence à 50 Hz, on peut observer le résultat de ce filtrage dans la Figure 1.A.3b.

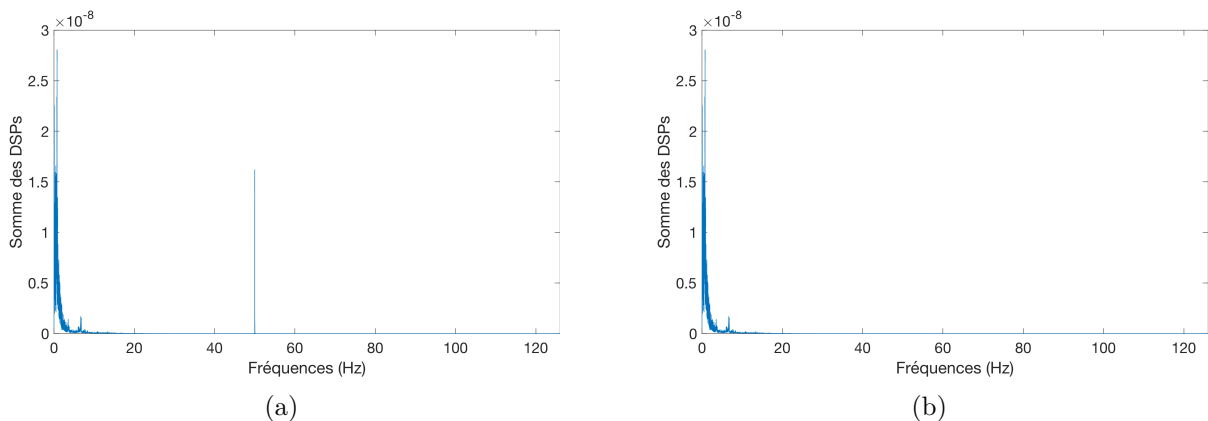


FIGURE 1.A.3 – (a) Somme des densités spectrales de puissance de l'ensemble des signaux d'un enregistrement SEEG - (b) Application d'un filtre notch pour éliminer la fréquence à 50 Hz.

1.A.4 Encadrement de la crise d'épilepsie :

Comme présenté dans la Figure 1.2e, une crise d'épilepsie peut être décomposée au moins en trois étapes. Afin de sélectionner des signaux contenant toutes les étapes de la crise, mais aussi une partie

du signal pre-ictal (qui est intéressant à prendre en compte pour exposer les particularités du début de crise), on propose le découpage suivant : des marqueurs indiquant le début de la crise ont été placés au préalable par les cliniciens. Puis on démarre notre jeu de données afin d'intégrer entre 30 et 50 secondes d'événements pré-ictaux : on le termine environ 10 à 30 secondes après la fin de la crise. Chez certains patients, les crises ont des durées et des comportements très similaires, il est alors possible de les représenter toutes avec le même nombre d'échantillons afin d'obtenir un jeu de données avec plusieurs crises.

1.A.5 Filtrage

Pour les signaux neuro-physiologiques, on sait que différentes fréquences portent différentes informations. Il est alors préférable de connaître la bande passante contenant le processus d'intérêt. Pour nous aider à choisir ces bandes, on peut se référer aux groupements usuels des bandes de fréquences dans le cerveau : delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz) et gamma (> 30 Hz). Malheureusement, les événements ictaux apparaissent généralement dans toutes les bandes [84] avec cependant des motifs caractéristiques généralement situés dans la bande gamma [14]. Comme les fréquences d'intérêt sont souvent spécifiques à chaque patient, on étudie les crises sur une bande large (20-100 Hz).

Il n'est pas nécessaire de réaliser d'autres prétraitements, en effet, il est courant pour des signaux EEG ou fMRI de supprimer des artefacts dus aux battements du cœur et aux clignements des yeux à l'aide d'une analyse en composantes indépendantes par exemple [18]. Cependant, les mesures locales de bonnes qualités réalisées par les électrodes SEEG ne comportent pas ce genre d'artefacts.

Chapitre 2

Connectivités fonctionnelles sous hypothèse de données gaussiennes

Comme il existe de nombreuses définitions et mesures de CF, il est important de pouvoir les comparer. Les exemples de mesures de CFs les plus courantes ont été présentées dans la Section 1.2.3. La corrélation constitue la mesure la plus intuitive. La cohérence, la PLV et l'AEC sont respectivement des extensions de la corrélation dans le domaine fréquentiel, en considérant la phase instantanée et l'amplitude des signaux.

Devant ce foisonnement de méthodes [192], il est intéressant d'étudier leurs points communs et leurs différences. C'est pourquoi nous proposons dans cette section d'étudier, dans le cas où l'on fait l'hypothèse que les signaux sont gaussiens et centrés, l'écriture de différentes mesures de CFs. Cette étude offre une nouvelle formule de moments d'ordres supérieurs, qui permet de décrire de nombreuses relations phase-amplitude entre deux signaux gaussiens comme fonctions des éléments de leur matrice de covariance. En plus de fournir un lien tangible entre PLV, AEC et corrélation, cette formule permet de calculer des mesures de CFs plus complexes considérant à la fois les phases et les amplitudes instantanées.

De nombreuses procédures font l'hypothèse que les signaux analysés ont un comportement gaussien. On peut citer par exemple les modèles graphiques [201] régularisés comme le lasso graphique [74]. Ces modèles, appliqués à des signaux multivariés, s'intéressent à des mesures de corrélation entre chaque paire de signaux en connaissance des autres signaux. L'idée derrière cette méthode est d'éliminer les interactions indirectes, qui sont des produits de combinaisons linéaires avec ces autres signaux. Nous proposons alors une extension de la formule générale pour les signaux multivariés gaussiens, permettant dans ce cas, de conditionner certaines mesures de CFs comme la PLV et l'AEC.

En Section 2.1, on présente dans un premier temps, le contexte derrière les signaux analytiques bivariés gaussiens ainsi que les fonctions spéciales utiles à ce chapitre. Ensuite, la formule générale de moments d'ordres supérieurs des relations phase-amplitude dans le cas gaussien est introduite. On propose la démonstration de cette formule en Section 2.2. Enfin, son extension dans le cas des signaux multivariés gaussien est fournie en Section 2.3.

2.1 Formules de moments d'ordres supérieurs des relations phase-amplitude dans le cas gaussien

2.1.1 Fonctions spéciales

Cette section présente brièvement trois familles de fonctions utiles pour la suite des calculs. Seules leurs propriétés utiles pour ce chapitre sont données. Ces fonctions sont : fonctions de Bessel modifiées, les intégrales elliptiques et la fonction hypergéométrique de Gauss.

Fonctions de Bessel modifiées

Les fonctions de Bessel modifiées sont de deux espèces [86], [104], respectivement $I_\kappa(x)$ et $K_\kappa(x)$. Si on se restreint aux nombres réels, $I_\kappa(x)$ est définie pour $x \in [0, +\infty]$ et $K_\kappa(x)$ pour $x \in]0, +\infty]$ ($x \in [0, +\infty]$ si $\kappa = 0$). Quand $\kappa \in \mathbb{N}$ est un entier naturel, les fonctions de Bessel modifiées sont les intégrales suivantes :

$$I_\kappa(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{x \cos(t)} \cos(\kappa t) dt, \quad (2.1)$$

$$K_\kappa(x) = \int_0^\infty e^{-x \cosh(t)} \cosh(\kappa t) dt. \quad (2.2)$$

Les Figures 2.1.(a) et 2.1.(b) représentent pour différentes valeurs de κ les fonctions $I_\kappa(x)$ et $K_\kappa(x)$.

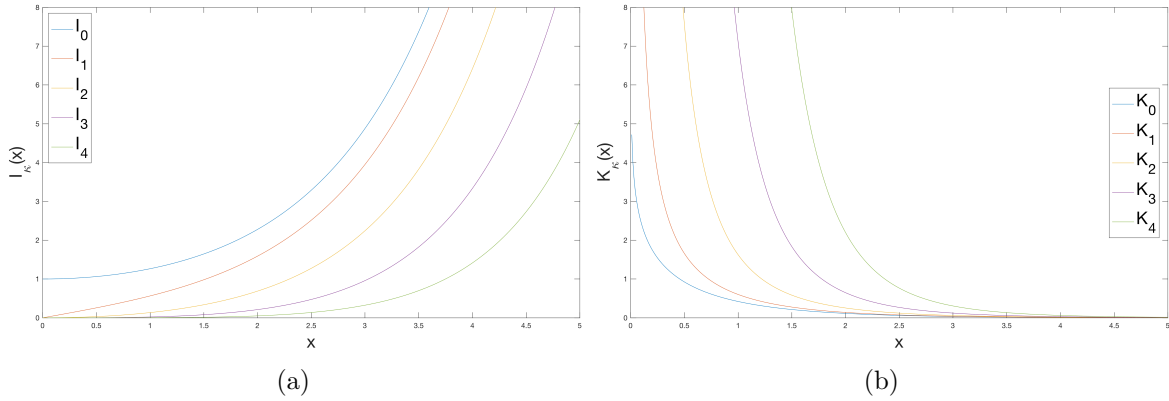


FIGURE 2.1 – (a) Fonction de Bessel modifiée de première espèce, $I_\kappa(x)$, pour $\kappa \in \{1, 2, 3, 4\}$, (b) Fonction de Bessel modifiée de deuxième espèce, $K_\kappa(x)$, pour $\kappa \in \{1, 2, 3, 4\}$

Intégrales elliptiques

Les intégrales elliptiques sont de trois espèces, seules deux nous intéressent pour la suite des calculs. Il s'agit de l'intégrale elliptique de première espèce $\mathcal{K}(x)$ définie pour $x \in [0, 1[$, et de seconde espèce $\mathcal{E}(x)$ définie pour $x \in [0, 1]$. Ces fonctions s'écrivent [86], [8] :

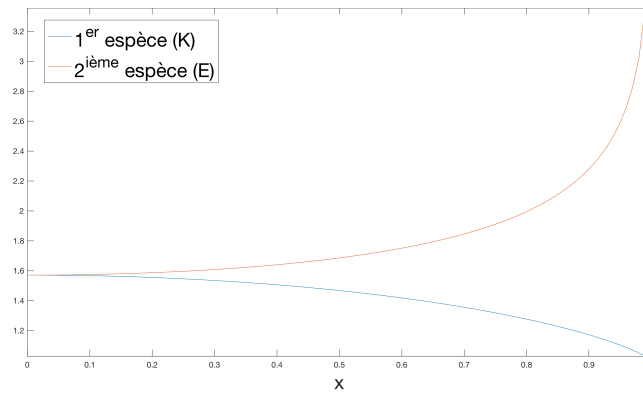
$$\mathcal{K}(x) = \int_0^1 \frac{1}{\sqrt{1-t^2}\sqrt{1-x^2t^2}} dt,$$

$$\mathcal{E}(x) = \int_0^1 \frac{\sqrt{1-x^2t^2}}{\sqrt{1-t^2}} dt. \quad (2.3)$$

La dérivée de l'intégrale elliptique de première espèce correspond à [8] :

$$\frac{d\mathcal{K}(x)}{dx} = \frac{\mathcal{E}(x) - (1-x^2)\mathcal{K}(x)}{x(1-x^2)}, \quad (2.4)$$

La Figure 2.2 représente les fonctions $\mathcal{K}(x)$ et $\mathcal{E}(x)$.


 FIGURE 2.2 – Intégrale elliptique de première et seconde espèce, respectivement notées $\mathcal{K}(x)$ et $\mathcal{E}(x)$

Fonction hypergéométrique de Gauss

La fonction hypergéométrique de Gauss est dépendante de trois paramètres b , c et d , et s'exprime sous la forme de la série hypergéométrique lorsqu'elle converge. Selon les valeurs prises par les paramètres, cette fonction correspond à de nombreuses fonctions usuelles ou spéciales. Elle est particulièrement utile pour représenter la solution de nombreuses intégrales. Elle se note [86] :

$${}_2F_1(b, c; d; y) = 1 + \frac{bc}{d}y + \frac{b(b+1)c(c+1)}{d(d+1)1(1+1)}y^2 + \frac{b(b+1)(b+2)c(c+1)(c+2)}{d(d+1)(d+2)1(1+1)(1+2)}y^3 \dots \quad (2.5)$$

Notons que y peut être une fonction $y = f(x)$. On a alors par exemple [86], [8] :

$${}_2F_1(-n, c; c; -x) = (1+x)^n, \quad \forall c, \quad (2.6)$$

$${}_2F_1\left(\frac{1}{2}, -\frac{1}{2}; \frac{1}{2}; \sin^2(x)\right) = \cos(x), \quad (2.7)$$

$${}_2F_1\left(\frac{1}{2}, \frac{1}{2}; 1; x^2\right) = \frac{2}{\pi}\mathcal{K}(x). \quad (2.8)$$

Cette fonction se dérive facilement, on a :

$$\frac{d[{}_2F_1(b, c; d; f(x))]}{dx} = \frac{df(x)}{dx} \left(\frac{bc}{d} {}_2F_1(b+1, c+1; d+1; f(x)) \right). \quad (2.9)$$

Certains ouvrages [86] ou logiciels [2] recensent des simplifications de la série ${}_2F_1(b, c; d; y)$ sous forme de fonctions usuelles ou spéciales. Le plus souvent, le glossaire des identités impliquant les fonctions hypergéométriques est rangé par valeurs de paramètres croissants $b \leq c \leq d$. Il existe alors de nombreuses formules de transformation afin de ranger par valeur croissante les paramètres on a par exemple (voir [86] pour une liste complète) :

$${}_2F_1(b, c; d; y) = (1-y)^{d-b-c} {}_2F_1(d-b, d-c; d; y). \quad (2.10)$$

2.1.2 Signal analytique bivarié gaussien

Signal analytique gaussien

On considère un signal temporel stationnaire gaussien et centré $\mathbf{s} \in \mathbb{R}^N$ composé de N échantillons. Chaque échantillon $s(n)$ est alors une réalisation d'une loi gaussienne centrée de variance σ_g^2 . Comme les échantillons sont assumés indépendants, l'index temporel peut être retiré $s(n) \sim s \forall n \in \{1, \dots, N\}$. La variable aléatoire s a pour densité de probabilité :

$$p_s(s) = \frac{1}{\sigma_g \sqrt{2\pi}} e^{-\frac{s^2}{2\sigma_g^2}}, \quad (2.11)$$

On considère maintenant la représentation analytique du signal \mathbf{s} , notée $\mathbf{z} = \mathbf{s} + \mathbf{i}\mathcal{H}[\mathbf{s}]$, avec $\mathcal{H}[\mathbf{s}]$ la transformée de Hilbert du signal \mathbf{s} . On rappelle que le nombre imaginaire est noté dans ce manuscrit $\mathbf{i} = \sqrt{-1}$. Dans le contexte d'un signal gaussien et centré, deux propriétés sur les variables aléatoires $s(n) \sim s$ et $\mathcal{H}[s](n) \sim \mathcal{H}[s] \forall n$ permettent de construire la densité de probabilité associée à chaque échantillon du signal \mathbf{z} [134], elles sont :

$$p_{\mathcal{H}[s]}(\mathcal{H}[s]) = p_s(\mathcal{H}[s]), \quad (2.12)$$

$$\mathbb{E}\{s\mathcal{H}[s]\} = 0. \quad (2.13)$$

Ces deux propriétés montrent que les signaux \mathbf{s} et $\mathcal{H}[\mathbf{s}]$ sont indépendants et suivent la même distribution que l'Équation (2.11). Ainsi, on peut facilement montrer, en posant $\sigma^2 = 2\sigma_g^2$ que les échantillons de $z(n) \sim z$ suivent une loi gaussienne complexe circulaire et symétrique qui a pour densité de probabilité :

$$p_z(z) = \frac{1}{\sigma^2\pi} e^{-\frac{|z|^2}{\sigma^2}}, \quad \forall z \in \mathbb{C}; \quad (2.14)$$

La représentation polaire du signal \mathbf{z} s'écrit $z(n) = a(n)e^{\mathbf{i}\phi(n)} \forall n \in \{1, \dots, N\}$ avec $a(n)$ et $\phi(n)$ qui sont respectivement l'amplitude et la phase instantanée du signal \mathbf{s} à l'instant n . Ils peuvent être calculés via le changement de variable suivant :

$$a(n) = \sqrt{s(n)^2 + \mathcal{H}[s](n)^2}, \quad (2.15)$$

$$\phi(n) = \text{atan2}(s(n), \mathcal{H}[s](n)), \quad (2.16)$$

avec atan2 la variante usuelle de la fonction arc tangente afin de récupérer la phase sur tout le cercle géométrique de $-\pi$ à $+\pi$. En effet, la fonction atan ne permet que de récupérer la phase sur l'intervalle $[-\frac{\pi}{2}, \frac{\pi}{2}]$, cela empêche la représentation des valeurs $z(n)$ telles que $|\arg(z(n))| \geq \frac{\pi}{2}$. Comme les éléments du signal \mathbf{z} sont indépendants, l'index temporel peut de nouveau être retiré avec $a(n) \sim a$ et $\phi(n) \sim \phi \forall n \in \{1, \dots, N\}$. La loi de probabilité conjointe entre la phase et l'amplitude a pour densité de probabilité :

$$p_{a,\phi}(a, \phi) = \frac{a}{\sigma^2\pi} e^{-\frac{a^2}{\sigma^2}}. \quad (2.17)$$

On peut en déduire les densités de probabilité respective des variables a_n et ϕ_n qui correspondent respectivement à une loi de Rayleigh et une loi uniforme sur l'intervalle $[-\pi, \pi]$, avec pour densité de probabilité :

$$p_a(a) = \frac{2a}{\sigma^2} e^{-\frac{a^2}{\sigma^2}}, \quad a \in [0, +\infty], \quad (2.18) \quad p_\phi(\phi) = \frac{1}{2\pi}, \quad \phi \in [-\pi, \pi]. \quad (2.19)$$

Extension au cas bivarié

On considère maintenant une mesure multivariée de I signaux sur N échantillons temporels, notée $\mathbf{S} \in \mathbb{R}^{I \times N}$. Chaque vecteur $\mathbf{s}(n) \forall n \in \{1, \dots, N\}$ est une réalisation d'une loi multivariée gaussienne et centrée. On a naturellement une représentation analytique de chaque signal $\mathbf{Z} \in \mathbb{R}^{I \times N}$, qui s'écrit $z_i(n) = a_i(n)e^{\mathbf{i}\phi_i(n)}$. On s'intéresse plus particulièrement à deux signaux analytiques $z_i(n)$ et $z_j(n)$, comme les échantillons sont indépendants on note respectivement les variables aléatoires $z_i(n) \sim z_i$ et $z_j(n) \sim z_j \forall n \in \{1, \dots, N\}$. Finalement, la matrice de covariance analytique entre ces deux signaux, écrite en coordonnées polaires s'écrit $k_{ij}e^{\mathbf{i}\mu_{ij}} = \mathbb{E}\{z_i z_j\}$. Conjointement, ces deux signaux suivent une loi de probabilité gaussienne complexe circulaire et symétrique bivariée qui a pour densité de probabilité [13] :

$$p_{z_i, z_j}(z_i, z_j) = \frac{1}{\pi^2 \det(\mathbf{K})} e^{-[z_i^*, z_j^*] \mathbf{K}^{-1} [z_i, z_j]^t}, \quad (2.20)$$

avec :

$$\mathbf{K} = \begin{pmatrix} k_{ii} & k_{ij}e^{\mathbf{i}\mu_{ij}} \\ k_{ij}e^{-\mathbf{i}\mu_{ij}} & k_{jj} \end{pmatrix}, \quad (2.21) \quad \det(\mathbf{K}) = k_{ii}k_{jj} - k_{ij}^2. \quad (2.22)$$

On peut alors de la même manière que dans le cas univarié, déterminer la densité de probabilité conjointe entre les amplitudes et les phases instantanées des échantillons des signaux \mathbf{z}_i et \mathbf{z}_j , qui s'écrivent :

$$\begin{aligned} p_{a_i, a_j, \phi_i, \phi_j}(a_i, a_j, \phi_i, \phi_j) &= \frac{a_i a_j}{\pi^2 \det(\mathbf{K})} e^{-(a_i e^{-i\phi_i}, a_j e^{-i\phi_j}) \mathbf{K}^{-1} (a_i e^{i\phi_i}, a_j e^{i\phi_j})^t}, \\ &= \frac{a_i a_j}{\pi^2 \det(\mathbf{K})} e^{-\frac{1}{\det(\mathbf{K})} [k_{jj} a_i^2 + k_{ii} a_j^2 - k_{ij} a_i a_j (e^{-i(\phi_i - \phi_j - \mu_{ij})} + e^{i(\phi_i - \phi_j - \mu_{ij})})]}, \\ &= \frac{a_i a_j}{\pi^2 \det(\mathbf{K})} e^{-\frac{1}{\det(\mathbf{K})} (k_{jj} a_i^2 + k_{ii} a_j^2 - 2k_{ij} a_i a_j \cos(\phi_i - \phi_j - \mu_{ij}))}. \end{aligned} \quad (2.23)$$

2.1.3 Formule des moments phases-amplitudes

À partir de la distribution de l'Équation (2.23), on va démontrer ici la proposition suivante :

Proposition 1 *Soit un signal multivarié gaussien centré et stationnaire, dont les éléments à un instant quelconque s'écrivent $s_i = a_i \cos(\phi_i)$, $\forall i \in \{1, \dots, I\}$ et \mathbf{K} la matrice de covariance analytique entre les signaux i et j est définie dans l'Équation (2.21). Il est possible de proposer une formule permettant de calculer l'espérance de certaines relations entre la phase et l'amplitude en fonction des éléments de la matrice de covariance analytique des signaux. Pour tout entiers $(m, p, q) \in (\mathbf{N})^3$, tel que $\gamma^+ = 2 + m + p + q$, $\gamma^- = 2 + m - p + q$ et sous la condition que $\gamma^- > 0$ on a :*

$$\begin{aligned} \mathbb{E}_{a_i, a_j, \phi_i, \phi_j} \left\{ (a_i a_j)^m \left(\frac{a_i}{a_j} \right)^p e^{iq(\phi_i - \phi_j)} \right\} &= \\ \frac{e^{iq\mu_{ij}}}{\det(\mathbf{K})} \left(\frac{\det(\mathbf{K})}{\sqrt{k_{ii} k_{jj}}} \right)^{m+2} \left(\sqrt{\frac{k_{ii}}{k_{jj}}} \right)^p \left(\frac{k_{ij}}{\sqrt{k_{ii} k_{jj}}} \right)^q \frac{\Gamma\left(\frac{\gamma^+}{2}\right) \Gamma\left(\frac{\gamma^-}{2}\right)}{\Gamma(q+1)} {}_2F_1 \left(\frac{\gamma^+}{2}, \frac{\gamma^-}{2}; q+1; \frac{k_{ij}^2}{k_{ii} k_{jj}} \right), \end{aligned} \quad (2.24)$$

$\Gamma(\bullet)$ correspond à la fonction gamma. Cette formule est démontrée en Section 2.2.

Des formules similaires ont déjà été proposées dans la littérature. Cependant, nous pensons que la formulation de l'Équation (2.24) est plus générale que la plupart des formules proposées précédemment car elle intègre en plus le rapport des amplitudes. Dans [142], est explicitée la relation pour tous m et q mais avec $p = 0$. Notons qu'on peut retrouver dans [81], [150], [142], des cas particuliers de l'Équation (2.24) pour certaines valeurs de (m, p, q) .

Exemple d'utilisation

On s'intéresse à appliquer la formule de l'Équation (2.24) pour trouver la PLV entre deux signaux suivant une distribution bivariable gaussienne et centrée. Notons que la PLV correspond au module de l'espérance de l'Équation (2.24) dans la situation où $m = 0$, $p = 0$ et $q = 1$. Ainsi on a :

$$|\mathbb{E}_{\phi_i, \phi_j} \{e^{i(\phi_i - \phi_j)}\}| = \frac{\pi k_{ij} \det(\mathbf{K})}{4(k_{ii} k_{jj})^{\frac{3}{2}}} {}_2F_1 \left(\frac{3}{2}, \frac{3}{2}; 2; \frac{k_{ij}^2}{k_{ii} k_{jj}} \right), \quad (2.25)$$

À partir de cette formule, un logiciel de calcul formel peut être utilisé afin de déterminer une représentation sous forme de fonctions usuelles ou spéciales de l'Équation (2.25). On peut aussi remarquer, grâce à la formulation de la dérivée des fonctions hypergéométriques de Gauss de l'Équation (2.9), et en posant $\rho^2 = \frac{k_{ij}^2}{k_{ii} k_{jj}}$ (avec ρ la corrélation analytique entre les signaux i et j) que :

$${}_2F_1 \left(\frac{3}{2}, \frac{3}{2}; 2; \rho^2 \right) = \frac{2}{\rho} \frac{d}{d\rho} \left({}_2F_1 \left(\frac{1}{2}, \frac{1}{2}; 1; \rho^2 \right) \right), \quad (2.26)$$

$$= \frac{4}{\pi \rho} \frac{d\mathcal{K}(\rho)}{d\rho}, \quad (2.27)$$

$$= \frac{4}{\pi \rho} \left(\frac{\mathcal{E}(\rho) - (1 - \rho^2)\mathcal{K}(\rho)}{\rho(1 - \rho^2)} \right), \quad (2.28)$$

Le passage de l'Équation (2.26) à l'Équation (2.27) est une application directe de la formule de l'Équation (2.8). La dérivée de la fonction $\mathcal{K}(\rho)$ est fournie par l'Équation (2.4). On peut alors maintenant en déduire, comme montré dans [142], [81], [150], l'expression suivante :

$$|\mathbb{E}_{\phi_i, \phi_j} \{e^{i(\phi_i - \phi_j)}\}| = \frac{1}{\rho} \left(\mathcal{E}(\rho) - (1 - \rho^2) \mathcal{K}(\rho) \right). \quad (2.29)$$

Table de résultats

On fournit dans le Tableau 2.1, pour différentes valeurs des paramètres (m, p, q) , des simplifications de l'équation générale en terme de fonctions usuelles ou spéciales. Notons que dans le cas où les signaux sont normalisés, impliquant $k_{ii} = k_{jj}$, on obtient une interprétation remarquable des intégrales elliptiques des deux premières espèces, comme espérance de l'inverse de la multiplication pour la première et du ratio des amplitudes pour la seconde. En effet, on a :

$$\mathbb{E}_{a_i, a_j} \left\{ \frac{1}{a_i a_j} \right\} = \mathcal{K}(k_{ij}), \quad (2.30) \quad \mathbb{E}_{a_i, a_j} \left\{ \frac{a_i}{a_j} \right\} = \mathcal{E}(k_{ij}). \quad (2.31)$$

On illustre les résultats 1 à 7 du Tableau 2.1 dans la Figure 2.3. Les représentations sont faites pour le cas normalisé $k_{ii} = k_{jj} = 1$ et en fonction de ρ . On remarque que dans tous les cas, les espérances obtenues sont des fonctions monotones de la corrélation. Cela montre que dans le cas gaussien, toutes les relations entre phases et amplitudes obtenues grâce à la formule de l'Équation (2.24), sont liées à la mesure de corrélation. On peut alors faire la proposition suivante :

Proposition 2 *Sous hypothèse qu'à chaque instant, le signal \mathbf{S} est une réalisation indépendante d'une loi multivariée Gaussienne, alors la PLV, l'AEC entre les signaux i et j sont des fonctions monotones de la corrélation ρ qui s'écrivent :*

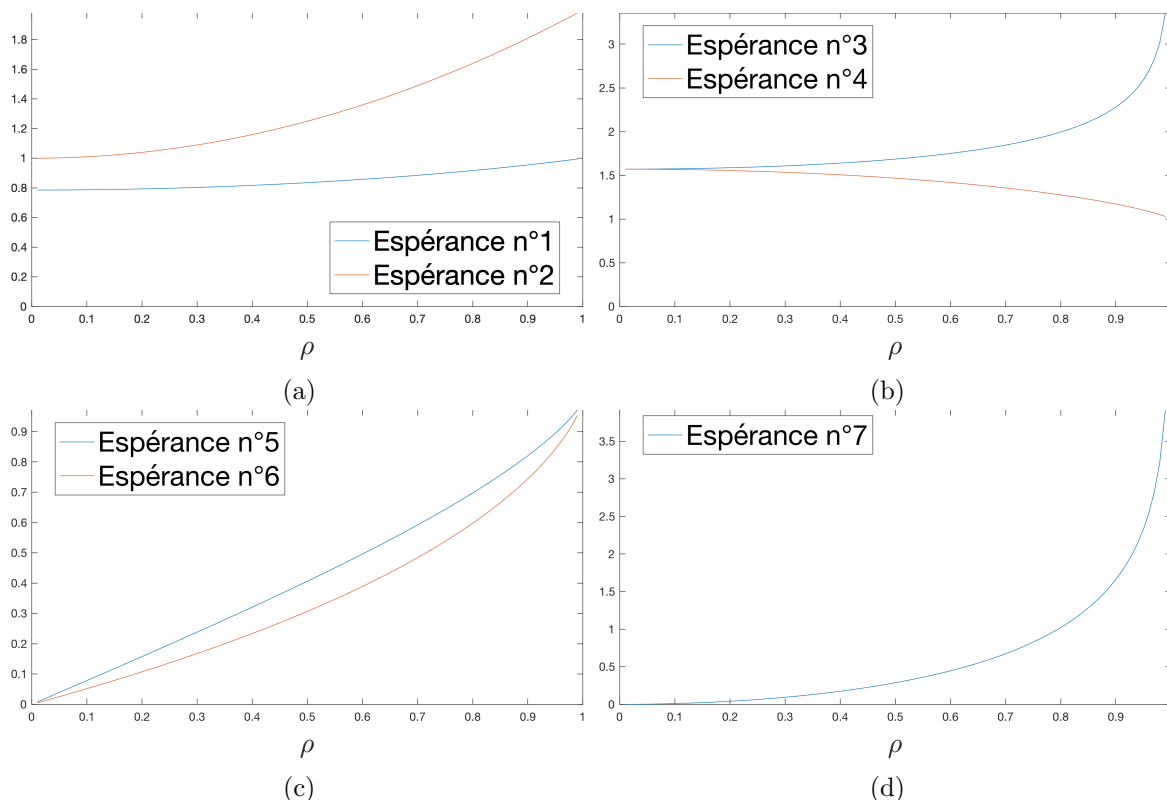
$$\text{PLV} = \frac{1}{\rho} (\mathcal{E}(\rho) - (1 - \rho^2) \mathcal{K}(\rho)), \quad (2.32)$$

$$\text{AEC} = \frac{1}{4 - \pi} \left(4\mathcal{E}(\rho) - 2(1 - \rho^2) \mathcal{K}(\rho) - \pi \right). \quad (2.33)$$

On précise que la corrélation ρ entre les signaux i et j peut être exprimée à partir de leur covariance analytique \mathbf{K} [134] telle que :

$$\rho = \frac{k_{ij}}{\sqrt{k_{ii} k_{jj}}}. \quad (2.34)$$

No.	Valeurs de $[m, p, q]$	Espérance	Cas classique avec $\rho = \frac{k_{ij}}{\sqrt{k_{ii}k_{jj}}}$
1	[1,0,0]	$\mathbb{E}_{a_i, a_j} \{a_i a_j\}$	$\sqrt{k_{ii}k_{jj}} \left(\mathcal{E}(\rho) - \frac{1}{2}(1 - \rho^2)\mathcal{K}(\rho) \right)$
2	[2,0,0]	$\mathbb{E}_{a_i, a_j} \{(a_i a_j)^2\}$	$k_{ii}k_{jj} + k_{ij}^2$
3	[-1,0,0]	$\mathbb{E}_{a_i, a_j} \left\{ \frac{1}{a_i a_j} \right\}$	$\frac{1}{\sqrt{k_{ii}k_{jj}}} \mathcal{K}(\rho)$
4	[0,1,0]	$\mathbb{E}_{a_i, a_j} \left\{ \frac{a_i}{a_j} \right\}$	$\sqrt{\frac{k_{ii}}{k_{jj}}} \mathcal{E}(\rho)$
5	[0,0,1]	$\mathbb{E}_{\phi_i, \phi_j} \{e^{i(\phi_i - \phi_j)}\}$	$\frac{1}{\rho} (\mathcal{E}(\rho) - (1 - \rho^2)\mathcal{K}(\rho)) e^{i\mu_{ij}}$
6	[0,0,2]	$\mathbb{E}_{\phi_i, \phi_j} \{e^{i2(\phi_i - \phi_j)}\}$	$\frac{1}{\rho} ((1 - \rho)\log(1 - \rho) + \rho) e^{i2\mu_{ij}}$
7	[-1,0,1]	$\mathbb{E}_{a_i, a_j, \phi_i, \phi_j} \left\{ \frac{e^{i(\phi_i - \phi_j)}}{a_i a_j} \right\}$	$\frac{-e^{i\mu_{ij}}}{\sqrt{k_{ii}k_{jj}}} \log(1 - \rho^2)$
8	[1,0,1]	$\mathbb{E}_{a_i, a_j, \phi_i, \phi_j} \{a_i a_j e^{i(\phi_i - \phi_j)}\}$	$k_{ij} e^{i\mu_{ij}}$
9	[1,1,0]	$\mathbb{E}_{a_i} \{a_i^2\}$	k_{ii}

 Tableau 2.1 – Réécriture de la formule générale de l'Équation (2.24) pour des valeurs particulières des paramètres (m, p, q)

 FIGURE 2.3 – Représentation de l'espérance obtenue pour les sept premières situations du Tableau (2.1) quand $k_{ii} = k_{jj} = 1$ et ρ varie entre 0 et 1. La figure (a) représente les espérances 1 et 2, la figure (b) les espérances 3 et 4, la figure (c) les espérances 5 et 6, la figure (d) l'espérance 7.

2.2 Démonstration de la formule des moments phases-amplitudes

L'objectif de cette section est de prouver la formule de l'Équation (2.24), ce qui correspond à calculer l'intégrale suivante :

$$\begin{aligned} \mathbb{E} \left\{ (a_i a_j)^m \left(\frac{a_i}{a_j} \right)^p e^{\dot{i}q(\phi_i - \phi_j)} \right\} &= \int_0^\infty \int_0^\infty \int_0^{2\pi} \int_0^{2\pi} (a_i a_j)^m \left(\frac{a_i}{a_j} \right)^p e^{\dot{i}q(\phi_i - \phi_j)} p(a_i, a_j, \phi_i, \phi_j) d\phi_j d\phi_i da_j da_i \\ &= \int_0^\infty \int_0^\infty \int_0^{2\pi} \int_0^{2\pi} (a_i a_j)^m \left(\frac{a_i}{a_j} \right)^p e^{\dot{i}q(\phi_i - \phi_j)} \frac{a_i a_j}{\pi^2 \det(\mathbf{K})} e^{-\frac{1}{\det(\mathbf{K})}(k_{jj}a_i^2 + k_{ii}a_j^2 - 2k_{ij}a_i a_j \cos(\phi_i - \phi_j - \mu_{ij}))} d\phi_j d\phi_i da_j da_i, \end{aligned} \quad (2.35)$$

avec $p(a_i, a_j, \phi_i, \phi_j)$ la distribution des phases et amplitudes conjointes de l'Équation (2.23).

2.2.1 Changement de variable

Afin de simplifier le calcul de l'intégrale de l'Équation (2.35), on propose le changement de variable suivant :

$$\begin{aligned} \psi &\equiv \phi_j - \phi_i \pmod{2\pi}, & \psi &\in [0, 2\pi], \\ v &= a_i a_j, & v &\in [0, \infty], \\ w &= \ln \left(\frac{a_i}{a_j} \right), & w &\in \mathbb{R}, \end{aligned} \quad (2.36)$$

La variable ψ correspond à la phase relative redistribuée sur l'intervalle $[0, 2\pi]$. Il s'agit de la variable d'intérêt dans le calcul de la PLV. Le second groupe de changement de variables consiste à réécrire le vecteur d'amplitudes (a_i, a_j) en coordonnées hyperboliques. Au même titre que des coordonnées polaires sont un couple de nombre (a, ϕ) permettant de localiser n'importe quel point dans le plan, les coordonnées hyperboliques permettent d'écrire n'importe quel point de coordonnées non-négatives sous forme d'un couple (v, w) où v correspond au produit des coordonnées et w est leur proportion logarithmique. À partir de ce changement de variable, le calcul de l'intégrale de l'Équation (2.35) devient :

$$\begin{aligned} \mathbb{E}_{a_i, a_j, \phi_i, \phi_j} \left\{ (a_i a_j)^m \left(\frac{a_i}{a_j} \right)^p e^{\dot{i}q(\phi_i - \phi_j)} \right\} &= \mathbb{E}_{v, w, \psi} \left\{ v^m e^{pw + \dot{i}q\psi} \right\}, \\ &= \int_0^\infty \int_0^{2\pi} \int_0^{2\pi} v^m e^{pw + \dot{i}q\psi} p_{v, w, \psi}(v, w, \psi) d\psi dw dv. \end{aligned} \quad (2.37)$$

On démontre alors la proposition suivante :

Proposition 3 *Soit un signal multivarié gaussien centré et stationnaire, dont les éléments à un instant quelconque s'écrivent $s_i = a_i \cos(\phi_i)$, $\forall i \in \{1, \dots, I\}$ et \mathbf{K} la matrice de covariance analytique entre les signaux i et j définie dans l'Équation (2.21). Alors la distribution marginale entre les variables ψ , v et w définie par l'Équation (2.36) correspondant respectivement à la phase relative, au produit des amplitudes et à la proportion logarithmique des amplitudes s'écrit :*

$$p_{v, w, \psi}(v, w, \psi) = \frac{v}{\pi \det(\mathbf{K})} e^{-\alpha v \cosh(w - \bar{\mu}) + \beta v \cos(\psi - \mu)}, \quad (2.38)$$

avec les paramètres suivants :

$$\begin{aligned} \alpha &= \frac{2\sqrt{k_{ii}k_{jj}}}{\det(\mathbf{K})}, & \beta &= \frac{2k_{ij}}{\det(\mathbf{K})}, \\ \bar{\mu} &= \ln \left(\sqrt{\frac{k_{ii}}{k_{jj}}} \right), & \mu &= \mu_{ij}. \end{aligned} \quad (2.39)$$

Démonstration : On rappelle la distribution conjointe des phases et amplitudes de l'Équation (2.23) :

$$p_{a_i, a_j, \phi_i, \phi_j}(a_i, a_j, \phi_i, \phi_j) = \frac{a_i a_j}{\pi^2 \det(\mathbf{K})} e^{-\frac{1}{\det(\mathbf{K})}(k_{jj} a_i^2 + k_{ii} a_j^2 - 2k_{ij} a_i a_j \cos(\phi_i - \phi_j - \mu_{ij}))}. \quad (2.40)$$

On procède d'abord au changement de variable intermédiaire suivant :

$$\varphi = \phi_i - \phi_j, \quad \varphi \in [-2\pi, 2\pi], \quad (2.41)$$

$$\phi_c = \phi_j = \phi_i - \varphi, \quad \phi_c \in [0, 2\pi], \quad (2.42)$$

$$v = a_i a_j, \quad (2.43)$$

$$w = \ln \left(\frac{a_i}{a_j} \right). \quad (2.44)$$

On retrouve le produit des amplitudes et la proportion logarithmique. Afin de rendre le changement de variable possible, deux variables sont introduites pour remplacer les phases ϕ_i et ϕ_j . φ correspond à la phase relative sur $[-2\pi, 2\pi]$ et ϕ_c est définie comme la phase ϕ_j .

La valeur absolue du déterminant de la matrice jacobienne (notée \mathbf{J}) de ces changements de variables se calcule simplement :

$$\begin{aligned} |\det(\mathbf{J})| &= \left| \det \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{e^{\frac{w}{2}}}{2\sqrt{v}} & \frac{e^{-\frac{w}{2}}}{2\sqrt{v}} \\ 0 & 0 & \frac{\sqrt{v}e^{\frac{w}{2}}}{2} & -\frac{\sqrt{v}e^{-\frac{w}{2}}}{2} \end{pmatrix} \right|, \\ &= \left| \det \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \det \begin{pmatrix} \frac{e^{\frac{w}{2}}}{2\sqrt{v}} & \frac{e^{-\frac{w}{2}}}{2\sqrt{v}} \\ \frac{\sqrt{v}e^{\frac{w}{2}}}{2} & -\frac{\sqrt{v}e^{-\frac{w}{2}}}{2} \end{pmatrix} \right|, \\ &= \frac{1}{2}. \end{aligned} \quad (2.45)$$

On obtient donc :

$$\begin{aligned} p_{v, w, \varphi, \phi_c}(v, w, \varphi, \phi_c) &= \frac{v}{2\pi^2 \det(\mathbf{K})} e^{-\frac{1}{\det(\mathbf{K})}(k_{jj} v e^w + k_{ii} v e^{-w} - 2k_{ij} v \cos(\varphi - \mu_{ij}))}, \\ &= \frac{v}{2\pi^2 \det(\mathbf{K})} e^{-\frac{1}{\det(\mathbf{K})} \left(\sqrt{k_{jj} k_{ii}} v e^{w - \ln \left(\sqrt{\frac{k_{ii}}{k_{jj}}} \right)} + \sqrt{k_{jj} k_{ii}} v e^{-w + \ln \left(\sqrt{\frac{k_{ii}}{k_{jj}}} \right)} - 2k_{ij} v \cos(\varphi - \mu_{ij}) \right)}, \end{aligned}$$

$$= \frac{v}{2\pi^2 \det(\mathbf{K})} e^{-\frac{1}{\det(\mathbf{K})} \left(2\sqrt{k_{jj} k_{ii}} v \cosh \left(w - \ln \left(\sqrt{\frac{k_{ii}}{k_{jj}}} \right) \right) - 2k_{ij} v \cos(\varphi - \mu_{ij}) \right)} \quad (2.46)$$

$$= \frac{v}{2\pi^2 \det(\mathbf{K})} e^{-\alpha v \cosh(w - \bar{\mu}) + \beta v \cos(\varphi - \mu)} \quad (2.47)$$

Le passage de l'Équation (2.46) à l'Équation (2.47) utilise la paramétrisation de l'Équation (2.39).

Afin de pouvoir démontrer la Proposition 3, il faut intégrer selon la variable ϕ_c et redistribuer la phase relative φ sur le cercle unité. L'élimination de la variable $\phi_c \in [0, 2\pi]$ n'est pas simple car elle est dépendante de la variable $\varphi \in [-2\pi, 2\pi]$. La Figure 2.4 montre l'ensemble des valeurs que peut prendre la variable ϕ_c en fonction de φ .

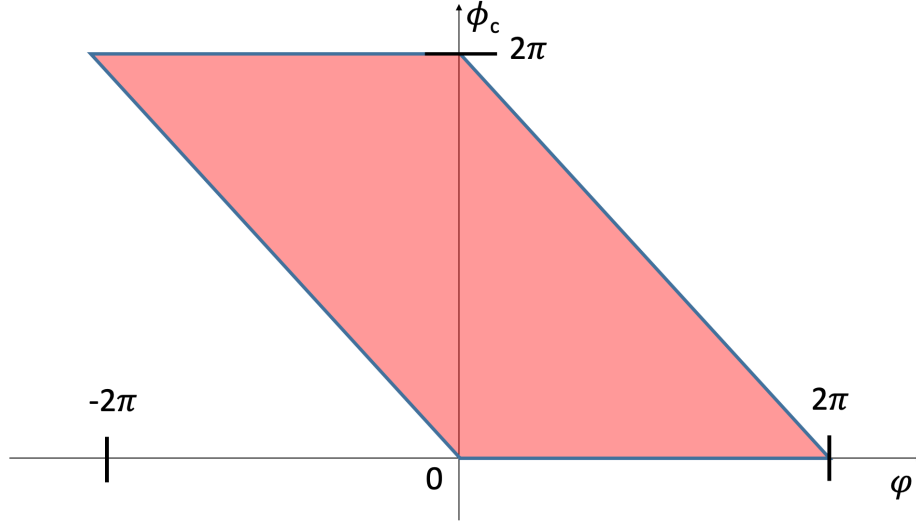


FIGURE 2.4 – Le rectangle rouge représente l'ensemble des valeurs que peut prendre la variable ϕ_c en fonction de φ

On a alors :

$$\begin{aligned}
 p_{v,w,\varphi}(v, w, \varphi) &= \begin{cases} \int_0^{2\pi-\varphi} p_{v,w,\varphi,\phi_c}(v, w, \varphi, \phi_c) d\phi_c, & \text{si } \varphi \in [0, 2\pi], \\ \int_{-\varphi}^{2\pi} p_{v,w,\varphi,\phi_c}(v, w, \varphi, \phi_c) d\phi_c, & \text{sinon,} \end{cases} \\
 &= \begin{cases} (2\pi - \varphi)p_{v,w,\varphi,\phi_c}(v, w, \varphi, \phi_c), & \text{si } \varphi \in [0, 2\pi], \\ (2\pi + \varphi)p_{v,w,\varphi,\phi_c}(v, w, \varphi, \phi_c), & \text{sinon,} \end{cases} \\
 &= (2\pi - |\varphi|)p_{v,w,\varphi,\phi_c}(v, w, \varphi, \phi_c). \tag{2.48}
 \end{aligned}$$

La redistribution du support de φ sur le cercle unité permet de définir la variable $\psi \in [0, 2\pi]$. Comme φ est une variable circulaire : φ et $\varphi - 2\pi$ sont équivalents et indiquent le même angle. On a donc :

$$\begin{aligned}
 p_{v,w,\psi}(v, w, \psi) &= p_{v,w,\varphi}(v, w, \psi) + p_{v,w,\varphi}(v, w, \psi - 2\pi), \\
 &= (2\pi - |\psi|)p_{v,w,\varphi,\phi_c}(v, w, \psi, \phi_c) + (2\pi - |\psi - 2\pi|)p_{v,w,\varphi,\phi_c}(v, w, \psi - 2\pi, \phi_c), \\
 &= (2\pi - \psi)p_{v,w,\varphi,\phi_c}(v, w, \psi, \phi_c) + (2\pi - (2\pi - \psi))p_{v,w,\varphi,\phi_c}(v, w, \psi, \phi_c), \\
 &= 2\pi p_{v,w,\varphi,\phi_c}(v, w, \psi, \phi_c), \\
 &= \frac{v}{\pi \det(\mathbf{K})} e^{-\alpha v \cosh(w - \bar{\mu}) + \beta v \cos(\psi - \mu)}. \tag{2.49}
 \end{aligned}$$

Ce qui démontre la Proposition 3. □

2.2.2 Loi marginale conditionnée au produit des amplitudes

Dans l'objectif de réduire le nombre de variables dans le calcul de l'intégrale de l'Équation (2.35), il s'avère utile de conditionner la distribution marginale de la phase relative ψ et de la proportion logarithmique w au produit des amplitudes v . On démontre alors la proposition suivante :

Proposition 4 Soit un signal multivarié gaussien centré et stationnaire, dont les éléments à un instant quelconque s'écrivent $s_i = a_i \cos(\phi_i)$, $\forall i \in \{1, \dots, I\}$ et \mathbf{K} la matrice de covariance analytique entre les signaux i et j définie dans l'Équation (2.21). Alors la distribution marginale entre les variables ψ et w définie par l'Équation (2.36) correspondant respectivement à la phase relative et à la proportion logarithmique des amplitudes conditionnées au produit des amplitudes v s'écrit :

$$p_{v,w,\psi}(w, \psi | v) = \frac{1}{4\pi K_0(\alpha v) I_0(\beta v)} e^{-\alpha v \cosh(w - \bar{\mu}) + \beta v \cos(\psi - \mu)}. \quad (2.50)$$

Les paramètres α , β , μ et $\bar{\mu}$ sont définis dans l'Équation (2.39), $I_0(\bullet)$ et $K_0(\bullet)$ sont les intégrales de Bessel modifiées de première et seconde espèce définies respectivement dans les Équations (2.1) et (2.2).

Démonstration : La démonstration de la Proposition 4 se base sur la distribution du produit de deux amplitudes v . Cette distribution est déjà connue [142], [13]. Cependant, on propose une nouvelle démonstration particulièrement courte à partir de la distribution $p_{v,w,\psi}(v, w, \psi)$:

$$\begin{aligned} p_v(v) &= \int \int_0^{2\pi} p_{v,w,\psi}(v, w, \psi) \, d\psi \, dw, \\ &= \frac{v}{\pi \det(\mathbf{K})} \int e^{-\alpha v \cosh(w - \bar{\mu})} \, dw \int_0^{2\pi} e^{\beta v \cos(\psi - \mu)} \, d\psi. \end{aligned} \quad (2.51)$$

$$(2.52)$$

On peut calculer chacune de ces deux intégrales à partir des définitions des fonctions de Bessel de première espèce (Équation (2.1)) et de seconde espèce (Équation (2.2)). En posant maintenant $\psi' = \psi - \mu$ on retrouve :

$$\begin{aligned} \int_0^{2\pi} e^{\beta v \cos(\psi - \mu)} \, d\phi &= \int_0^{2\pi} e^{\beta v \cos(\psi')} \, d\psi', \\ &= 2\pi I_0(\beta v). \end{aligned} \quad (2.53)$$

En posant $w' = w - \bar{\mu}$ on a :

$$\begin{aligned} \int e^{-\alpha v \cosh(w - \bar{\mu})} \, dw &= 2 \int_0^\infty e^{-\alpha v \cosh(w')} \, dw', \\ &= 2K_0(\alpha v). \end{aligned} \quad (2.54)$$

Ainsi on retrouve facilement le résultat de [142] :

$$p_v(v) = \frac{4v}{\det(\mathbf{K})} K_0(\alpha v) I_0(\beta v). \quad (2.55)$$

On peut désormais calculer la densité de probabilité marginale des variables ψ et w connaissant la variable v , qui s'écrit :

$$p_{w,\psi}(w, \psi | v) = \frac{p_{v,w,\psi}(v, w, \psi)}{p_v(v)} = \frac{1}{4\pi K_0(\alpha v) I_0(\beta v)} e^{-\alpha v \cosh(w - \bar{\mu}) + \beta v \cos(\psi - \mu)}. \quad (2.56)$$

Permettant de prouver la Proposition 4. □

2.2.3 La loi de Von-Mises hyperbolique

Toujours dans l'objectif de réduire le nombre de variables dans le calcul de l'intégrale de l'Équation (2.35), on peut montrer que la variable ψ définie par l'Équation (2.36) correspondant à la phase relative lorsqu'elle est conditionnée au produit des amplitudes v suit une loi de von-Mises [104] :

$$p_\psi(\psi | v) = \int p_{w,\psi}(w, \psi | v) \, dw = \frac{1}{2\pi I_0(\beta v)} e^{\beta v \cos(\psi - \mu)}, \quad (2.57)$$

et que l'espérance de la variable $e^{im\psi}$, sachant que ψ suit une loi de Von-Mises $\forall m \in \mathbb{N}^+$ s'écrit :

$$\mathbb{E}_\psi \left\{ e^{im\psi} \mid v \right\} = \frac{I_m(\beta v)}{I_0(\beta v)} e^{im\mu}. \quad (2.58)$$

Ces résultats sont connus [104], [13], une démonstration de l'Équation (2.58) est proposée en Annexe 2.A.1. Il est alors remarquable que des résultats similaires existent pour la distribution de la proportion logarithmique w conditionnée au produit des amplitudes v . Cela est résumé dans la proposition suivante :

Proposition 5 *Soit un signal multivarié gaussien centré et stationnaire, dont les éléments à un instant quelconque s'écrivent $s_i = a_i \cos(\phi_i)$, $\forall i \in \{1, \dots, I\}$ et \mathbf{K} la matrice de covariance analytique entre les signaux i et j définie dans l'Équation (2.21). La distribution de la variable w définie par l'Équation (2.36) correspondant à la proportion logarithmique des amplitudes lorsqu'elle est conditionnée au produit des amplitudes v suit une loi de Von-Mises hyperbolique [19] :*

$$p_w(w \mid v) = \int_0^{2\pi} p_{w,\psi}(w, \psi \mid v) dw = \frac{1}{2K_0(\alpha v)} e^{-\alpha v \cosh(w - \bar{\mu})}. \quad (2.59)$$

L'espérance de la variable e^{lw} , sachant que w suit une loi de Von-Mises hyperbolique, $\forall l \in \mathbb{N}^+$ correspond à :

$$\mathbb{E}_w \left\{ e^{lw} \mid v \right\} = \frac{K_l(\alpha v)}{K_0(\alpha v)} e^{l\bar{\mu}}. \quad (2.60)$$

Les paramètres α et $\bar{\mu}$ sont définis dans l'Équation (2.39), $K_0(\bullet)$ est l'intégrale de Bessel modifiée de seconde espèce définie dans l'Équation (2.2).

Comme la distribution de Von-mises hyperbolique a déjà été étudiée [19], le résultat de l'Équation (2.60) n'est pas nouveau et on le redémontre en Annexe 2.A.2. On montre cependant ici qu'il permet de calculer l'espérance du ratio d'amplitudes conditionné au produit des amplitudes. En effet, on rappelle que :

$$\mathbb{E}_w \left\{ e^{lw} \mid v \right\} = \mathbb{E}_{\frac{a_i}{a_j}} \left\{ \left(\frac{a_i}{a_j} \right)^l \mid a_i a_j \right\}. \quad (2.61)$$

Les lois de Von-Mises et de Von-Mises hyperboliques sont respectivement représentées dans les Figures 2.5a et 2.5b pour $v = 1$. On peut voir ces distributions comme des gaussiennes sur respectivement un support circulaire [104] ou hyperbolique. Pour la distribution de Von-Mises, β varie de 0 à 4 et $\mu = 0$. On remarque que si les phases ϕ_i et ϕ_j proviennent de signaux indépendants (cas $\beta = k_{ij} = 0$), alors leur différence ψ suit une loi uniforme. Au contraire, plus les signaux sont dépendants, plus cette différence de phases devient localisée. Pour la distribution de Von-Mises hyperbolique, α varie de 0 à 4 et $\bar{\mu} = 0$. Pour $k_{ii} = k_{jj} = 1$, la proportion logarithmique devient d'autant plus localisée que les signaux sont corrélés ($\alpha = \frac{2}{1-k_{ij}^2}$). $\alpha = 0$ correspond à un cas limite non réaliste quand a_i et a_j proviennent de signaux indépendants ($k_{ij} = 0$ impliquant $\alpha = \frac{2}{\sqrt{k_{ii}k_{jj}}}$), alors cette proportion tend vers une loi uniforme quand la variance des signaux tend vers l'infini.

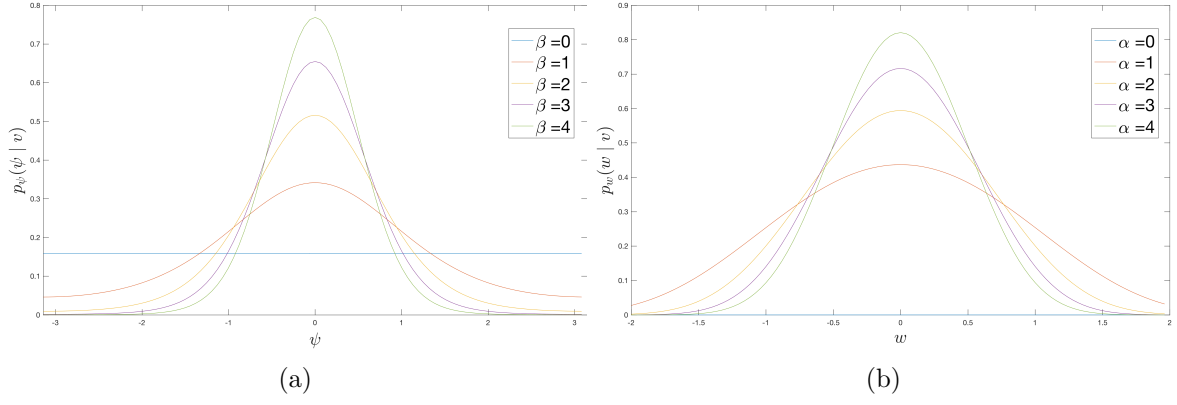


FIGURE 2.5 – (a) Distribution de Von Mises, pour différentes valeurs de β , $\mu = 0$ et $v = 1$. (b) Distribution de Von Mises hyperbolique, pour différentes valeurs de α , $\bar{\mu} = 0$ et $v = 1$.

2.2.4 Réduction au calcul d'une intégrale à une seule variable

Il est alors possible de réduire l'intégrale quadruple de l'Équation (2.35) à une intégrale sur uniquement la variable v en remarquant que :

$$\begin{aligned} \mathbb{E}_{v,w,\psi} \left\{ v^m e^{pw+iq\psi} \right\} &= \int_0^1 v^m \left[\int_0^{2\pi} \int_0^{2\pi} e^{pw+iq\psi} p_{w,\psi}(w, \psi | v) d\psi dw \right] p_v(v) dv, \\ &= \int_0^1 v^m \left[\int e^{pw} p_w(w | v) dw \int_0^{2\pi} e^{iq\psi} p_\psi(\psi | v) d\psi \right] p_v(v) dv, \\ &= \mathbb{E}_v \{ v^m \mathbb{E}_w \{ e^{pw} | v \} \mathbb{E}_\psi \{ e^{iq\psi} | v \} \}. \end{aligned} \quad (2.62)$$

Comme le résultat de $\mathbb{E}_w \{ e^{pw} | v \}$ est donné dans l'Équation (2.60), le résultat de $\mathbb{E}_\psi \{ e^{iq\psi} | v \}$ est fourni dans l'Équation (2.58) et que la distribution de $p_v(v)$ est connue (Équation (2.55)) on obtient finalement :

$$\mathbb{E}_{v,w,\psi} \left\{ v^m e^{pw+iq\psi} \right\} = \frac{4e^{p\bar{\mu}+iq\mu}}{\det(\mathbf{K})} \int_0^1 v^{m+1} K_p(\alpha v) I_q(\beta v) dv. \quad (2.63)$$

Cette formule intégrale peut se réécrire à l'aide de fonctions spéciales, en utilisant les règles de calcul de [86]. Il s'agit de l'équation 6.576.5 qui permet d'écrire en posant $\gamma^+ = 2 + m + p + q$ et $\gamma^- = 2 + m - p + q$:

$$\int_0^1 v^{m+1} K_p(\alpha v) I_q(\beta v) dv = \frac{2^m \beta^p}{\alpha^{p+m+1}} \frac{\Gamma\left(\frac{\gamma^+}{2}\right) \Gamma\left(\frac{\gamma^-}{2}\right)}{\Gamma(q+1)} {}_2F_1\left(\frac{\gamma^+}{2}, \frac{\gamma^-}{2}; q+1; \frac{\beta^2}{\alpha^2}\right), \quad (2.64)$$

$$\text{avec } m+1-p+q > 0, \quad (2.65)$$

$$\text{et } \alpha > \beta. \quad (2.66)$$

On remarque que selon l'inégalité de Cauchy-Schwartz, on a toujours $\alpha > \beta$ car $k_{ii}k_{jj} > k_{ij}^2$. Il suffit alors de respecter la condition de l'Équation (2.65) pour garantir l'existence d'une solution. Finalement, en remplaçant chaque variable β , α , μ , $\bar{\mu}$ par leurs valeurs en fonction de k_{ii} , k_{jj} , k_{ij} et μ_{ij} , on peut retrouver le résultat de la Proposition 1.

2.3 Extension aux signaux multivariés gaussiens

2.3.1 La corrélation partielle

Introduction

Soit une matrice $\mathbf{Z} \in \mathbb{R}^{I \times T}$ (complexe ou réelle), composée de I signaux $\mathbf{z}_i \in \mathbb{R}^T$ (notée pour simplifier \mathbf{z}_i). L'objectif de cette section est de montrer comment on peut calculer la corrélation entre

deux signaux i et j en prenant en compte l'existence des autres signaux de la matrice \mathbf{Z} [201]. On introduit les notations suivantes : $\mathring{\mathbf{Z}} \in \mathbb{R}^{2 \times T}$ est la matrice contenant les deux signaux d'intérêts \mathbf{z}_i et \mathbf{z}_j , la matrice contenant tous les autres signaux est $\check{\mathbf{Z}} \in \mathbb{R}^{(I-2) \times T}$. Afin de trouver la combinaison linéaire des éléments de $\check{\mathbf{Z}}$ qui explique au mieux le signal \mathbf{z}_i , on réalise la régression suivante :

$$\underset{\mathbf{b}_i}{\operatorname{argmin}} \quad \|\mathbf{z}_i - \check{\mathbf{Z}}\mathbf{b}_i\|_F^2. \quad (2.67)$$

La solution notée \mathbf{b}_i^* est :

$$\mathbf{b}_i^* = (\check{\mathbf{Z}}^*\check{\mathbf{Z}})^{-1}\check{\mathbf{Z}}^*\mathbf{z}_i. \quad (2.68)$$

Ainsi, on peut générer un signal conditionné ϵ_i qui consiste à éliminer du signal \mathbf{z}_i , sa meilleure explication linéaire en termes des signaux de la matrice $\check{\mathbf{Z}}$.

$$\epsilon_i = \mathbf{z}_i - \check{\mathbf{Z}}\mathbf{b}_i^* = (\mathbf{I} - \check{\mathbf{Z}}(\check{\mathbf{Z}}^*\check{\mathbf{Z}})^{-1}\check{\mathbf{Z}}^*)\mathbf{z}_i. \quad (2.69)$$

Ainsi on peut définir la notion de corrélation partielle :

Definition 7 : Corrélation partielle (ou conditionnée)

La corrélation partielle [201], notée $\rho\rho$, est la corrélation entre les signaux résidus ϵ_i et ϵ_j connaissant les autres signaux $\check{\mathbf{Z}}$. Elle s'écrit :

$$\rho\rho_{ij} = \frac{\mathbb{E}\{\epsilon_i\epsilon_j\}}{\sqrt{\mathbb{E}\{\epsilon_i^2\}\mathbb{E}\{\epsilon_j^2\}}}. \quad (2.70)$$

Calcul

La covariance entre les signaux ϵ_i et ϵ_j peut s'écrire en fonction des signaux de la matrice \mathbf{Z} telle que :

$$\begin{aligned} \mathbb{E}\{\epsilon_i\epsilon_j\} &= \mathbf{z}_i^*(\mathbf{I} - \check{\mathbf{Z}}(\check{\mathbf{Z}}^*\check{\mathbf{Z}})^{-1}\check{\mathbf{Z}}^*)(\mathbf{I} - \check{\mathbf{Z}}(\check{\mathbf{Z}}^*\check{\mathbf{Z}})^{-1}\check{\mathbf{Z}}^*)\mathbf{z}_j, \\ &= \mathbf{z}_i^*\mathbf{z}_j - 2\mathbf{z}_i^*\check{\mathbf{Z}}(\check{\mathbf{Z}}^*\check{\mathbf{Z}})^{-1}\check{\mathbf{Z}}^*\mathbf{z}_j + \mathbf{z}_i^*\check{\mathbf{Z}}(\check{\mathbf{Z}}^*\check{\mathbf{Z}})^{-1}\check{\mathbf{Z}}^*\check{\mathbf{Z}}(\check{\mathbf{Z}}^*\check{\mathbf{Z}})^{-1}\check{\mathbf{Z}}^*\mathbf{z}_j, \\ &= \mathbf{z}_i^*\mathbf{z}_j - \mathbf{z}_i^*\check{\mathbf{Z}}(\check{\mathbf{Z}}^*\check{\mathbf{Z}})^{-1}\check{\mathbf{Z}}^*\mathbf{z}_j. \end{aligned} \quad (2.71)$$

Cependant, le calcul de $\rho\rho_{ij}$ pour chaque paire de signaux peut être une tâche coûteuse, car cela demande pour chaque cas des multiplications et une inversion matricielle. Heureusement, il existe une astuce pour n'effectuer qu'une seule inversion matricielle afin d'obtenir toutes les valeurs de $\rho\rho_{ij}$ $\forall i \in \{1, \dots, I\}$ et $\forall j \in \{1, \dots, I\}$. On considère dans un premier temps, la matrice de covariance entre l'ensemble des signaux qui peut être décomposée en quatre blocs comme suit :

$$\mathbf{K} = \begin{pmatrix} \mathring{\mathbf{Z}} \\ \check{\mathbf{Z}} \end{pmatrix} \begin{pmatrix} \mathring{\mathbf{Z}} & \check{\mathbf{Z}} \end{pmatrix} = \begin{pmatrix} \mathring{\mathbf{Z}}^*\mathring{\mathbf{Z}} & \mathring{\mathbf{Z}}^*\check{\mathbf{Z}} \\ \check{\mathbf{Z}}^*\mathring{\mathbf{Z}} & \check{\mathbf{Z}}^*\check{\mathbf{Z}} \end{pmatrix}. \quad (2.72)$$

On obtient alors $\mathring{\mathbf{Z}}^*\mathring{\mathbf{Z}}$ la matrice de covariance empirique des signaux d'intérêts, $\check{\mathbf{Z}}^*\check{\mathbf{Z}}$ la matrice de covariance empirique des autres signaux et $\mathring{\mathbf{Z}}^*\check{\mathbf{Z}}$ la matrice d'intercovariance empirique entre les signaux d'intérêts et les autres. On introduit la matrice de précision Θ qui correspond à l'inverse de la matrice de covariance. Selon la formule d'inversion par bloc [201], on peut lier les éléments de la matrice Θ aux éléments de la matrice de covariance \mathbf{K} et qui s'écrivent :

$$\Theta = \mathbf{K}^{-1} = \begin{pmatrix} \mathring{\mathbf{Z}}^*\mathring{\mathbf{Z}} & \mathring{\mathbf{Z}}^*\check{\mathbf{Z}} \\ \check{\mathbf{Z}}^*\mathring{\mathbf{Z}} & \check{\mathbf{Z}}^*\check{\mathbf{Z}} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{E}^{-1} & -\mathbf{E}^{-1}\mathbf{G} \\ -\mathbf{F}\mathbf{E}^{-1} & (\check{\mathbf{Z}}^*\check{\mathbf{Z}})^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{G} \end{pmatrix}. \quad (2.73)$$

Avec :

$$\mathbf{E} = \check{\mathbf{Z}}^* \check{\mathbf{Z}} - \check{\mathbf{Z}}^* \check{\mathbf{Z}} (\check{\mathbf{Z}}^* \check{\mathbf{Z}})^{-1} \check{\mathbf{Z}}^* \check{\mathbf{Z}}, \quad (2.74)$$

$$\mathbf{F} = (\check{\mathbf{Z}}^* \check{\mathbf{Z}})^{-1} \check{\mathbf{Z}}^* \check{\mathbf{Z}}, \quad (2.75)$$

$$\mathbf{G} = \check{\mathbf{Z}}^* \check{\mathbf{Z}} (\check{\mathbf{Z}}^* \check{\mathbf{Z}})^{-1}. \quad (2.76)$$

À partir de cette formule, deux remarques peuvent être faites, la première porte sur la matrice \mathbf{E} . En effet, grâce à la formulation de $\mathbb{E}\{\epsilon_i \epsilon_j\}$ donné dans l'Équation (2.71) on a :

$$E_{ij} = \mathbf{z}_i^* \mathbf{z}_j - \mathbf{z}_i^* \check{\mathbf{Z}} (\check{\mathbf{Z}}^* \check{\mathbf{Z}})^{-1} \check{\mathbf{Z}}^* \mathbf{z}_j, \quad (2.77)$$

$$= \mathbb{E}\{\epsilon_i \epsilon_j\}. \quad (2.78)$$

La seconde porte sur le fait que ce n'est pas directement la matrice \mathbf{E} qui est contenue dans la matrice de précision mais son inverse. Comme \mathbf{E} est une matrice carrée de dimension 2, on a simplement :

$$\mathbf{E} = \det(\mathbf{E}) \begin{pmatrix} \theta_{jj} & -\theta_{ij} \\ -\theta_{ij} & \theta_{ii} \end{pmatrix}. \quad (2.79)$$

Alors le calcul de la corrélation partielle peut se faire à partir des éléments de la matrice de précision telle que :

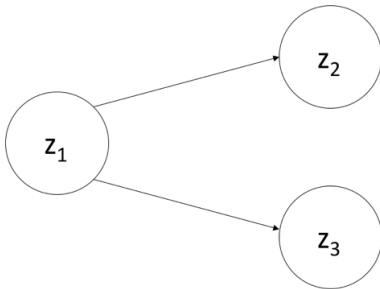
$$p\rho_{ij} = \frac{E_{ij}}{\sqrt{E_{ii}E_{jj}}} = -\frac{\theta_{ij}}{\sqrt{\theta_{ii}\theta_{jj}}}. \quad (2.80)$$

Illustration

On illustre la corrélation partielle de manière similaire aux travaux de [58], l'objectif étant de montrer les différences principales entre cette mesure et la corrélation classique sur un modèle simple. En effet, il ne s'agit pas d'une mesure préférable à la corrélation dans le cas multivarié, mais qui permet d'observer des relations différentes. On considère ici trois signaux $\mathbf{z} \in \mathbb{R}^{3 \times N}$, avec $N = 10000$. Ces trois signaux sont des réalisations du processus suivant :

$$\mathbf{z}_1 = \eta_1, \quad \mathbf{z}_2 = \frac{5\mathbf{z}_1 + \eta_2}{6}, \quad \mathbf{z}_3 = \frac{5\mathbf{z}_1 + \eta_3}{6}, \quad (2.81)$$

avec $\eta_i \simeq \mathcal{N}(0, 1)$ un vecteur de N échantillons, généré à partir de tirages indépendants selon une loi normale. On explicite les relations entre ces trois signaux grâce au graphe représenté dans la Figure 2.6. Une flèche entre le signal \mathbf{z}_i et en direction du signal \mathbf{z}_j indique que le signal \mathbf{z}_i génère en partie le signal \mathbf{z}_j . La corrélation ρ et la corrélation partielle $p\rho$ sont fournies à côté :



$$\rho_{12} = 0.98, \quad p\rho_{12} = 0.7,$$

$$\rho_{13} = 0.98, \quad p\rho_{13} = 0.7,$$

FIGURE 2.6 – Graphe associé au premier processus

$$\rho_{23} = 0.96, \quad p\rho_{23} = -0.02.$$

On marque en rouge les principales différences entre les deux résultats obtenus. On remarque que, en connaissance du signal \mathbf{z}_1 , la corrélation partielle entre les signaux \mathbf{z}_2 et \mathbf{z}_3 devient quasi nulle. Ce

résultat est logique, comme le signal 1 a permis de générer les signaux 2 et 3, les signaux conditionnés ϵ_2 et ϵ_3 sont alors uniquement composés respectivement des bruits η_2 et η_3 . Il s'agit du résultat principal permettant d'illustrer la suppression de corrélations peu pertinentes grâce à la corrélation partielle.

Cependant, la corrélation partielle n'a pas pour conséquence que de supprimer des connexions. On considère le cas suivant :

$$\mathbf{z}_1 = \frac{5\mathbf{z}_2 + 5\mathbf{z}_3 + \eta_1}{11}, \quad \mathbf{z}_2 = \eta_2, \quad \mathbf{z}_3 = \eta_3. \quad (2.82)$$

On explicite les relations entre ces trois signaux grâce au graphe représenté dans la Figure 2.7, et les résultats sont présentés à coté :

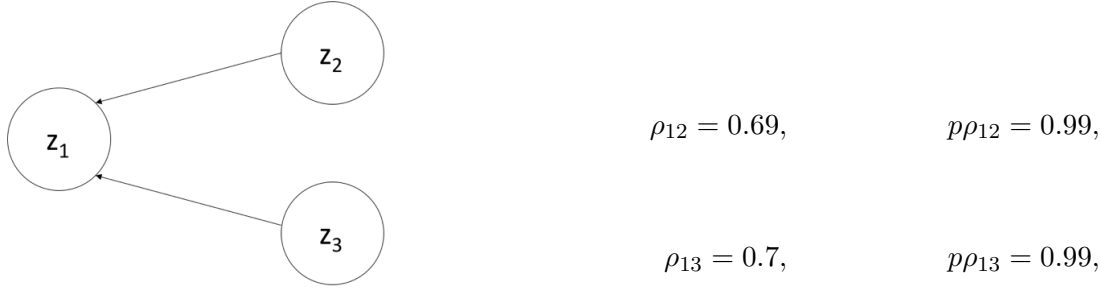


FIGURE 2.7 – Graphe associé au second processus

Ici la connaissance du signal \mathbf{z}_1 permet de créer une connexion entre les signaux \mathbf{z}_2 et \mathbf{z}_3 . Des signaux qui a priori sont indépendants, interagissent ensemble pour former le signal 1, expliquant la forte corrélation partielle entre les signaux 2 et 3.

2.3.2 Conditionnement de la formule des moments phases-amplitudes

Un conditionnement de la formule de l'Équation (2.24) est possible. Les signaux ϵ_i et ϵ_j définis par l'Équation (2.69) sont les résidus des signaux analytiques \mathbf{z}_i et \mathbf{z}_j fournis grâce à la connaissance des autres signaux \mathbf{Z} . Les éléments de ces résidus suivent une distribution gaussienne complexe circulaire et symétrique bivariée [201] telle que :

$$p_{\epsilon_i, \epsilon_j}(\epsilon_i, \epsilon_j) = \frac{1}{\pi^2 \det(\mathbf{E})} e^{-[\epsilon_i^*, \epsilon_j^*] \mathbf{E}^{-1} [\epsilon_i, \epsilon_j]^t}. \quad (2.83)$$

$\mathbf{E} \in \mathbb{R}^{2 \times 2}$ est la matrice de covariance empirique du signal ϵ . La formule des moments de l'Équation (2.24) peut alors s'appliquer aux signaux ϵ_i et ϵ_j , fournissant des relations phases amplitudes en connaissance des autres signaux $\check{\mathbf{Z}}$. Pour simplifier les calculs, on a observé à partir de l'Équation (2.80) qu'il était possible d'écrire les éléments de la matrice \mathbf{E} en fonction de la matrice de précision $\Theta = \mathbf{K}^{-1}$. On a donc :

$$\det(\mathbf{E}) = \frac{1}{\theta_{ii}\theta_{jj} - \theta_{ij}^2}, \quad (2.84)$$

$$\mathbf{E} = \frac{1}{\theta_{ii}\theta_{jj} - \theta_{ij}^2} \begin{pmatrix} \theta_{jj} & -\theta_{ij} e^{i\nu_{ij}} \\ -\theta_{ij} e^{-i\nu_{ij}} & \theta_{ii} \end{pmatrix} = \begin{pmatrix} \frac{\theta_{jj}}{\theta_{ii}\theta_{jj} - \theta_{ij}^2} & \frac{\theta_{ij}}{\theta_{ii}\theta_{jj} - \theta_{ij}^2} e^{i(\nu_{ij} + \pi)} \\ \frac{\theta_{ij}}{\theta_{ii}\theta_{jj} - \theta_{ij}^2} e^{-i(\nu_{ij} + \pi)} & \frac{\theta_{ii}}{\theta_{ii}\theta_{jj} - \theta_{ij}^2} \end{pmatrix}. \quad (2.85)$$

Ici, les éléments de la matrice \mathbf{E} sont écrits en coordonnées polaires en fonction des éléments de la matrice de précision. On obtient ainsi la même représentation que la matrice de covariance bivariée analytique \mathbf{K} de l'Équation (2.21). Alors la même démonstration que celle décrite dans la Section 2.2 peut être appliquée en remplaçant les éléments de la matrice \mathbf{K} par ceux de la matrice \mathbf{E} . On peut ainsi proposer un conditionnement de la formule des moments phases-amplitudes de l'Équation (2.24).

Proposition 6 Soit un signal multivarié gaussien centré et stationnaire, dont les éléments à un instant quelconque s'écrivent $s_i = a_i \cos(\phi_i)$, $\forall i \in \{1, \dots, I\}$ et Θ la matrice de précision analytique associée à ce processus définie dans l'Équation (2.73). Pour tout entiers $(m, p, q) \in (\mathbf{N})^3$, tel que $\gamma^+ = 2 + m + p + q$, $\gamma^- = 2 + m - p + q$ et sous la condition que $\gamma^- > 0$, la formule des moments phases-amplitudes entre les signaux i et j conditionnée aux autres signaux s'écrit :

$$\mathbb{E}_{a_i, a_j, \phi_i, \phi_j} \left\{ (a_i a_j)^m \left(\frac{a_i}{a_j} \right)^p e^{iq(\phi_i - \phi_j)} \mid \check{\mathbf{Z}} \right\} = (\theta_{ii} \theta_{jj} - \theta_{ij}^2) e^{iq(\nu_{ij} + \pi)} \left(\frac{1}{\sqrt{\theta_{ii} \theta_{jj}}} \right)^{m+2} \left(\sqrt{\frac{\theta_{ii}}{\theta_{jj}}} \right)^p \left(\frac{\theta_{ij}}{\sqrt{\theta_{ii} \theta_{jj}}} \right)^q \frac{\Gamma\left(\frac{\gamma^+}{2}\right) \Gamma\left(\frac{\gamma^-}{2}\right)}{\Gamma(q+1)} {}_2F_1\left(\frac{\gamma^+}{2}, \frac{\gamma^-}{2}; q+1; \frac{\theta_{ij}^2}{\theta_{ii} \theta_{jj}}\right). \quad (2.86)$$

Pour certaines valeurs des paramètres (m, p, q) , on obtient une liste de formules qui sont présentées dans le Tableau 2.2. On précise une limite notable de la formule de l'Équation (2.86). En effet, on obtient ici des formules de conditionnement selon les autres signaux $\check{\mathbf{Z}}$, c'est-à-dire en considérant à la fois leurs phases et leurs amplitudes. Or, dans le cas de la PLV par exemple, qui ne considère que des relations entre la phase des signaux, on pourrait espérer une formule qui conditionne uniquement selon la phase des autres signaux. Malheureusement, même dans le cas gaussien, des solutions analytiques dans ce contexte sont souvent impossibles [147]. Cependant cela permet tout de même de fournir une formule de PLV et d'AEC conditionnelle via la proposition suivante :

Proposition 7 Sous hypothèse qu'à chaque instant le signal \mathbf{S} est une réalisation indépendante d'une loi multivariée Gaussienne alors la PLV, l'AEC entre les signaux i et j conditionnées aux autres signaux analytiques $\check{\mathbf{Z}}$ sont des fonctions monotones de la corrélation partielle $p\rho$ qui se notent pPLV et pAEC et s'écrivent :

$$\text{pPLV} = \frac{1}{p\rho} (\mathcal{E}(p\rho) - (1 - p\rho^2)\mathcal{K}(p\rho)), \quad (2.87)$$

$$\text{pAEC} = \frac{1}{4 - \pi} (4\mathcal{E}(p\rho) - 2(1 - p\rho^2)\mathcal{K}(p\rho) - \pi). \quad (2.88)$$

On précise que la corrélation partielle $p\rho$ entre les signaux i et j peut être exprimée à partir de la matrice de précision analytique associée à ce processus Θ [134] telle que :

$$p\rho = -\frac{\theta_{ij}}{\sqrt{\theta_{ii} \theta_{jj}}}. \quad (2.89)$$

Remarque : La pPLV et la pAEC sont des mesures proches de la corrélation partielle dans le contexte de signaux gaussiens. Cependant, ces formules sont des fonctions des éléments de la matrice de précision analytique. Dans un contexte plus général (non gaussien), il s'agit donc de mesures différentes de la corrélation partielle.

2.4 Conclusion

Ce chapitre présente l'expression des moments d'ordres supérieurs entre la phase et l'amplitude de deux signaux gaussiens, en fonction des éléments de leur matrice de covariance analytique. Ce résultat est plus général que les formules déjà existantes car il intègre en plus le ratio des amplitudes. Il montre aussi, que dans le cas gaussien, les mesures de CFs comme la PLV ou l'AEC entre deux signaux sont des fonctions monotones de leur corrélation. Après avoir introduit le concept de corrélation partielle, nous proposons un conditionnement de la formule générale. Cette extension peut être utile pour certains modèles comme le lasso graphique, qui suppose aussi que les signaux suivent une distribution gaussienne.

Dans le chapitre suivant, nous allons proposer deux approches pour faire de la PLV partielle et régularisée. L'une de ces approches se base sur les travaux de ce chapitre.

No.	Valeurs de [m, p, q]	Espérance	Cas classique avec $p\rho = \frac{\theta_{ij}}{\sqrt{\theta_{ii}\theta_{jj}}}$
1	[1,0,0]	$\mathbb{E}_{a_i, a_j} \left\{ a_i a_j \mid \check{\mathbf{Z}} \right\}$	$\frac{\sqrt{\theta_{ii}\theta_{jj}}}{\det(\Theta)} \left(\mathcal{E}(p\rho) - \frac{1}{2}(1 - p\rho^2)\mathcal{K}(p\rho) \right)$
2	[2,0,0]	$\mathbb{E}_{a_i, a_j} \left\{ (a_i a_j)^2 \mid \check{\mathbf{Z}} \right\}$	$\frac{\theta_{ii}\theta_{jj} + \theta_{ij}^2}{\det(\Theta)^2}$
3	[-1,0,0]	$\mathbb{E}_{a_i, a_j} \left\{ \frac{1}{a_i a_j} \mid \check{\mathbf{Z}} \right\}$	$\frac{\det(\Theta)}{\sqrt{\theta_{ii}\theta_{jj}}} \mathcal{K}(p\rho)$
4	[0,1,0]	$\mathbb{E}_{a_i, a_j} \left\{ \frac{a_i}{a_j} \mid \check{\mathbf{Z}} \right\}$	$\sqrt{\frac{\theta_{ii}}{\theta_{jj}}} \mathcal{E}(p\rho)$
5	[0,0,1]	$\mathbb{E}_{\phi_i, \phi_j} \left\{ e^{i(\phi_i - \phi_j)} \mid \check{\mathbf{Z}} \right\}$	$\frac{1}{p\rho} (\mathcal{E}(p\rho) - (1 - p\rho^2)\mathcal{K}(p\rho)) e^{i(\nu_{ij} + \pi)}$
6	[0,0,2]	$\mathbb{E}_{\phi_i, \phi_j} \left\{ e^{i2(\phi_i - \phi_j)} \mid \check{\mathbf{Z}} \right\}$	$\frac{1}{p\rho} ((1 - p\rho)\log(1 - p\rho) + p\rho) e^{i2(\nu_{ij} + \pi)}$
7	[-1,0,1]	$\mathbb{E}_{a_i, a_j, \phi_i, \phi_j} \left\{ \frac{e^{i(\phi_i - \phi_j)}}{a_i a_j} \mid \check{\mathbf{Z}} \right\}$	$\frac{-\det(\Theta) e^{i(\nu_{ij} + \pi)}}{\sqrt{\theta_{ii}\theta_{jj}}} \log(1 - p\rho^2)$
8	[1,0,1]	$\mathbb{E}_{a_i, a_j, \phi_i, \phi_j} \left\{ a_i a_j e^{i(\phi_i - \phi_j)} \mid \check{\mathbf{Z}} \right\}$	$\frac{\theta_{ij} e^{i(\nu_{ij} + \pi)}}{\det(\Theta)}$
9	[1,1,0]	$\mathbb{E}_{a_i} \left\{ a_i^2 \mid \check{\mathbf{Z}} \right\}$	$\frac{\theta_{ii}}{\det(\Theta)}$

Tableau 2.2 – Réécriture de la formule générale conditionnée de l'Équation (2.86) pour des valeurs particulières des paramètres (m, p, q)

Annexe

2.A Démonstrations

2.A.1 Démonstration de l'Équation (2.58)

$$\begin{aligned}
 \mathbb{E}_\psi \{ e^{im\psi} \mid v \} &= \int_0^{2\pi} \frac{e^{im\psi}}{2\pi I_0(\beta v)} e^{\beta v \cos(\psi-\mu)} d\psi, \\
 &= \int_0^{2\pi} \frac{e^{im(\psi+\mu)}}{2\pi I_0(\beta v)} e^{\beta v \cos(\psi)} d\psi, \\
 &= \frac{e^{im\mu}}{2\pi I_0(\beta v)} \int_0^{2\pi} [\cos(m\psi) + \sin(m\psi)] e^{\beta v \cos(\psi)} d\psi, \\
 &= \frac{e^{im\mu}}{2\pi I_0(\beta v)} \left[\int_0^{2\pi} \cos(m\psi) e^{\beta v \cos(\psi)} d\psi - \int_{-\pi}^{\pi} \sin(m\psi) e^{-\beta v \cos(\psi)} d\psi \right], \quad (2.90) \\
 &= \frac{I_m(\beta v)}{I_0(\beta v)} e^{im\mu}. \quad (2.91)
 \end{aligned}$$

Précisons que l'Équation (2.90) est constituée de deux intégrales. La première correspond à la définition d'une fonction de Bessel de première espèce définie dans l'Équation (2.1). La seconde intégrale est obtenue après un changement de variable simple $\psi \leftarrow \psi - \pi$ et permet de mettre en avant l'intégration d'une fonction impaire dont le résultat est nul.

2.A.2 Démonstration de l'Équation (2.60)

$$\begin{aligned}
 \mathbb{E}_w \{ e^{lw} \mid v \} &= \int \frac{e^{lw}}{2K_0(\alpha v)} e^{\alpha v \cosh(w-\bar{\mu})} dw, \\
 &= \int \frac{e^{l(w+\bar{\mu})}}{2K_0(\alpha v)} e^{\alpha v \cosh(w)} dw, \\
 &= \frac{e^{l\bar{\mu}}}{2K_0(\alpha v)} \int [\cosh(lw) + \sinh(lw)] e^{\alpha v \cosh(w)} dw, \\
 &= \frac{e^{l\bar{\mu}}}{2K_0(\alpha v)} \left[\int \cosh(lw) e^{\alpha v \cosh(w)} dw + \int \sinh(w) e^{-\alpha v \cosh(w)} dw \right], \quad (2.92) \\
 &= \frac{K_l(\alpha v)}{K_0(\alpha v)} e^{l\bar{\mu}}. \quad (2.93)
 \end{aligned}$$

Précisons que l'Équation (2.92) est constituée de deux intégrales. La première correspond à la définition d'une fonction de Bessel de seconde espèce définie dans l'Équation (2.2). La seconde est celle d'une fonction impaire dont le résultat est nul.

Chapitre 3

Estimation de la Phase-Locking-Value partielle et régularisée

Une manière classique d’identifier les CFs est de mesurer la ressemblance de phase entre chaque paire de signaux via la PLV [192]. Cependant, la PLV est une mesure bivariée qui ne prend pas en compte l’existence des autres signaux analysés. Afin d’ignorer les couplages potentiels indirects et les dépendances linéaires cachées, une analyse du conditionnement de la PLV est proposée dans ce chapitre. On montre son application sur des signaux sEEG. La mesure conditionnée de la PLV est notée pPLV (pour PLV partielle).

Il existe déjà deux formes de pPLV dans la littérature : dans [170], une pPLV non paramétrique est proposée, permettant d’étendre la mesure empirique de la PLV aux cas multivariés. Dans [13], une pPLV paramétrique est définie, supposant que le signal suit une distribution gaussienne multivariée. Cependant, la formulation proposée est complexe et la première contribution de ce chapitre est d’en soumettre une nouvelle, plus concise à partir des travaux du chapitre précédent.

Comme c’est le cas pour les corrélations partielles, les mesures de pPLV devraient être parcimonieuses car une grande partie des interactions indirectes sont éliminées. Cependant, les estimations de la pPLV proposées jusqu’à présent n’imposent pas ces contraintes aux résultats. Notre deuxième contribution est alors d’adapter le lasso graphique [74], un modèle génératif gaussien initialement destiné à produire des réseaux parcimonieux à partir d’une matrice de corrélation, dans le contexte de la pPLV. Bien que le lasso graphique ait été utilisé sur les matrices de densité spectrale [108], à notre connaissance, il n’a pas encore été étendu aux mesures de synchronie de phases. En fait, nous proposons deux extensions parcimonieuses de la pPLV avec le lasso graphique basé sur l’estimation de la pPLV non paramétrique et paramétrique.

Les CFs, dans le cas de l’épilepsie, sont des mesures dynamiques. Il s’agit d’une série chronologique construite sur des fenêtres temporelles glissantes sur les signaux enregistrés. Dans le contexte du lasso graphique, ce fenêtrage permet de considérer des fenêtres temporelles de courtes durées et de rendre l’hypothèse de signaux stationnaires plus réaliste. Il semble naturel de s’attendre à ce que les estimations des réseaux fonctionnels entre deux instants consécutifs varient peu. C’est pourquoi, nous proposons d’imposer spécifiquement une contrainte de lissage temporel dans les deux procédures d’estimation de la pPLV. On utilise alors la version dynamique du lasso Graphique [90], [60] pour inférer un graphe dynamique de pPLV représentatif de l’évolution temporelle du réseau cortical. Afin d’implémenter la version dynamique du lasso Graphique on opte pour la méthode des multiplicateurs de direction alternée (ADMM) élaborée dans [90].

Dans ce chapitre, nous proposons dans un premier temps de décrire les deux versions existantes de la pPLV dans la Section 3.1 ainsi que, comme contribution, la procédure pour régulariser ces mesures. Ensuite, les mesures de pPLV intégrant le lissage temporel et la parcimonie sont explicitées dans la Section 3.2 permettant de définir une nouvelle mesure de CFs. Dans cette même section, on fournit l’implémentation permettant d’obtenir ces graphes dynamiques de pPLV. Enfin, pour évaluer les performances de nos deux approches, nous exécutons nos algorithmes sur des signaux tests générés par des oscillateurs de Roessler. Ce sont des références habituelles pour modéliser des synchronies de phases entre les signaux. Une discussion sur les résultats de l’expérience pour chaque méthode est fournie

en Section 3.3. Enfin, on termine cette dernière section, en démontrant la pertinence des mesures dynamiques et parcimonieuses de la pPLV pour estimer la connectivité fonctionnelle de signaux iEEG réels enregistrés pour plusieurs crises d'épilepsie d'un patient.

3.1 Phase-Locking-Value conditionnelle

Comme la mesure de PLV définie dans l'Équation (1.10) est bivariable, elle ne peut pas faire la distinction entre les interactions conditionnellement dépendantes et indépendantes de la phase des signaux. La PLV partielle (notée pPLV) assure que les interactions conditionnelles causées pas les dépendances linéaires entre les signaux sont éliminées. Dans cette section, deux formulations régularisées de la pPLV sont proposées. Une version paramétrique faisant l'hypothèse que les signaux suivent une distribution gaussienne multivariée et une version non paramétrique. La régularisation des mesures de pPLV est faite à l'aide du critère du lasso graphique.

3.1.1 Le lasso graphique

On utilise le cadre du lasso graphique pour construire les estimations de la pPLV régularisée. La méthode du lasso graphique permet d'obtenir des mesures de corrélation partielle régularisée. Elle a aussi été introduite pour estimer une matrice de précision dans le cas où la matrice de covariance est singulière. Il s'agit à la base d'une régularisation ℓ_1 mais qui peut se généraliser à d'autres termes de régularisation. Elle est définie comme suit :

Definition 8 : lasso graphique

On considère le signal $\mathbf{S} \in \mathbb{R}^{I \times N}$ stationnaire et réel ayant une distribution gaussienne multivariée centrée, la matrice de covariance empirique associée à ce signal s'écrit :

$$\hat{\mathbf{K}}^{standard} = \frac{1}{N} \mathbf{S} \mathbf{S}^t, \quad (3.1)$$

L'objectif du lasso graphique est d'estimer une matrice de précision régularisée $\Theta \in \mathbb{R}^{I \times I}$. Ceci est réalisé en minimisant la fonction de log-vraisemblance négative et régularisée de la distribution gaussienne multivariée centrée [74] :

$$\operatorname{argmin}_{\Theta \in \mathbb{S}_{++}^p} -\ln(\det(\Theta)) + \operatorname{Tr}(\Theta \hat{\mathbf{K}}^{standard}) + 2\lambda \operatorname{Reg}(\Theta). \quad (3.2)$$

λ est le paramètre associé à la fonction de régularisation convexe $\operatorname{Reg}(\bullet)$ et \mathbb{S}_{++}^p est l'ensemble des matrices semi-définies positives. La matrice de précision régularisée minimisant ce critère est notée $\hat{\Theta}^{standard}$. Alors la corrélation partielle régularisée se note :

$$p\rho(\lambda) = -\frac{\hat{\theta}_{ij}^{standard}}{\sqrt{\hat{\theta}_{ii}^{standard} \hat{\theta}_{jj}^{standard}}}. \quad (3.3)$$

Démonstration : On redémontre ici le résultat connu de l'Équation (3.2), dans l'objectif de comprendre comment ce critère est obtenu à partir de la distribution du signal \mathbf{S} . On considère dans un premier temps la formule de Bayes exprimant la distribution de la matrice de précision régularisée Θ conditionnée à un instant du signal mesuré $\mathbf{s}(n) \sim \mathbf{s}$.

$$p(\Theta | \mathbf{s}) = \frac{p(\mathbf{s} | \Theta)p(\Theta)}{p(\mathbf{s})}. \quad (3.4)$$

On sait que \mathbf{s} est une réalisation d'une loi gaussienne multivariée centrée dont le seul paramètre est une matrice de covariance. Alors, si la matrice de précision Θ est connue (on rappelle que cette matrice correspond à l'inverse de la matrice de covariance), on a :

$$p(\mathbf{s} | \Theta) = \left(\frac{\det(\Theta)}{(2\pi)^N} \right)^{\frac{1}{2}} e^{-\frac{1}{2} \mathbf{s}^t \Theta \mathbf{s}}. \quad (3.5)$$

Au contraire, si la matrice de précision n'est pas connue on ne peut pas faire d'a priori sur la distribution de $p(\mathbf{s})$. On considère alors toutes les mesures \mathbf{s} comme également probables et on fixe $p(\mathbf{s}) = 1$. Finalement, comme on privilégie une solution régularisée de notre matrice de précision, on définit la distribution de probabilité suivante :

$$p(\Theta) = \frac{e^{-\lambda \mathbf{Reg}(\Theta)}}{C_\lambda}, \quad \text{avec} \quad C_\lambda = \int e^{-\lambda \mathbf{Reg}(\Theta)} d\Theta. \quad (3.6)$$

Alors la fonction de log-vraisemblance de la loi de probabilité $p(\Theta | \mathbf{s})$ s'écrit :

$$\begin{aligned} \log vr &= \ln \left(\prod_{n=1}^N p(\Theta | \mathbf{s}(n)) \right), \\ &= \ln \left(\prod_{n=1}^N p(\mathbf{s}(n) | \Theta) p(\Theta) \right), \\ &= \ln \left[\prod_{n=1}^N \frac{1}{C_\lambda} \left(\frac{\det(\Theta)}{(2\pi)^N} \right)^{\frac{1}{2}} e^{-\frac{1}{2} \mathbf{s}^t(n) \Theta \mathbf{s}(n) - \lambda \mathbf{Reg}(\Theta)} \right], \\ &= \sum_{n=1}^N \left[-\ln(C_\lambda (2\pi)^{\frac{N}{2}}) + \frac{1}{2} \ln(\det(\Theta)) - \frac{1}{2} \mathbf{s}^t(n) \Theta \mathbf{s}(n) - \lambda \mathbf{Reg}(\Theta) \right], \\ &= -N \ln(C_\lambda (2\pi)^{\frac{N}{2}}) + \frac{N}{2} \ln(\det(\Theta)) - \frac{1}{2} \text{Tr} \left[\Theta \left(\sum_{n=1}^N \mathbf{s}(n) \mathbf{s}^t(n) \right) \right] - N \lambda \mathbf{Reg}(\Theta), \\ &= -N \ln(C_\lambda (2\pi)^{\frac{N}{2}}) + \frac{N}{2} \ln(\det(\Theta)) - \frac{1}{2} \text{Tr}(\Theta \mathbf{S}^t \mathbf{S}) - N \lambda \mathbf{Reg}(\Theta), \\ &= -N \ln(C_\lambda (2\pi)^{\frac{N}{2}}) + \frac{N}{2} \left[\ln(\det(\Theta)) - \text{Tr}(\Theta \hat{\mathbf{K}}^{standard}) - 2\lambda \mathbf{Reg}(\Theta) \right]. \end{aligned} \quad (3.7)$$

Maximiser la fonction de log-vraisemblance $\log vr$ est alors équivalent à minimiser le critère de l'Équation (3.2) sachant que $-N \ln(C_\lambda (2\pi)^{\frac{N}{2}})$ est une constante ne dépendant pas de Θ . \square

Comme montré dans [105], la corrélation partielle peut être modifiée en remplaçant dans l'Équation (3.2) la matrice de covariance empirique par la matrice de corrélation empirique. Cette modification est particulièrement intéressante dans le cas où des différences d'amplitudes entre les signaux de \mathbf{S} existent (des signaux qui en moyenne ont des amplitudes plus élevées que d'autres). Dans ce cas on note :

$$\left(\hat{\mathbf{K}}^{standard} \right)_{ij} = \frac{\hat{\mathbf{k}}_{ij}^{standard}}{\sqrt{\hat{\mathbf{k}}_{ii}^{standard} \hat{\mathbf{k}}_{jj}^{standard}}}, \quad (3.8)$$

la matrice de corrélation empirique associée aux signaux \mathbf{S} . Ainsi, la matrice de précision régularisée Θ minimisant le critère de l'Équation (3.2) quand la matrice de covariance empirique $\hat{\mathbf{K}}^{standard}$ est remplacée par la matrice de corrélation empirique $\hat{\mathbf{K}}^{standard}$ est notée $\hat{\Theta}^{standard}$. C'est cette matrice qui permet de calculer la corrélation partielle modifiée notée :

$$\overline{p\rho}(\lambda) = -\frac{\hat{\theta}_{ij}^{standard}}{\sqrt{\hat{\theta}_{ii}^{standard} \hat{\theta}_{jj}^{standard}}}. \quad (3.9)$$

3.1.2 Phase-Locking-Value partielle paramétrique

Comme montré dans la Proposition 7, une mesure de la pPLV paramétrique sous hypothèse que les signaux sont gaussiens peut être déduite. Nous proposons désormais une version régularisée de la pPLV paramétrique.

Proposition 8 *On considère le signal $\mathbf{S} \in \mathbb{R}^{I \times N}$ stationnaire et réel ayant une distribution gaussienne multivariée centrée. On note par $\mathbf{Z} \in \mathbb{C}^{I \times N}$ la représentation analytique du signal \mathbf{S} . La matrice de covariance analytique empirique associée à ce signal s'écrit :*

$$\hat{\mathbf{K}}^{\text{analytique}} = \frac{1}{N} \mathbf{Z} \mathbf{Z}^*, \quad (3.10)$$

Une matrice de précision régularisée peut être obtenue en minimisant la fonction de log-vraisemblance négative et régularisée de la distribution gaussienne complexe circulaire multivariée centrée [74] :

$$\operatorname{argmin}_{\Theta \in \mathbb{S}_{++}^p} -\ln(\det(\Theta)) + \operatorname{Tr}(\Theta \hat{\mathbf{K}}^{\text{analytique}}) + \lambda \operatorname{Reg}(\Theta). \quad (3.11)$$

λ est le paramètre associé à la fonction de régularisation convexe $\operatorname{Reg}(\bullet)$ et \mathbb{S}_{++}^p est l'ensemble des matrices semi-définies positives. La matrice de précision régularisée minimisant ce critère est notée $\hat{\Theta}^{\text{analytique}} \in \mathbb{C}^{I \times I}$. Alors par application de la Proposition 7, la pPLV paramétrique régularisée se note :

$$\text{pPLV}^{\text{param}}(\lambda) = \mathcal{F} \left(\frac{|\hat{\theta}_{ij}^{\text{analytique}}|}{\sqrt{\hat{\theta}_{ii}^{\text{analytique}} \hat{\theta}_{jj}^{\text{analytique}}}} \right), \quad (3.12)$$

avec $\mathcal{F}(\bullet)$ une fonction faisant intervenir les intégrales elliptiques des deux premières espèces qui sont explicitées dans les Équations (2.3) et (2.3) :

$$\mathcal{F}(x) = \frac{1}{x} \left(\mathcal{E}(x) - (1-x^2) \mathcal{K}(x) \right). \quad (3.13)$$

Démonstration : On justifie ici l'utilisation du critère de l'Équation (3.11) pour obtenir une matrice de précision régularisée. Il s'agit d'une démonstration très similaire à la Définition 8. Elle se base sur la fonction de log-vraisemblance de la loi $p(\Theta | \mathbf{z})$, soit la probabilité d'obtenir la matrice de précision Θ connaissant \mathbf{z} à l'instant du signal analytique mesuré. L'application du théorème de Bayes donné dans l'Équation (3.4) revient à définir les lois de probabilités $p(\mathbf{z} | \Theta)$, $p(\mathbf{z})$ et $p(\Theta)$. Comme montré dans [134], la représentation analytique d'un signal multivarié gaussien centré suit une distribution gaussienne complexe circulaire multivariée centrée avec comme unique paramètre sa matrice de covariance analytique. Cette loi, écrite en fonction de la matrice de précision analytique (à la place de sa matrice de covariance) se note :

$$p(\mathbf{z} | \Theta) = \frac{\det(\Theta)}{\pi^N} e^{-\mathbf{z}^t \Theta \mathbf{z}}. \quad (3.14)$$

Si la matrice de précision n'est pas connue on ne peut pas faire d'a priori sur la distribution de $p(\mathbf{z})$ et on fixe $p(\mathbf{z}) = 1$. Finalement, comme on privilégie une solution régularisée de notre matrice de précision, on utilise de nouveau la loi $p(\Theta)$ grâce à l'Équation (3.6). Alors par application des étapes de calcul menant à l'Équation (3.7), la fonction de log-vraisemblance de la loi de probabilité $p(\Theta | \mathbf{z})$ s'écrit :

$$\log vr = -N \ln(C_\lambda \pi^N) + N \left[\ln(\det(\Theta)) - \operatorname{Tr}(\Theta \hat{\mathbf{K}}^{\text{analytique}}) - \lambda \operatorname{Reg}(\Theta) \right]. \quad (3.15)$$

Alors maximiser la fonction de log-vraisemblance $\log vr$ est équivalent à minimiser le critère de l'Équation (3.11). \square

Comparaison avec la littérature

On précise qu'il a déjà été proposé dans [12] une formule pour calculer la pPLV paramétrique sans régularisation, et qui s'écrit :

$$\text{pPLV}^{\text{papram}}(0)_{ij} = \left| \frac{\pi}{2} (1 - \tilde{\chi}_{ij}) \left[\chi_{ij}^{\frac{3}{2}} {}_2F_1 \left(\frac{3}{4}, \frac{5}{4}, 1, \chi_{ij}^2 \right) + \frac{3}{4} \chi_{ij}^{\frac{5}{2}} {}_2F_1 \left(\frac{5}{4}, \frac{7}{4}, 2, \chi_{ij}^2 \right) \right] \right|, \quad (3.16)$$

Avec

$$\tilde{\chi}_{ij} = \frac{\hat{\theta}_{ii}^{analytique} \hat{\theta}_{jj}^{analytique}}{(\hat{\theta}_{ij}^{analytique})^2}, \text{ et } \chi_{ij} = \left(\frac{(\hat{\theta}_{ij}^{analytique})^2}{2\hat{\theta}_{ii}^{analytique} \hat{\theta}_{jj}^{analytique} - (\hat{\theta}_{ij}^{analytique})^2} \right). \quad (3.17)$$

La formule proposée dans la Proposition 7 et donc plus concise (il s'agit de deux expressions différentes de la même fonction), nous proposons en plus ici une version régularisée de ce score.

Modification utilisant la matrice de corrélation analytique

De la même manière que pour la corrélation partielle régularisée, où une alternative est proposée dans [105] remplaçant la matrice de covariance par la matrice de corrélation empirique pour le calcul de la matrice de précision régularisée. On propose aussi de s'intéresser à cette modification dans le cadre du score $pPLV^{param}(\lambda)$. Dans ce cas, on note la matrice de corrélation empirique analytique associée aux signaux \mathbf{Z} :

$$(\hat{\mathbf{K}}^{analytique})_{ij} = \frac{\hat{\mathbf{k}}_{ij}^{analytique}}{\sqrt{\hat{\mathbf{k}}_{ii}^{analytique} \hat{\mathbf{k}}_{jj}^{analytique}}}. \quad (3.18)$$

Ainsi, la matrice de précision régularisée Θ minimisant le critère de l'Équation (3.11) quand la matrice de covariance empirique $\hat{\mathbf{K}}^{analytique}$ est remplacée par la matrice de corrélation empirique $\hat{\mathbf{K}}^{analytique}$ est notée $\hat{\Theta}^{analytique}$. C'est cette matrice qui permet de calculer la pPLV paramétrique partielle modifiée notée :

$$\overline{pPLV}^{param}(\lambda) = \mathcal{F} \left(\frac{|\hat{\theta}_{ij}^{analytique}|}{\sqrt{\hat{\theta}_{ii}^{analytique} \hat{\theta}_{jj}^{analytique}}} \right). \quad (3.19)$$

3.1.3 Phase-Locking-Value partielle non paramétrique

On propose une mesure non paramétrique de la pPLV régularisée, c'est à dire qu'il n'est plus nécessaire de supposer que les signaux mesurés suivent une distribution multivariée gaussienne.

Proposition 9 *On considère le signal $\mathbf{S} \in \mathbb{R}^{I \times N}$ stationnaire et réel. On note par $\mathbf{Z} \in \mathbb{C}^{I \times N}$ la représentation analytique du signal \mathbf{S} . La représentation polaire d'un élément du signal analytique est $z_i(n) = a_i(n)e^{i\phi_i(n)}$ et fait intervenir la phase instantanée $\phi_i(n)$ du signal \mathbf{s}_i à l'instant n . On introduit alors le signal de phase $\Phi \in \mathbb{C}^{I \times N}$ correspondant au signal \mathbf{Z} dépourvu de ses amplitudes instantanées $a_i(n)$ tel que $\Phi_i(n) = e^{i\phi_i(n)}$. La matrice de covariance empirique associée de ce signal de phase s'écrit :*

$$\hat{\mathbf{K}}^{phase} = \frac{1}{N} \Phi \Phi^*, \quad (3.20)$$

Une matrice de précision régularisée peut alors être obtenue en minimisant la fonction de log-vraisemblance négative et régularisée de la distribution gaussienne complexe circulaire multivariée centrée [74] :

$$\underset{\Theta \in \mathbb{S}_{++}^p}{\operatorname{argmin}} -\ln(\det(\Theta)) + \operatorname{Tr}(\Theta \hat{\mathbf{K}}^{phase}) + \lambda \operatorname{Reg}(\Theta). \quad (3.21)$$

λ est le paramètre associé à la fonction de régularisation convexe $\operatorname{Reg}(\bullet)$ et \mathbb{S}_{++}^p est l'ensemble des matrices semi-définies positives. La matrice de précision régularisée minimisant ce critère est notée $\hat{\Theta}^{phase} \in \mathbb{C}^{I \times I}$ alors la pPLV non-paramétrique régularisée se note :

$$pPLV^{nonpar}(\lambda) = \frac{|\hat{\theta}_{ij}^{phase}|}{\sqrt{\hat{\theta}_{ii}^{phase} \hat{\theta}_{jj}^{phase}}}. \quad (3.22)$$

Justification : Il ne s'agit pas ici d'une démonstration, car le critère de l'Équation (3.21) n'est pas choisi ici dans l'objectif de maximiser la vraisemblance de la distribution $p(\Theta | \Phi)$. En effet, par application du théorème de Bayes donné dans l'Équation (3.4), on peut définir les lois de probabilités $p(\Phi | \Theta)$, $p(\Phi)$ et $p(\Theta)$. Malheureusement, même dans le cas gaussien, il n'existe pas de distribution permettant de décrire la loi $p(\Phi | \Theta)$ sous forme analytique. On fait alors comme si Φ était gaussien. On va cependant montrer que le critère de l'Équation (3.21) possède des propriétés intéressantes afin d'obtenir une matrice de précision régularisée pertinente.

a) *pPLV non-paramétrique sans régularisation :* On montre dans un premier temps que l'application de la Proposition 9 sans régularisation imposée sur la matrice de précision ($\lambda = 0$), correspond exactement à la définition de la pPLV non-paramétrique définie dans [170]. Celle-ci se base sur la définition de la PLV empirique.

Definition 9 : PLV empirique

Le calcul empirique de la PLV définit dans l'Équation (1.10) se note [104] :

$$\text{PLV} = \left| \frac{1}{N} \sum_{n=1}^N e^{i(\phi_i(n) - \phi_j(n))} \right| = |\hat{k}_{ij}^{phase}|. \quad (3.23)$$

On remarque qu'il s'agit ici simplement d'une mesure de covariance des signaux de phase Φ . La pPLV empirique proposée dans [170] applique la même stratégie de conditionnement que pour la covariance partielle correspondant à la Définition 7 sur les signaux de phase.

Definition 10 : pPLV non-paramétrique sans régularisation

Le calcul de la pPLV non-paramétrique [170] se base sur la matrice de précision non régularisée obtenue à partir de la matrice de covariance des signaux de phase. Elle se note :

$$\hat{\Theta}^{Schelter} = (\hat{\mathbf{K}}^{phase})^{-1}, \quad (3.24)$$

alors :

$$\text{pPLV}^{Schelter} = \frac{|\hat{\theta}_{ij}^{Schelter}|}{\sqrt{\hat{\theta}_{ii}^{Schelter} \hat{\theta}_{jj}^{Schelter}}}. \quad (3.25)$$

Il est simple de démontrer qu'il s'agit exactement de la même définition que celle de la pPLV dans la Proposition 9 quand $\lambda = 0$. En effet, dans ce cas, la dérivée de la fonction de coût du critère de l'Équation (3.21) est nulle quand $\hat{\Theta}^{phase} = \hat{\Theta}^{Schelter}$ soit :

$$\text{pPLV}^{nonpar}(0) = \text{pPLV}^{Schelter}. \quad (3.26)$$

b) *Un critère applicable pour toutes matrices semi-définie positives :* La matrice $\hat{\mathbf{K}}^{phase}$ n'est pas la matrice de covariance empirique d'un signal gaussien. Cependant, elle est relative à une matrice de covariance, et est donc semi-définie positive. On peut alors utiliser le critère de l'Équation (3.21) qui reste convexe si la matrice $\hat{\mathbf{K}}^{phase}$ est semi-définie positive avec $\lambda > 0$ (quand $\lambda = 0$ le critère est convexe uniquement si $\hat{\mathbf{K}}^{phase}$ est définie positive) [59].

c) *Un critère avantageux :* Plus généralement, la fonction de coût du critère de l'Équation (3.21) non régularisé notée :

$$D_{kl}(\Theta, \hat{\mathbf{K}}^{phase}) = -\ln(\det(\Theta)) + \text{Tr}(\Theta \hat{\mathbf{K}}^{phase}), \quad (3.27)$$

correspond à une divergence de Bregman [39], [165] entre les matrices Θ et $\hat{\mathbf{K}}^{phase}$ construite à partir de la fonction $u(\Theta) = -\ln(\det(\Theta))$. Dans ce cas, il s'agit de la divergence de Kullback–Leibler qui calcule l'entropie relative entre les distributions gaussiennes centrées $\mathcal{N}(0, \Theta)$ et $\mathcal{N}(0, \mathbf{K}^{phase})$. Cette fonction (avec ou sans addition de régularisation) a pour principal avantage de contraindre la matrice de précision Θ à appartenir au cône des matrices semi-définie positives [36], [165]. Cette propriété

est indispensable pour s'assurer que la mesure de pPLV non-paramétrique correspondra bien à une mesure de CF comprise entre 0 et 1. De plus, cette mesure a l'avantage de ne pas être dépendante de la matrice $(\hat{\mathbf{K}}^{phase})^{-1}$ contrairement à la distance de Frobenius qui est également une divergence de Bregman. En effet, appliquée au contexte des matrices de précision, cette dernière distance permet d'obtenir le critère alternatif à celui proposé dans l'Équation (3.21), et qui s'écrit :

$$D_F(\Theta, \hat{\mathbf{K}}^{phase}) = \| (\hat{\mathbf{K}}^{phase})^{-1} - \Theta \|_F^2. \quad (3.28)$$

Ce critère implique donc que la matrice $\hat{\mathbf{K}}^{phase}$ ne soit pas singulière, ce qui peut être une condition limitante pour des applications réelles. \square

Comme remarque finale sur la pPLV non-paramétrique, on précise que la matrice $\hat{\mathbf{K}}^{phase}$ correspond à une matrice de corrélation. En effet, les signaux de phases sont naturellement normalisés et les éléments diagonaux de la matrice $\hat{\mathbf{K}}^{phase}$ sont unitaires. Il n'est donc pas nécessaire de proposer une alternative à la pPLV non-paramétrique à partir de la matrice de corrélation (contrairement au cas du lasso graphique et de la pPLV paramétrique).

3.1.4 Principale différence entre les pPLV paramétriques et non paramétriques

Hormis l'hypothèse sur les signaux, les deux mesures de pPLV ont une autre différence majeure. En effet elles sont conditionnées à des signaux différents. Pour le cas paramétrique, le conditionnement se fait selon les autres signaux analytiques en reprenant les notations de la Section 2.3.1, on a :

$$\text{pPLV}_{ij}^{param} = \mathbb{E}_{\phi_i, \phi_j} \left\{ e^{i(\phi_i - \phi_j)} \mid \check{\mathbf{Z}} \right\}. \quad (3.29)$$

Tandis que la version non paramétrique réalise un conditionnement selon les autres signaux de phase :

$$\text{pPLV}_{ij}^{nonpar} = \mathbb{E}_{\phi_i, \phi_j} \left\{ e^{i(\phi_i - \phi_j)} \mid \check{\check{\mathbf{Z}}} \right\}, \quad (3.30)$$

avec $\check{\check{\mathbf{Z}}}$ la phase des signaux de la matrice $\check{\mathbf{Z}}$. Les auteurs de [13] ont montré, dans le cas non régularisé et sur un modèle de signaux non-gaussiens, que la version paramétrique produit une estimation avec une variance moindre que la version non paramétrique. Cependant, dans le cas paramétrique, les amplitudes des signaux qui sont contenus dans la matrice $\check{\mathbf{Z}}$ interviendront dans le calcul de la pPLV. Cela peut être limitant dans notre situation, car on utilise une mesure de phase dans l'objectif de ne plus prendre en compte les amplitudes des signaux.

3.2 Application aux mesures de connectivités fonctionnelles

3.2.1 Régularisation pertinente dans le calcul dynamique de CFs

On cherche maintenant à adapter le terme de régularisation $\mathbf{Reg}(\bullet)$ des Équations (3.2), (3.11) et (3.21) pour l'application spécifique de l'estimation dynamique des CFs pour des données neuro-physiologiques. On cherche seulement à se concentrer, dans un contexte clinique, sur les CFs les plus importantes, par conséquent une régularisation promouvant la parcimonie est imposée aux graphes de CFs. Aussi, comme on s'attend à des mesures de CFs qui évoluent de manière continue dans le temps, l'a priori de lissage temporel est ajouté au problème d'optimisation de l'Équation (3.21). Comme le calcul dynamique des CFs (présenté dans la Section 1.2.1) est réalisé sur des fenêtres glissantes de courte durée (avec du recouvrement), cela revient à considérer la série temporelle $\mathbf{S}^{(t)}$ constituée de T segments de N échantillons, i.e. $\{s_n^{(t)}(n), t \in \{1, \dots, T\}\}$ and $n \in \{1, \dots, N\}$. Comme les segments sont recouvrants, deux séries de signaux $\mathbf{S}^{(t)}$ et $\mathbf{S}^{(t+1)}$ seront corrélées et donc conduisent à des mesures de CFs a priori lisses. À partir de cette segmentation, la contrainte de lissage revient à imposer des variations lentes entre les graphes de CFs adjacents.

On obtient alors le critère régularisé suivant.

Proposition 10

Soit une série de matrices de covariance (ou de corrélation) $\{\mathbf{K}^{(1)}, \dots, \mathbf{K}^{(T)}\}$ pouvant être calculée à partir de la série temporelle $\mathbf{S}^{(t)}$, sa représentation analytique $\mathbf{Z}^{(t)}$ ou le signal de phase correspondant $\Phi^{(t)} \forall t \in \{1, \dots, T\}$. L'addition des contraintes de parcimonie et de lissage temporel au critère du graphical lasso des Équations (3.2), (3.11) et (3.21) revient à appliquer le lasso graphique dynamique [90], où la série de matrices de précision $\{\Theta^{(1)}, \dots, \Theta^{(T)}\}$ sont solutions du problème d'optimisation suivant.

$$\underset{\Theta^{(1)}, \dots, \Theta^{(T)} \in \mathbb{S}_{++}^p}{\operatorname{argmin}} \sum_{t=1}^T \left[-\ln(\det(\Theta^{(t)})) + \operatorname{Tr}(\Theta^{(t)} \hat{\mathbf{K}}^{(t)}) + \lambda \|\Theta^{(t)}\|_{od,1} \right] + \gamma \sum_{\tau=1}^{T-1} g(\Theta^{(\tau+1)} - \Theta^{(\tau)}). \quad (3.31)$$

Ici, λ et γ sont respectivement les hyperparamètres des fonctions de régularisation parcimonieuses $\|\Theta\|_{od,1} = \sum_{i,j \neq i} |\Theta_{ij}|$ et du lissage temporel. Pour ce dernier terme, on applique une régularisation group-lasso pour lisser et encourager une segmentation de la série temporelle [90] :

$$g(\Theta) = \sum_{i=1}^I \|\Theta_i\|_F^2; \quad (3.32)$$

3.2.2 Résumé des différentes mesures de CFs introduites

Le Tableau 3.1 récapitule les différentes mesures de CFs que l'on peut obtenir à partir de la Définition 8 et des Propositions 8 et 9 et du critère régularisé de l'Équation (3.31). Elles sont au nombre de 5 avec 3 mesures d'intérêt de pPLV régularisée qui sont : une mesure de pPLV paramétrique $\overline{\text{pPLV}}^{\text{param}}$ définie dans l'Équation (3.12), une mesure de pPLV paramétrique modifiée utilisant la matrice de corrélation analytique $\overline{\text{pPLV}}^{\text{param}}$ définie dans l'Équation (3.19) et enfin une mesure de pPLV non-paramétrique $\overline{\text{pPLV}}^{\text{nonpar}}$ définie dans l'Équation (3.22). Mais aussi 2 mesures de corrélation partielle régularisée : ρ définie dans l'Équation (3.3) et une mesure modifiée utilisant la matrice de corrélation $\overline{\rho}$ définie dans l'Équation (3.9).

3.2.3 Implémentation

Afin d'implémenter la version dynamique du lasso graphique, on opte pour la méthode des multiplicateurs de direction alternée (ADMM) élaborée dans [90]. On s'aide en plus des recommandations de [37] pour définir le critère d'arrêt et la mise à jour du paramètre de l'ADMM. L'algorithme de [90] a été élaboré dans le cadre de mesure de corrélation partielle avec des matrices de covariance \mathbf{K} et de précision Θ réelles. Cependant, pour les trois mesures de pPLV, il s'agit de l'application du critère graphical lasso avec des matrices complexes [187]. La principale contribution apportée à cette implémentation est donc d'adapter l'implémentation de [90] au cas où les matrices \mathbf{K} et Θ sont complexes.

On résume l'ensemble des étapes pour trouver une solution au problème de l'Équation (3.31) dans l'Algorithme 1. Les étapes de cet algorithme sont détaillées dans l'Annexe 3.A.

Voir Prop. ou Def.	Matrice de covariance pour le critère de l'Équation (3.31) $\forall t \in \{1, \dots, T\}$	Matrice de précision solution du critère de l'Équation (3.31) $\forall t \in \{1, \dots, T\}$	Mesure de CF obtenue $\forall t \in \{1, \dots, T\}$
Def. 8	$\hat{\mathbf{K}}^{standard,(t)} = \frac{1}{N} \mathbf{S}^{(t)} \mathbf{S}^{(t)*}$	$\hat{\Theta}^{standard,(t)}$	$pp^{(t)}(\lambda, \gamma)$ $= -\frac{\hat{\theta}_{ij}^{standard,(t)}}{\sqrt{\hat{\theta}_{ii}^{standard,(t)} \hat{\theta}_{jj}^{standard,(t)}}}$
Def. 8	$\left(\hat{\mathbf{K}}^{standard,(t)} \right)_{ij}$ $= \frac{\hat{k}_{ij}^{standard,(t)}}{\sqrt{\hat{k}_{ii}^{standard,(t)} \hat{k}_{jj}^{standard,(t)}}}$	$\hat{\Theta}^{standard,(t)}$	$\overline{pp}^{(t)}(\lambda, \gamma)$ $= -\frac{\hat{\theta}_{ij}^{standard,(t)}}{\sqrt{\hat{\theta}_{ii}^{standard,(t)} \hat{\theta}_{jj}^{standard,(t)}}}$
Prop. 8	$\hat{\mathbf{K}}^{analytique,(t)} = \frac{1}{N} \mathbf{Z}^{(t)} \mathbf{Z}^{(t)*}$	$\hat{\Theta}^{analytique,(t)}$	$pPLV^{param,(t)}(\lambda, \gamma)$ $= \mathcal{F} \left(\frac{ \hat{\theta}_{ij}^{analytique,(t)} }{\sqrt{\hat{\theta}_{ii}^{analytique,(t)} \hat{\theta}_{jj}^{analytique,(t)}}} \right)$
Prop. 8	$\left(\hat{\mathbf{K}}^{analytique,(t)} \right)_{ij}$ $= \frac{\hat{k}_{ij}^{analytique,(t)}}{\sqrt{\hat{k}_{ii}^{analytique,(t)} \hat{k}_{jj}^{analytique,(t)}}}$	$\hat{\Theta}^{analytique,(t)}$	$\overline{pPLV}^{param,(t)}(\lambda, \gamma)$ $= \mathcal{F} \left(\frac{ \hat{\theta}_{ij}^{analytique,(t)} }{\sqrt{\hat{\theta}_{ii}^{analytique,(t)} \hat{\theta}_{jj}^{analytique,(t)}}} \right)$
Prop. 9	$\hat{\mathbf{K}}^{phase,(t)} = \frac{1}{N} \Phi^{(t)} \Phi^{(t)*}$	$\hat{\Theta}^{phase,(t)}$	$pPLV^{nonpar,(t)}(\lambda, \gamma)$ $= \frac{ \hat{\theta}_{ij}^{phase,(t)} }{\sqrt{\hat{\theta}_{ii}^{phase,(t)} \hat{\theta}_{jj}^{phase,(t)}}}$

Tableau 3.1 – Récapitulation des différentes étapes pour obtenir 5 mesures conditionnelles de CFs, les notations sont précisées dans les définitions ou propositions associées à la première colonne.

Algorithm 1 Résout le problème de l'Équation (3.31) grâce à l'ADMM

Require: $\hat{\mathbf{K}}$, les paramètres λ et γ , itération maximale κ_{max} , une première supposition $\Theta_{(0)}$, $\mathbf{B}_{(0)}$, $\mathbf{C}_{(0)}$, $\mathbf{D}_{(0)}$, $\mathbf{U}_{B,(0)}$, $\mathbf{U}_{C,(0)}$, $\mathbf{U}_{D,(0)}$
 $k = 1$
 $\varpi_1 = \varpi_t = 2$ et $\varpi_t = 3 \forall t \in \{2, \dots, T-1\}$
 $\epsilon_R = \epsilon_S = 0$

while $\kappa < \kappa_{max}$ or ($\|\mathbf{ResR}_{(\kappa)}\|_F \leq \epsilon_R$ & $\|\mathbf{ResS}_{(\kappa)}\|_F \leq \epsilon_S$) **do**

for $t = \{1, \dots, T\}$ **do**

$$\mathbf{F}_{(\kappa)}^{(t)} = \mathbf{B}_{(\kappa)}^{(t)} + \mathbf{C}_{(\kappa)}^{(t)} + \mathbf{D}_{(\kappa)}^{(t)} - \mathbf{U}_{B,(\kappa)}^{(t)} - \mathbf{U}_{C,(\kappa)}^{(t)} - \mathbf{U}_{D,(\kappa)}^{(t)}$$

Calculer $\mathbf{W}_{(\kappa)}^{(t)} \Sigma_{(\kappa)}^{(t)} \mathbf{W}_k^{*(t)}$ la SVD de $\left(\frac{3\delta}{\varpi_t} \frac{\mathbf{F}_{(\kappa)}^{(t)} + \mathbf{F}_{(\kappa)}^{*(t)}}{2} - \hat{\mathbf{K}}^{(t)} \right)$.

$$\Theta_{(\kappa+1)}^{(t)} = \frac{\varpi_N}{6\delta} \mathbf{W}_{(\kappa)}^{*(t)} \left(\Sigma_{(\kappa)}^{(t)} + \sqrt{\left(\Sigma_{(\kappa)}^{(t)} \right)^2 + \frac{12\delta}{\varpi_N} \mathbf{I}} \right) \mathbf{W}_{(\kappa)}^{(t)}$$

$$\mathbf{B}_{(\kappa+1)}^{(t)} = \text{ST}_{\frac{\lambda}{\delta}} \left(\Theta_{(\kappa+1)}^{(t)} + \mathbf{U}_{B,(\kappa)}^{(t)} \right), \quad (\text{Avec } \text{ST}_{\frac{\lambda}{\delta}}(\bullet) \text{ précisée dans l'Équation (3.75), et}$$

$\forall i \neq j$)

$$\mathbf{U}_{B,k+1}^{(t)} = \mathbf{U}_{B,(\kappa)}^{(t)} + \Theta_{(\kappa+1)}^{(t)} - \mathbf{B}_{(\kappa+1)}^{(t)}$$

end for

for $\tau = \{1, \dots, T-1\}$ **do**

$$\mathbf{H}_k^{(\tau)} = \Theta_{(\kappa+1)}^{(\tau)} + \mathbf{U}_{C,(\kappa)}^{(\tau)}$$

$$\mathbf{G}_{(\kappa)}^{(\tau+1)} = \Theta_{(\kappa+1)}^{(\tau+1)} + \mathbf{U}_{D,(\kappa)}^{(\tau+1)}$$

$$\begin{pmatrix} \mathbf{C}_{(\kappa+1)}^{(\tau)} \\ \mathbf{D}_{(\kappa+1)}^{(\tau+1)} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \mathbf{G}_{(\kappa)}^{(\tau+1)} + \mathbf{H}_k^{(\tau)} - \text{STB}_{\frac{2\gamma}{\delta}}(\mathbf{G}_{(\kappa)}^{(\tau+1)} - \mathbf{H}_k^{(\tau)}) \\ \mathbf{G}_{(\kappa)}^{(\tau+1)} + \mathbf{H}_k^{(\tau)} + \text{STB}_{\frac{2\gamma}{\delta}}(\mathbf{G}_{(\kappa)}^{(\tau+1)} - \mathbf{H}_k^{(\tau)}) \end{pmatrix} \quad (\text{STB}_{\frac{2\gamma}{\delta}}(\bullet) \text{ voir l'Équation (3.78)})$$

$$\begin{pmatrix} \mathbf{U}_{C,k+1}^{(\tau)} \\ \mathbf{U}_{D,k+1}^{(\tau+1)} \end{pmatrix} = \begin{pmatrix} \mathbf{U}_{C,(\kappa)}^{(\tau)} + \Theta_{(\kappa+1)}^{(\tau)} - \mathbf{C}_{(\kappa+1)}^{(\tau)} \\ \mathbf{U}_{D,(\kappa)}^{(\tau+1)} + \Theta_{(\kappa+1)}^{(\tau+1)} - \mathbf{D}_{(\kappa+1)}^{(\tau+1)} \end{pmatrix}$$

end for

$$\mathbf{ResR}_{(\kappa+1)} = \sum_{t=1}^T \mathbf{B}^{(t)} - \Theta^{(t)} + \sum_{\tau=1}^{T-1} \mathbf{C}^{(\tau)} + \mathbf{D}^{(\tau+1)} - \Theta^{(\tau)} - \Theta^{(\tau+1)}$$

$$\mathbf{ResS}_{(\kappa+1)} = \delta \sum_{t=1}^T \mathbf{B}_{(\kappa+1)}^{(t)} - \mathbf{B}_{(\kappa)}^{(t)} + \delta \sum_{\tau=1}^{T-1} \mathbf{C}_{(\kappa+1)}^{(\tau)} - \mathbf{C}_{(\kappa)}^{(\tau)} + \mathbf{D}_{(\kappa+1)}^{(\tau+1)} - \mathbf{D}_{(\kappa)}^{(\tau+1)}$$

$$\delta = \begin{cases} 2\delta & \text{si } \|\mathbf{ResR}_{(\kappa)}\|_F \geq 10 \|\mathbf{ResS}_{(\kappa)}\|_F \\ \frac{\delta}{2} & \text{si } \|\mathbf{ResS}_{(\kappa)}\|_F \geq 10 \|\mathbf{ResR}_{(\kappa)}\|_F \\ \delta & \text{sinon} \end{cases}$$

$$\kappa = \kappa + 1$$

$$\epsilon_a = \sum_{t=1}^T \|\mathbf{U}_{B,(\kappa+1)}^{(t)}\|_F^2 + \sum_{\tau=1}^{T-1} \|\mathbf{U}_{C,(\kappa+1)}^{(\tau)}\|_F^2 + \|\mathbf{U}_{D,(\kappa+1)}^{(\tau)}\|_F^2$$

$$\epsilon_c = \sum_{t=1}^T \|\mathbf{B}_{(\kappa+1)}^{(t)}\|_F^2 + \sum_{\tau=1}^{T-1} \|\mathbf{C}_{(\kappa+1)}^{(\tau)}\|_F^2 + \|\mathbf{D}_{(\kappa+1)}^{(\tau)}\|_F^2$$

$$\epsilon_d = \sum_{t=1}^T \|\Theta_{(\kappa+1)}^{(t)}\|_F^2 + \sum_{\tau=1}^{T-1} \|\Theta_{(\kappa+1)}^{(\tau)}\|_F^2 + \|\Theta_{(\kappa+1)}^{(\tau)}\|_F^2$$

$$\epsilon_R = 10^{-4} \left(\sqrt{TI^2 + 2(I^2(T-1))} + \sqrt{\max(\epsilon_c, \epsilon_d)} \right)$$

$$\epsilon_S = 10^{-4} \left(\sqrt{TI^2 + 2(I^2(T-1))} + \sqrt{\epsilon_a} \right)$$

end while

3.3 Comparaison sur un modèle des estimateurs de pPLV régularisés

L'objectif de cette section est de comparer sur un modèle mathématique les mesures de pPLV régularisées proposées dans ce chapitre qui sont résumées dans la Section 3.2.2. On observe aussi les résultats obtenus à partir des deux mesures de corrélation partielle régularisée (voir la Section 3.2.2). Le Tableau 3.1 récapitule ces 5 mesures de CFs. Les points majeurs que l'on cherche à éclaircir à partir de cette expérience sont :

- Les régularisations de parcimonie et de cohérence temporelle imposées aux mesures de pPLV sont pertinentes pour obtenir des graphes de CFs interprétables et robustes aux bruits des signaux sEEG.
- L'une des 3 mesures de pPLV proposées est plus adaptée pour mesurer le niveau de synchronie des phases entre chaque paire de signaux.
- Les particularités des mesures de pPLV régularisées, proposées par rapport aux mesures de corrélation partielle déjà existante.

Dans un premier temps, on présente le modèle de signaux sEEG dans la Section 3.3.1. Puis, on met en lumière l'intérêt des régularisations proposées dans la Section 3.3.2. Enfin on discute de l'impact de plusieurs bruits ou transformations (retards, modulation d'amplitude) sur les estimations des mesures de CFs régularisées dans la Section 3.3.3.

3.3.1 Modèle de signaux avec couplage de phase

Modèle

On modélise une série temporelle $\mathbf{S}^{(t)} \in \mathbb{R}^{I \times N}$ avec $t \in \{1, \dots, T\}$. On fixe $I = 3$, le nombre d'échantillon de chaque signal à $N = 500$ et le nombre de segments à $T = 90$. Afin de modéliser les oscillateurs couplés entre les signaux, on propose le modèle d'oscillation de Roessler. Ce modèle est couramment employé pour générer des signaux réalistes avec des phases synchrones ([170], [12], [13]) et il s'écrit :

$$\begin{pmatrix} \frac{d}{dn}(s_i^{(t)}(n)) \\ \frac{d}{dn}(x_i^{(t)}(n)) \\ \frac{d}{dn}(y_i^{(t)}(n)) \end{pmatrix} = \begin{pmatrix} -w_1 y_i^{(t)}(n) + \sigma \eta + \sum_{j \neq i} \epsilon_{ij}^{(t)} (s_i^{(t)}(n) - s_j^{(t)}(n)) \\ w_j s_i^{(t)}(n) + \alpha x_i^{(t)}(n) \\ b + (s_i^{(t)}(n) - c) y_i^{(t)}(n) \end{pmatrix}, \quad (3.33)$$

On fixe les paramètres propres à l'oscillateur de Roessler de manière similaire à ([170], [12], [13]), on a $a = 0.2$, $b = 0.2$, $c = 10$, $w_1 = 1.03$, $w_2 = 1.01$, $w_3 = 0.99$ et $\sigma = 3.5$ et η provenant d'un tirage aléatoire selon une loi normale. Le paramètre essentiel étant le taux de couplage entre les signaux i et j est noté $\epsilon_{ij}^{(t)}$. Si l'on souhaite synchroniser le signal i et j à l'instant (t) , alors $\epsilon_{ij}^{(t)} \in [0.05, 0.3]$ est tiré selon une distribution uniforme, sinon $\epsilon_{ij}^{(t)} = 0$. Afin d'intégrer l'Équation (3.33), on utilise la boîte à outil SDETools de Matlab, avec un intervalle d'échantillonnage de 0.02 pour produire 11000 échantillons. Les 1000 premiers sont supprimés car ils correspondent au régime transitoire. Le signal est sous-échantillonné tous les 20 échantillons afin d'avoir $N = 500$.

La Figure 3.1(a) affiche les réseaux temporels de synchronie de notre modèle, où les taux de couplage ϵ_{13} , ϵ_{23} , ϵ_{12} sont mis à zéro alternativement sur les périodes de temps $t = 1, \dots, 30$, $t = 31, \dots, 60$ et $t = 61, \dots, 90$. L'objectif des mesures de CFs proposées étant de récupérer le graphe de CFs de la Figure 3.1(a) correspondant à chaque segment t . On favorise une représentation des résultats sous forme de matrice de CFs (voir la Section 1.2.2). Ainsi, la Figure 3.1(b) représente la matrice de CFs correspondante à l'évolution idéale des pPLV entre chaque paire de signaux modélisés en fonction de t .

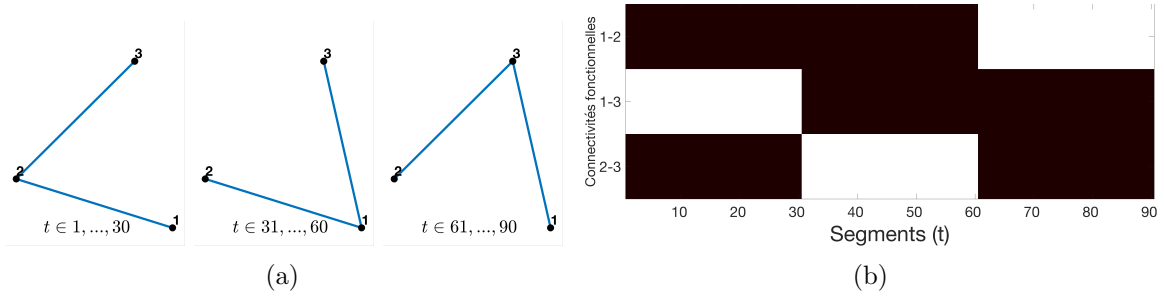


FIGURE 3.1 – (a) Les trois états de couplage en fonction de t . (b) Matrice de CFs idéale (la couleur noire indique la présence d’une synchronie entre les signaux i et j à l’instant t).

Choix des paramètres de la méthode

Il est difficile de comparer deux méthodes différentes utilisant le lasso graphique, car elles sont dépendantes des paramètres sélectionnés (λ, γ). On propose une normalisation des données qui, dans le cadre de nos conditions expérimentales, permet d’obtenir des résultats comparables en utilisant le même jeu de paramètres (λ, γ). Elle est appliquée à toutes matrices de covariance (ou corrélation) empiriques $\hat{\mathbf{K}}^{(t)} \forall t \in \{1, \dots, T\}$ nécessaires pour calculer la matrice de précision empirique (qui sont $\hat{\mathbf{K}}^{standard,(t)}$, $\hat{\mathbf{K}}^{standard,(t)}$, $\hat{\mathbf{K}}^{analytique,(t)}$, $\hat{\mathbf{K}}^{analytique,(t)}$ du Tableau 3.1). La normalisation apportée aux matrices de covariance ou de corrélation est :

$$\hat{\mathbf{K}}^{(t)} \leftarrow \frac{1}{\alpha^{(t)}} \hat{\mathbf{K}}^{(t)}, \quad \alpha^{(t)} = \frac{\sum_{t=1}^T \|\hat{\mathbf{K}}^{(t)}\|_{od,F}^2}{\sum_{t=1}^T \|\hat{\mathbf{K}}^{phase,(t)}\|_{od,F}^2}. \quad (3.34)$$

Avec $\|\mathbf{X}\|_{od,F}^2$ le carré de la norme de Frobenius appliqué seulement sur les éléments non diagonaux de la matrice \mathbf{X} . Cette normalisation permet aux éléments non diagonaux de la matrice $\hat{\mathbf{K}}^{(t)}$ d’avoir en moyenne des valeurs similaires à ceux de la matrice de covariance des signaux de phase $\hat{\mathbf{K}}^{phase,(t)}$. Cette dernière a été choisie car il s’agit d’une matrice de corrélation non affectée par l’amplitude des signaux. Ainsi, l’impact de l’amplitude moyenne des signaux sur chaque matrice de covariance est minimisé, sachant qu’il s’agit du principal facteur influençant les paramètres (λ, γ). On précise que cette normalisation n’est pas nécessaire dans le cas $\hat{\mathbf{K}}^{(t)} = \hat{\mathbf{K}}^{phase,(t)}$, dans ce cas $\alpha^{(t)} = 1 \forall t$. Pour toutes les expériences avec ce modèle, on fixe empiriquement $\lambda = 0.1$ et $\gamma = 1$.

3.3.2 Intérêt des régularisations proposées

Le tableau 3.2 présente deux scores permettant d’évaluer la performance d’estimation de deux mesures de pPLV, les mesures $pPLV^{nonpar}$ et $pPLV^{param}$. Les autres mesures de CFs ne sont pas comparées dans cette partie, l’objectif étant ici de présenter l’intérêt des régularisations proposées. Les deux mesures de pPLV sont alors calculées sans régularisation (notées $pPLV(0, 0)$), avec seulement la régularisation parcimonieuse (notées $pPLV(\lambda, 0)$) et avec à la fois la régularisation temporelle et parcimonieuse (notées $pPLV(\lambda, \gamma)$).

Le premier score considéré est le niveau de fausses détections. L’objectif est d’observer si les mesures de pPLV sont bien faibles ou nulles quand il n’y a pas de couplage entre les signaux i et j (ϵ_{ij}). Ce score se note :

$$\text{SumNeg} = \sum_{t=1}^T \sum_{(i,j) \in \mathcal{S}(t)} pPLV_{ij}^{(t)}. \quad (3.35)$$

Avec $\{\mathcal{S}(t) = (i, j) \mid i > j \ \& \ \epsilon_{ij}^{(t)} = 0\}$ l’ensemble des paires de signaux i, j non synchronisées en fonction des segments t (éléments blancs dans la Figure. 3.1(b)). Pour les deux mesures de pPLV, l’addition de la régularisation réduit le risque de faux positifs. Cependant, comme mentionné dans [105], le lasso graphique souffre de certains biais qui conduisent à une sous-évaluation systématique des valeurs de couplage, rendant la détection des paires où les signaux sont faiblement synchronisés plus

difficiles à détecter. Par conséquent, nous proposons un deuxième score qui mesure le contraste entre les valeurs de pPLV estimées sur les paires couplées (éléments noirs de la Figure. 3.1(b)) versus les paires non couplées :

$$\mathbf{Contrast} = \hat{m}_{\{Pos\}} - 2\hat{\sigma}_{\{Pos\}} - (\hat{m}_{\{Neg\}} + 2\hat{\sigma}_{\{Neg\}}), \quad (3.36)$$

$$\begin{aligned} \hat{m}_{\{Pos\}} &= \frac{1}{2T} \sum_{t=1}^T \sum_{(i,j) \notin \mathcal{S}(t)} pPLV_{ij}^{(t)}, & \hat{\sigma}_{\{Pos\}}^2 &= \frac{1}{2T} \sum_{t=1}^T \sum_{(i,j) \notin \mathcal{S}(t)} (pPLV_{ij}^{(t)} - \hat{m}_{\{Pos\}})^2, \\ \hat{m}_{\{Neg\}} &= \frac{1}{T} \sum_{t=1}^T \sum_{(i,j) \in \mathcal{S}(t)} pPLV_{ij}^{(t)}, & \hat{\sigma}_{\{Neg\}}^2 &= \frac{1}{2T} \sum_{t=1}^T \sum_{(i,j) \in \mathcal{S}(t)} (pPLV_{ij}^{(t)} - \hat{m}_{\{Neg\}})^2. \end{aligned}$$

Ce score mesure l'écart entre les valeurs des deux ensembles en utilisant comme intervalle de confiance, un substitut basé sur la moyenne empirique \hat{m} et l'écart-type $\hat{\sigma}$. Pour les deux méthodes, une bonne mesure de contraste (positive) est seulement obtenue quand les régularisations parcimonieuses et temporelles sont actives ensemble. Il semble que les performances liées aux scores **Contrast** et **SumNeg** de la méthode paramétrique $pPLV^{param}$ sont meilleures que la méthode non-paramétrique $pPLV^{nonpar}$. Ce résultat est en accord avec [13] qui montre que la version paramétrique produit une estimation avec une variance moindre que la version non paramétrique (les auteurs étudient seulement le cas non régularisé).

Method	Score	(0, 0)	(λ , 0)	(λ , γ)
$pPLV^{nonpar}$	SumNeg	14	4.4	1.9
	Contrast	-0.18	-0.15	0.15
$pPLV^{param}$	SumNeg	12.5	2.3	1.2
	Contrast	-0.14	-0.06	0.17

Tableau 3.2 – Performance des méthodes $pPLV^{nonpar}$ et $pPLV^{param}$ sans régularisation (0, 0), avec seulement la parcimonie (λ , 0) et avec la régularisation parcimonieuse et temporelle (λ , γ) sur le modèle de Roessler.

Les Figures 3.2 montrent les matrices de CFs obtenues en fonction des segments t pour les deux mesures de pPLV et pour les 3 configurations : $pPLV(0, 0)$, $pPLV(\lambda, 0)$, $pPLV(\lambda, \gamma)$. La palette de couleurs est mise à l'échelle pour être la même pour toutes les méthodes. Quand aucune régularisation n'est considérée $pPLV(0, 0)$ les mesures de CFs produisent des valeurs pPLV erratiques qui résultent de la variabilité (aléatoire) de $\epsilon_{ij}^{(t)}$. Les Figures 3.2(c)-(d) montrent clairement comment la contrainte de parcimonie diminue la valeur des estimations de la pPLV dans les régions de couplage inactif. De plus, comme mentionné ci-dessus, les biais induits par le lasso graphique abaissent globalement les valeurs estimées. Finalement, le cas $pPLV(\lambda, \gamma)$ qui est affiché dans la Figure 3.2(e)-(f) fournit une matrice de CFs très proche de la matrice de CFs idéale de la Figure 3.1(d). Ceci est une illustration claire des avantages des régularisations conjointes parcimonieuses et temporelles pour l'estimation de la pPLV. Quelques différences sont notables entre les résultats obtenus par la méthode $pPLV^{param}$ et $pPLV^{nonpar}$, moins de fausses activations sont observables pour la méthode paramétrique (c'est particulièrement visible entre les Figures 3.2(e) et (f)).

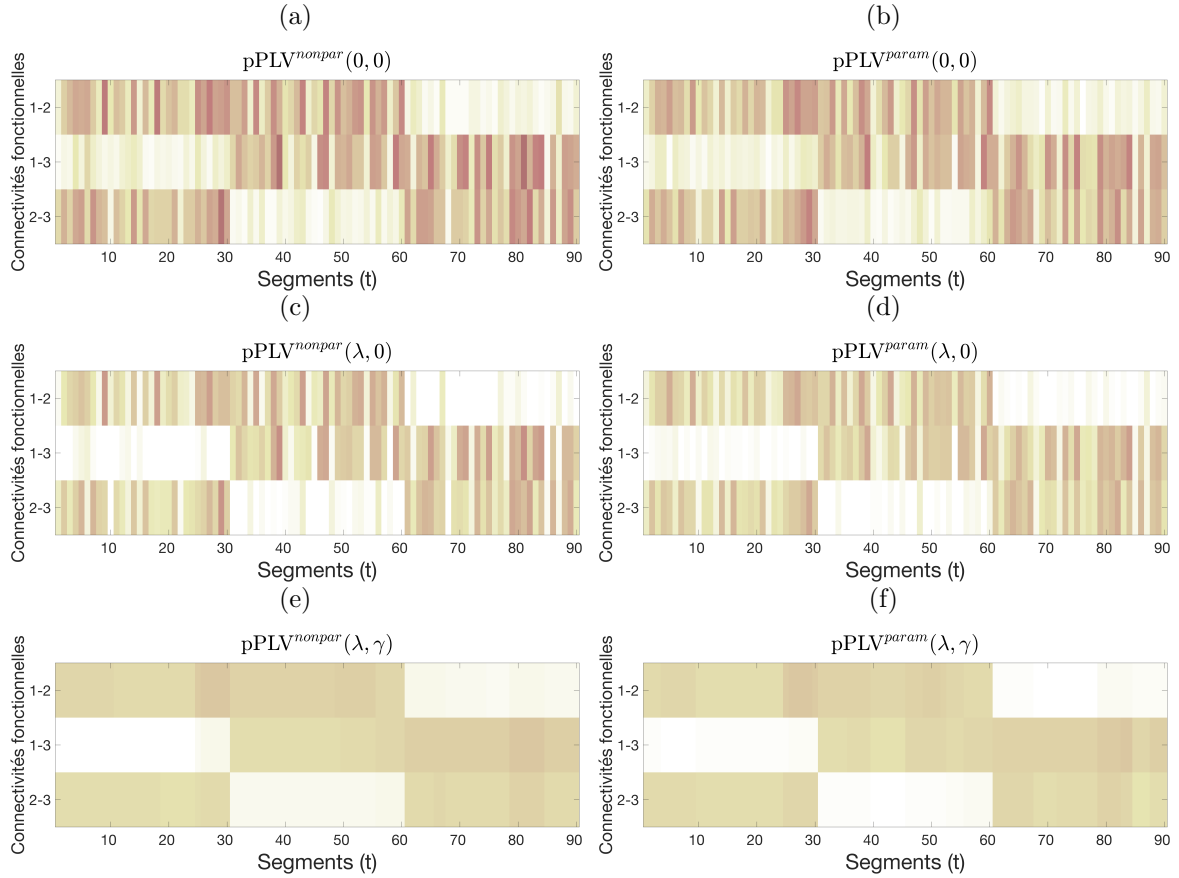


FIGURE 3.2 – Exemples de matrice de CFs obtenues par les méthodes (a) $pPLV^{nonpar}(0,0)$ - (b) $pPLV^{param}(0,0)$ - (c) $pPLV^{nonpar}(\lambda,0)$ - (d) $pPLV^{param}(\lambda,0)$ - (e) $pPLV^{nonpar}(\lambda,\gamma)$ - (f) $pPLV^{param}(\lambda,\gamma)$.

3.3.3 Application des méthodes sur le modèle

Modèle simple

On applique les différentes méthodes de calcul de la CF au modèle décrit dans la Section 3.3.1. On reste sur les mêmes conditions expérimentales que dans la Section 3.3.2. Des exemples de réalisation des signaux $s_i^{(t)}$ pour $t = 1$ et respectivement $i = 2$ et $i = 3$ sont fournis dans les Figures 3.3.(a) et 3.3.(b).

La Figure 3.3.(c) représentent un exemple de matrice de CFs pour le cas non régularisé $pPLV^{nonpar}(0,0)$. Il s'agit du seul estimateur utilisé pour montrer l'impact du manque de régularisation. On choisit la pPLV non paramétrique, car c'est elle qui dans le cas non régularisé donne les résultats les plus "consistants" pour l'ensemble des expériences de cette section. Ensuite les Figures 3.3.(d)-(h) montre les résultats obtenus pour chaque mesure de CFs du Tableau 3.1, soit les estimateurs $pPLV^{nonpar}(\lambda,\gamma)$, $pPLV^{param}(\lambda,\gamma)$, $\overline{pPLV^{param}}(\lambda,\gamma)$, $p\rho(\lambda,\gamma)$ et $\overline{p\rho}(\lambda,\gamma)$. Dans cette situation, la seule différence notable est dans l'ajout de la régularisation. Pour tous les estimateurs régularisés les résultats obtenus sont proches de la matrice de CFs idéale donnée dans la Figure 3.1(b).

Le Tableau 3.3 présente les deux scores (**SumNeg** et **Contrast**) évaluant la performance d'estimation de 6 mesures de pPLV sur 100 réalisations de signaux obtenus via le modèle de Roessler. Les observations de la Figure 3.3 sont confortées par les résultats du tableau, avec un fort **Contrast** et un faible **SumNeg** score pour chaque méthode régularisée. Dans ce cas, les méthodes avec corrélation partielle semblent plus adaptées pour détecter le couplage entre les signaux que les méthodes de pPLV.

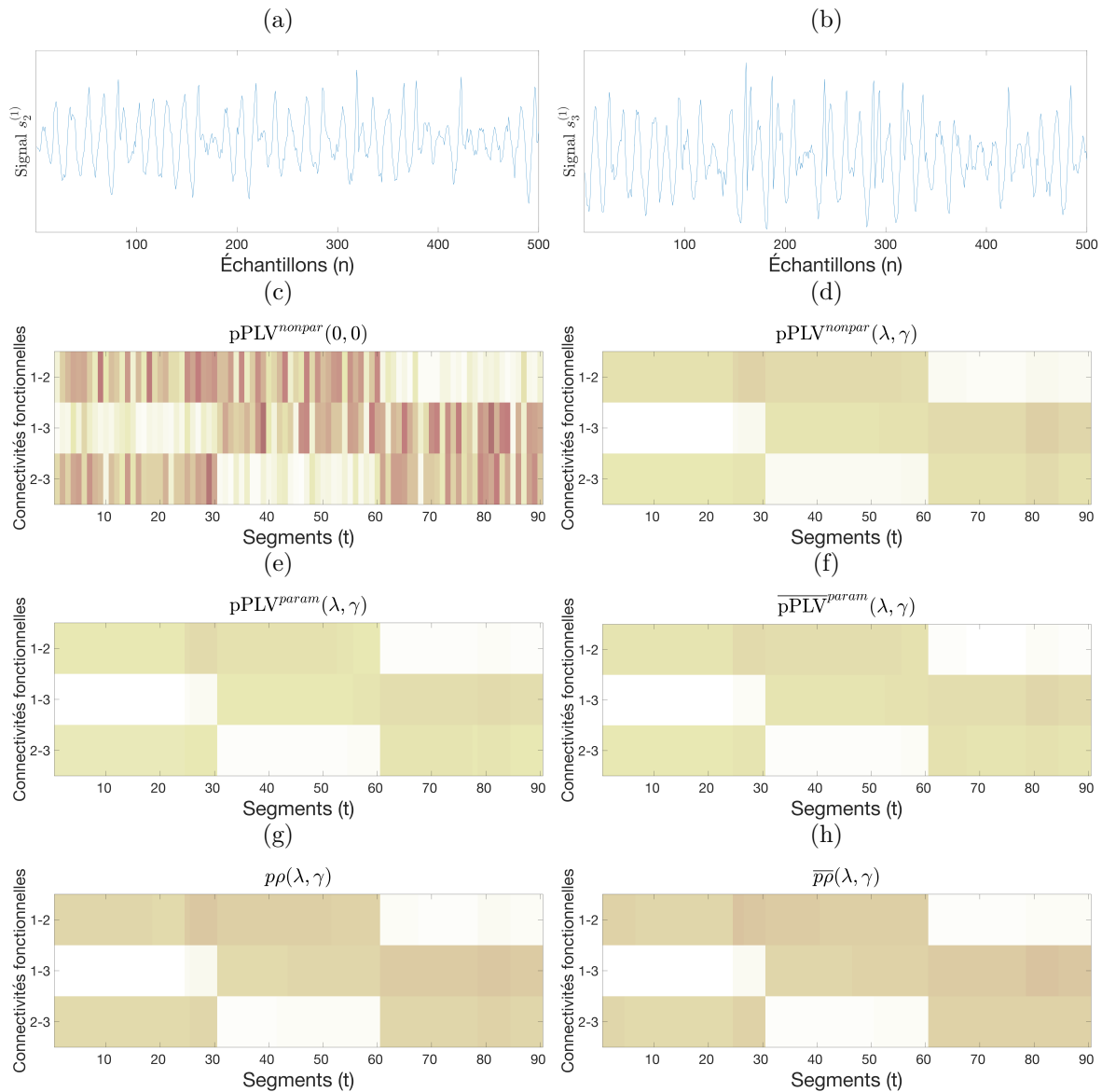


FIGURE 3.3 – (a), (b), Exemples de signaux provenant du modèle de Roessler pour $t = 1$ et $i = 2$ et 3. - Matrices de CFs obtenues respectivement à partir des méthodes (c) $pPLV^{nonpar}(0,0)$, (d) $pPLV^{nonpar}(\lambda, \gamma)$, (e) $pPLV^{param}(\lambda, \gamma)$, (f) $\overline{pPLV}^{param}(\lambda, \gamma)$, (g) $p\rho(\lambda, \gamma)$ et (h) $\overline{p\rho}(\lambda, \gamma)$.

	$pPLV^{nonpar}(0,0)$	$pPLV^{nonpar}$	$pPLV^{param}$	\overline{pPLV}^{param}	$p\rho$	$\overline{p\rho}$
SumNeg	14	1.9	1.2	1	1.3	1.2
Contrast	-0.18	0.16	0.17	0.18	0.22	0.24

Tableau 3.3 – Performances de la méthode $pPLV^{nonpar}(0,0)$ (cas non régularisé) et des 5 mesures de CFs étudiées avec régularisations (λ, γ) sur le modèle de Roessler sans transformation.

Impact du retard

À partir des signaux obtenus grâce au modèle décrit dans la Section 3.3.1, on applique la modification suivante :

$$\tilde{s}_i^{(t)}(n) = s_i^{(t)}(n - m_i), \quad (3.37)$$

Il s'agit d'un retard constant, on utilise $m_1 = 0$, $m_2 = 5$ et $m_3 = 10$. Ce type d'effet apparait dans les signaux psychophysiologiques. En effet, en fonction de la position des électrodes aux groupes de neurones sources, les signaux peuvent être reçus avec du retard. Des exemples de réalisation des signaux $s_i^{(t)}$ et de la modification apportée $\tilde{s}_i^{(t)}(n)$ pour $t = 1$ et respectivement $i = 2$ et $i = 3$ sont fournis dans les Figures 3.4.(a) et 3.4.(b).

Les Figures 3.4.(c)-(h) représentent un exemple de matrice de CFs pour le cas non régularisé $\text{pPLV}^{\text{nonpar}}(0, 0)$ et pour 5 mesures de CFs du Tableau 3.1. Dans cette situation, seules les mesures de pPLV permettent de détecter le couplage des signaux (c'est à dire les méthodes $\text{pPLV}^{\text{nonpar}}$, $\text{pPLV}^{\text{param}}$ et $\overline{\text{pPLV}}^{\text{param}}(\lambda, \gamma)$). On remarque que la corrélation partielle est fortement impactée par cette transformation, montrant une première différence de comportement entre les mesures de pPLV et de corrélation partielle. Cette différence est explicable car l'objectif d'une mesure comme la PLV ou la pPLV est de mesurer la stabilité de la différence de phase dans le temps. L'ajout d'un retard constant modifie principalement la valeur de la différence de phase, mais ce changement a un impact mineur sur la stabilité de cette différence dans le temps.

Le tableau 3.4 présente les deux scores (**SumNeg** et **Contrast**) évaluant la performance d'estimation de 6 mesures de pPLV sur 100 réalisations de signaux obtenus via le modèle de Roessler avec retard constant. Dans ce cas, les valeurs négatives de **Contrast** montrent bien que les méthodes utilisant la corrélation partielle ne sont plus adaptées.

Les observations des Figures 3.4 sont confortées par les résultats du tableau, avec un fort **Contrast** et un faible **SumNeg** score pour chaque méthode régularisée. Dans ce cas, les méthodes avec corrélation partielle ne semblent plus adaptées pour détecter le couplage entre les signaux que les méthodes de pPLV.

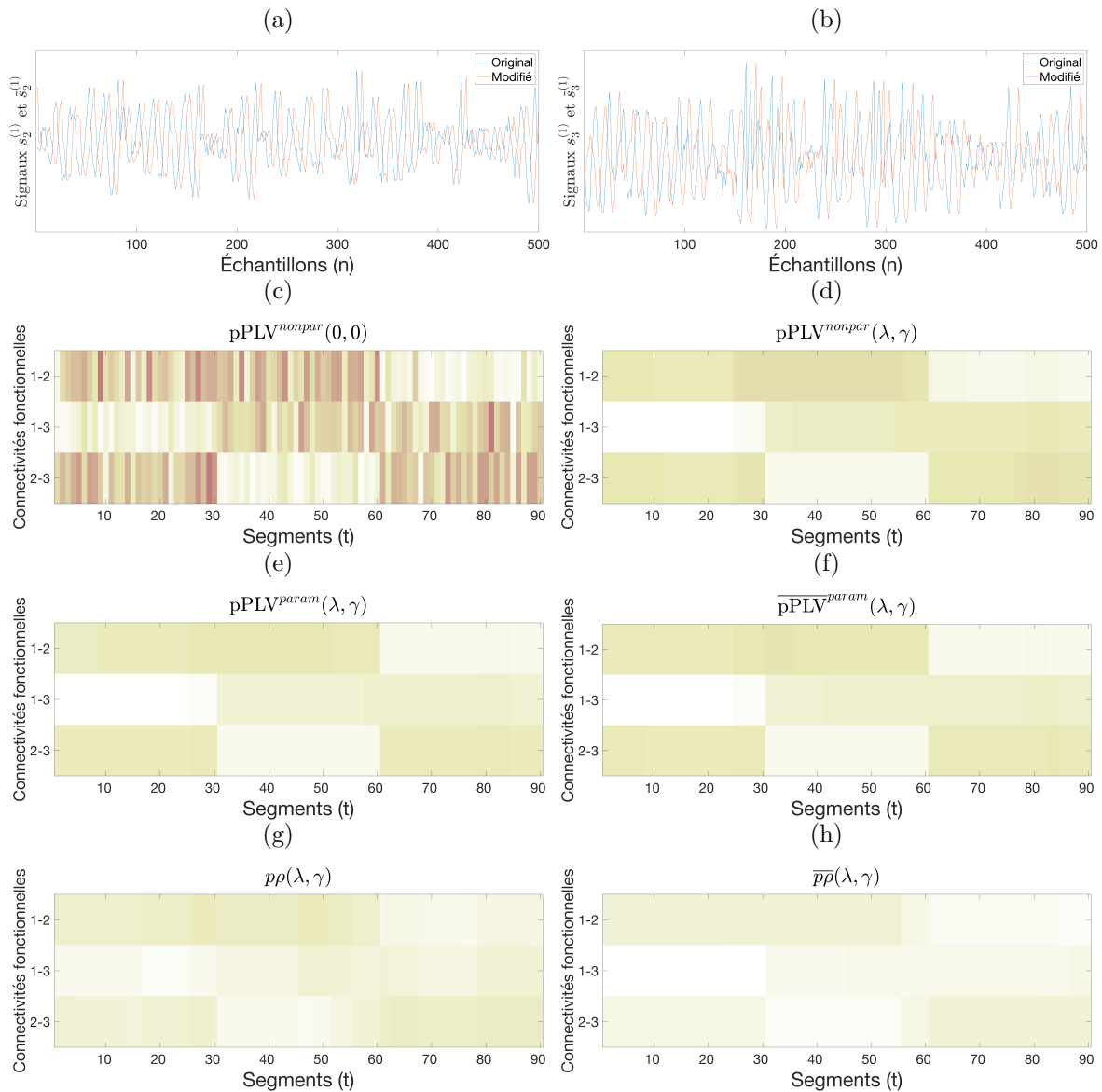


FIGURE 3.4 – (a), (b), Exemples de signaux provenant du modèle de Roessler et de sa transformation avec un retard constant pour $t = 1$ et $i = 2$ et 3 . - Matrices de CFs obtenues respectivement à partir des méthodes (c) $\text{pPLV}^{\text{nonpar}}(0, 0)$, (d) $\text{pPLV}^{\text{nonpar}}(\lambda, \gamma)$, (e) $\text{pPLV}^{\text{param}}(\lambda, \gamma)$, (f) $\overline{\text{pPLV}}^{\text{param}}(\lambda, \gamma)$, (g) $p\rho(\lambda, \gamma)$ et (h) $\overline{p\rho}(\lambda, \gamma)$.

	$\text{pPLV}^{\text{nonpar}}(0, 0)$	$\text{pPLV}^{\text{nonpar}}$	$\text{pPLV}^{\text{param}}$	$\overline{\text{pPLV}}^{\text{param}}$	$p\rho$	$\overline{p\rho}$
SumNeg	13	2.7	3	2.9	7.3	3.3
Contrast	-0.18	0.09	0.06	0.07	-0.07	-0.05

Tableau 3.4 – Performances de la méthode $\text{pPLV}^{\text{nonpar}}(0, 0)$ (cas non régularisé) et des 5 mesures de CFs étudiées avec régularisations (λ, γ) sur le modèle de Roessler avec retard.

Impact d'une différence d'amplitude constante

À partir des signaux obtenus grâce au modèle décrit dans la Section 3.3.1, on applique maintenant la transformation suivante :

$$\tilde{s}_i^{(t)}(n) = \alpha_i s_i^{(t)}(n), \quad (3.38)$$

les signaux sont ici multipliés par une amplitude constante. On a $\alpha_1 = 1$, $\alpha_2 = 2$ et $\alpha_3 = 3$. Encore une fois, ce facteur d'amplitude est choisi à partir de l'observation des signaux neurophysiologiques. En effet, selon la distance d'une électrode aux groupes de neurones sources, le signal enregistré peut être plus ou moins fort. Des exemples de réalisation des signaux $s_i^{(t)}$ et de la modification apportée $\tilde{s}_i^{(t)}(n)$ pour $t = 1$ et respectivement $i = 2$ et $i = 3$ sont fournis dans les Figures 3.5.(a) et 3.5.(b).

Les Figures 3.5.(c)-(h) représentent un exemple de matrice de CFs pour le cas non régularisé $\text{pPLV}^{\text{nonpar}}(0,0)$ et pour 5 mesures de CFs du Tableau 3.1. On remarque que la corrélation partielle $p\rho$ et la pPLV paramétrique $\text{pPLV}^{\text{param}}$ sont impactées par cette modification et semblent produire des résultats biaisés, notamment entre les segments 30 et 60 et les signaux 2 et 3. Cette erreur d'estimation, due à un problème d'échelle entre les signaux, est connue pour la corrélation partielle [105]. Cela explique que l'utilisation du lasso graphique à partir de la matrice de corrélation a été introduite. En effet, la méthode $\overline{p\rho}$ fournit une matrice de CF pertinente. Le fait que la méthode $\text{pPLV}^{\text{param}}$ ne marche pas très bien dans ce cas est dû à la remarque de la Section 3.1.4 car les signaux sont conditionnés par rapport aux autres signaux analytiques, et non par rapport à la phase de ces autres signaux. Cela implique que l'amplitude des autres signaux peut impacter le résultat de la pPLV paramétrique. L'ajout de la modification $\overline{\text{pPLV}}^{\text{param}}$ qui annule l'effet des amplitudes moyennes des signaux permet de nouveau de récupérer la bonne matrice de CFs.

Le tableau 3.5 présente les deux scores (**SumNeg** et **Contrast**) évaluant la performance d'estimation de 6 mesures de pPLV sur 100 réalisations de signaux obtenus via le modèle de Roessler avec amplitudes constantes. On remarque les performances dégradées des méthodes non modifiées (n'utilisant pas la matrice de corrélation) $\text{pPLV}^{\text{param}}$ et $p\rho$. En effet, elles sont associées à un score de **Contrast** faible et un score **SumNeg** élevé.

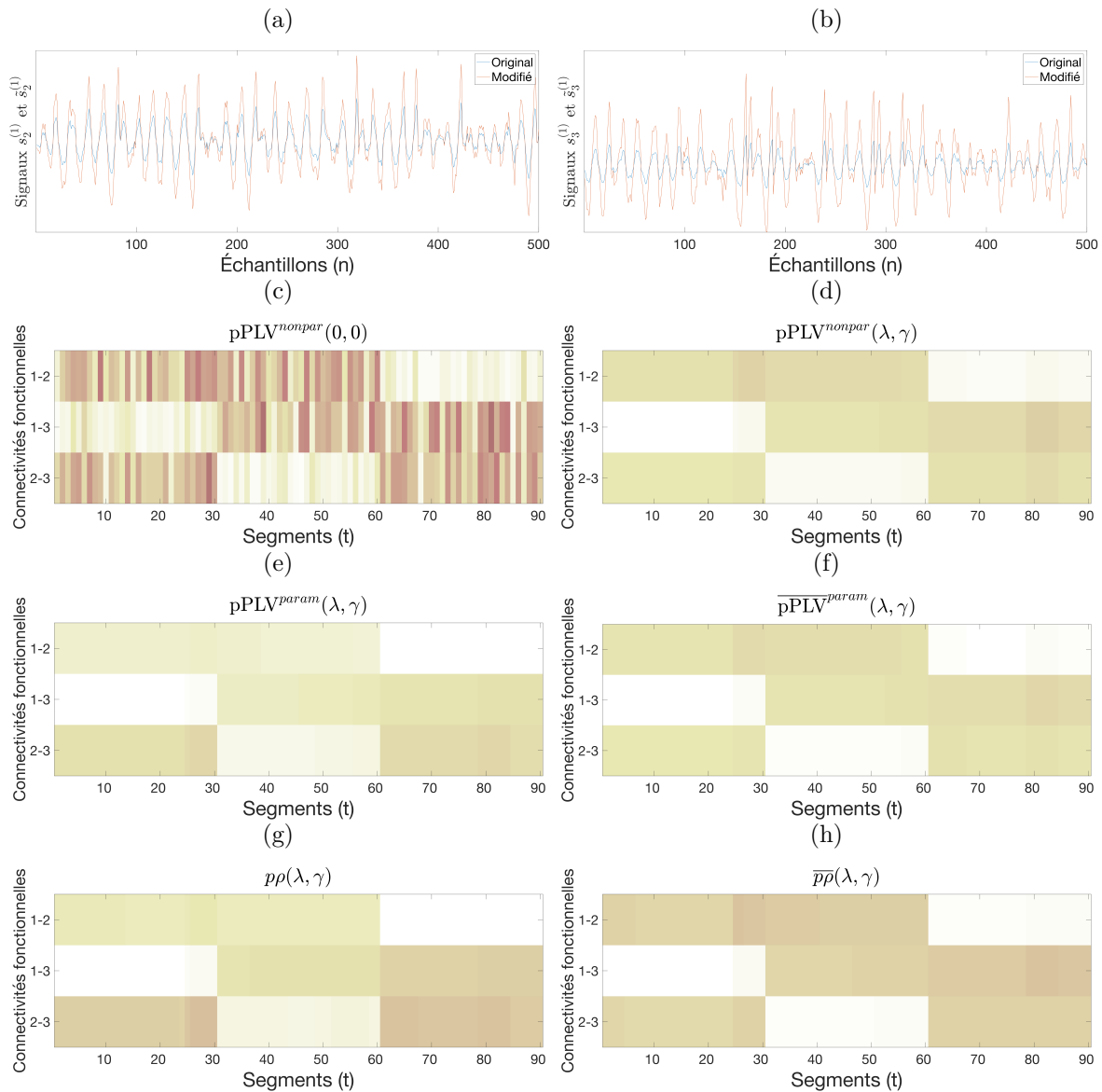


FIGURE 3.5 – (a), (b), Exemples de signaux provenant du modèle de Roessler et de sa transformation avec une amplitude constante pour $t = 1$ et $i = 2$ et 3 . - Matrices de CFs obtenues respectivement à partir des méthodes (c) $pPLV^{nonpar}(0,0)$, (d) $pPLV^{nonpar}(\lambda, \gamma)$, (e) $pPLV^{param}(\lambda, \gamma)$, (f) $\overline{pPLV}^{param}(\lambda, \gamma)$, (g) $p\rho(\lambda, \gamma)$ et (h) $\overline{p\rho}(\lambda, \gamma)$.

	$pPLV^{nonpar}(0,0)$	$pPLV^{nonpar}$	$pPLV^{param}$	\overline{pPLV}^{param}	$p\rho$	$\overline{p\rho}$
SumNeg	14	1.9	3	1	3.4	1.2
Contrast	-0.18	0.16	0	0.18	0.01	0.24

Tableau 3.5 – Performance de la méthode $pPLV^{nonpar}(0,0)$ (cas non régularisé) et des 5 mesures de CFs étudiées avec régularisations (λ, γ) sur le modèle de Roessler avec amplitudes constantes.

Impact de la modulation d'amplitude

À partir des signaux obtenus grâce au modèle décrit dans la Section 3.3.1, on applique maintenant la transformation suivante :

$$\tilde{s}_i^{(t)}(n) = \left[1 + \frac{h}{2}(\sin(2\pi f_i + \xi_i) - 1) \right] s_i^{(t)}(n), \quad (3.39)$$

On utilise le cas $h = 1$, avec $f_1 = f_2 = f_3 = 10^{-3}Hz$ afin de moduler avec un signal basse fréquence et $\xi_1 = 2$, $\xi_2 = 4$, $\xi_3 = 0$. Ce bruit modélise les variations d'amplitudes importantes qui peuvent apparaître au cours du temps sur les signaux sEEG réels. Des exemples de réalisation des signaux $s_i^{(t)}$ et de la modification apportée $\tilde{s}_i^{(t)}(n)$ pour $t = 1$ et respectivement $i = 2$ et $i = 3$ sont fournis dans les Figures 3.6.(a) et 3.6.(b).

Les Figures 3.6.(c)-(h) représentent un exemple de matrice de CFs pour le cas non régularisé $\text{pPLV}^{\text{nonpar}}(0, 0)$ et pour 5 mesures de CFs du Tableau 3.1. Dans cette situation, les résultats sont clairs, toutes les méthodes régularisées échouent à obtenir un profil de CFs satisfaisant, hormis pour la pPLV non paramétrique $\text{pPLV}^{\text{nonpar}}(\lambda, \gamma)$. En effet, comme montré dans la Section 3.1.2 c'est uniquement dans cette dernière situation que l'impact de l'amplitude des signaux est totalement éliminé.

Le Tableau 3.6 présente les deux scores (**SumNeg** et **Contrast**) évaluant la performance d'estimation de six mesures de pPLV sur 100 réalisations de signaux obtenus via le modèle de Roessler avec modulation d'amplitudes. On remarque que seul la pPLV non paramétrique $\text{pPLV}^{\text{nonpar}}$ à un **Contrast** positif, indiquant qu'il s'agit de la seule méthode permettant de détecter des couplages cohérents avec ce genre de transformation.

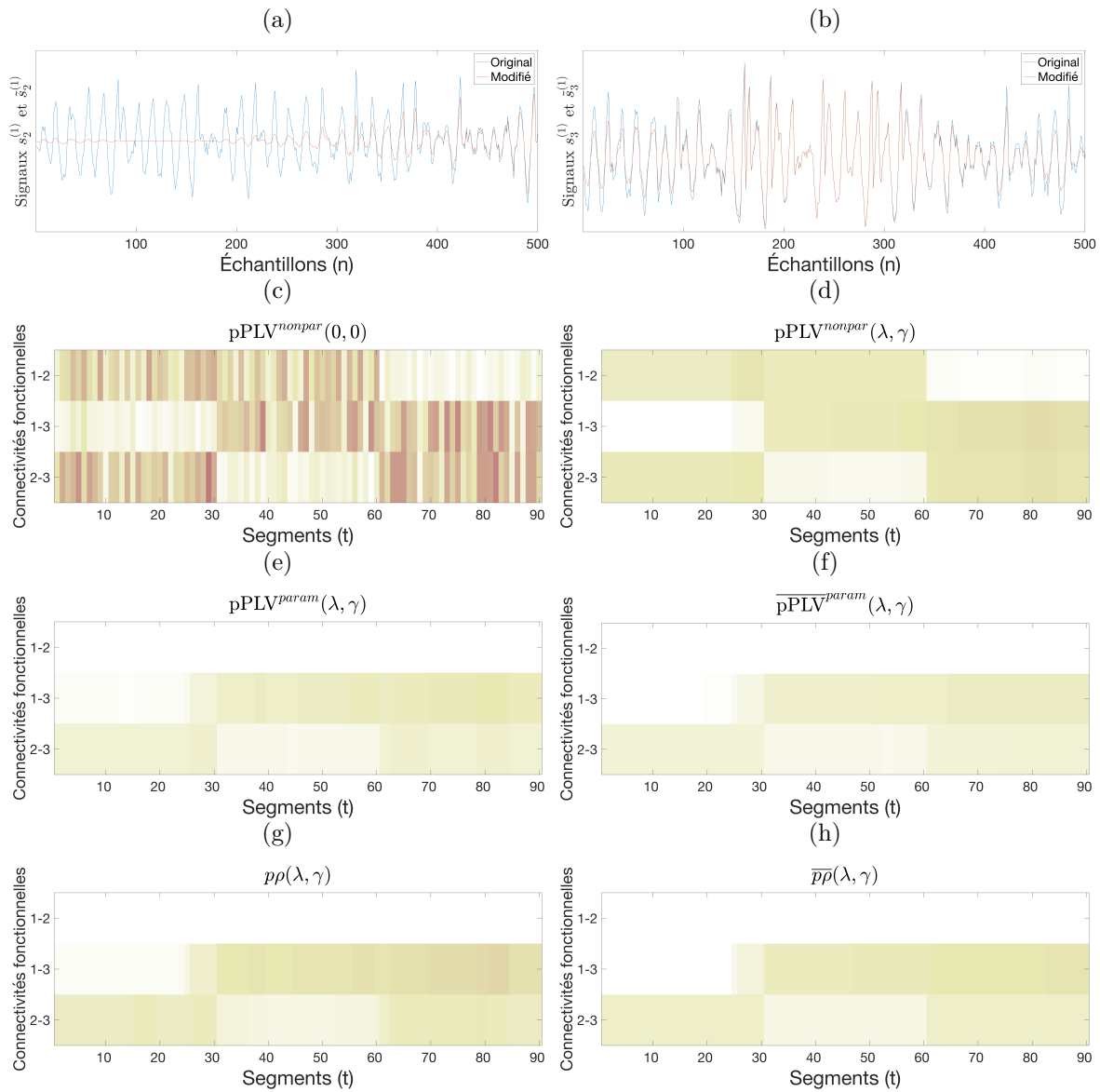


FIGURE 3.6 – (a), (b), Exemples de signaux provenant du modèle de Roessler et de sa transformation avec une modulation d’amplitude pour $t = 1$ et $i = 2$ et 3 . - Matrices de CFs obtenues respectivement à partir des méthodes (c) $pPLV^{nonpar}(0, 0)$, (d) $pPLV^{nonpar}(\lambda, \gamma)$, (e) $pPLV^{param}(\lambda, \gamma)$, (f) $\overline{pPLV}^{param}(\lambda, \gamma)$, (g) $p\rho(\lambda, \gamma)$ et (h) $\overline{p\rho}(\lambda, \gamma)$.

	$pPLV^{nonpar}(0, 0)$	$pPLV^{nonpar}$	$pPLV^{param}$	\overline{pPLV}^{param}	$p\rho$	$\overline{p\rho}$
SumNeg	13	2.2	4.6	3.8	5.8	4.8
Contrast	-0.18	0.1	-0.18	-0.15	-0.24	-0.2

Tableau 3.6 – Performance de la méthode $pPLV^{nonpar}(0, 0)$ (cas non régularisé) et des 5 mesures de CFs étudiées avec régularisations (λ, γ) sur le modèle de Roessler avec modulation d’amplitude.

Impact de la sur-modulation

Pour aller plus loin, on considère un objectif différent afin de détecter le couplage entre les signaux lorsqu'un phénomène de modulation avec changement de signe (ou sur-modulation) est appliqué. On utilise ici la transformation de l'Équation (3.39) dans le cas $h = 2$ et avec $f_1 = 0Hz$, $f_2 = 2.10^{-3}Hz$, $f_3 = 4.10^{-3}Hz$, $\xi_1 = \pi/2$, $\xi_2 = 0$ et $\xi_3 = 0$.

On peut justifier l'intérêt de ce bruit si l'on étudie des signaux large bande (comme dans le cas de l'épilepsie où les signaux sEEG peuvent être étudiés entre 20 et 100 Hz). Dans ce cas, les basses fréquences peuvent faire apparaître ce genre de phénomène. On note que dans cet exemple les mesures de pPLV fourniront un résultat nul. En effet, les changements de signes introduits produiront des signaux de phase quasi-orthogonaux. Dans cette situation, l'utilisation du second moment de la pPLV est préférable pour supprimer l'impact des changements de signe introduit par la sur-modulation. Le second moment de la pPLV est obtenu simplement, dans le cas non paramétrique, en modifiant le calcul de la matrice $\hat{\mathbf{K}}^{phase,(t)}$ de l'Équation (3.20) et notée $\hat{\mathbf{K}}^{phase2,(t)}$ tel que :

$$(\hat{\mathbf{K}}^{phase2,(t)})_{ij} = \frac{1}{N} \sum_{N=1}^N e^{i2(\phi_i^{(t)}(n) - \phi_j^{(t)}(n))}. \quad (3.40)$$

Pour montrer l'impact de ce changement, considérons deux signaux de N échantillons tels que leurs phases relatives soient stables $\phi_i(n) - \phi_j(n) = \psi \forall n$. Maintenant, on considère un changement de signe appliqué sur les $N/2$ derniers échantillons (on assume N paires), alors les calculs du premier et du second moment de la corrélation de phase donne :

$$\begin{aligned} (\hat{\mathbf{K}}^{phase,(t)})_{ij} &= \frac{1}{N} \sum_{n=1}^N e^{i(\phi_i^{(t)}(n) - \phi_j^{(t)}(n))}, & (\hat{\mathbf{K}}^{phase2,(t)})_{ij} &= \frac{1}{N} \sum_{n=1}^N e^{i2(\phi_i^{(t)}(n) - \phi_j^{(t)}(n))}, \\ &= \frac{1}{N} \sum_{n=1}^{N/2} e^{i\psi} + \frac{1}{N} \sum_{n=1}^{N/2} e^{i(\psi+\pi)}, & &= \frac{1}{N} \sum_{n=1}^{N/2} e^{i2\psi} + \frac{1}{N} \sum_{n=1}^{N/2} e^{i2(\psi+\pi)}, \\ &= 0. & (3.41) &= e^{i2\psi}. & (3.42) \end{aligned}$$

Dans cette situation, l'utilisation de l'ordre 1 indique que les signaux i et j ne sont pas couplés. À l'ordre 2, on obtient un résultat dont le module vaut 1, soit au contraire des signaux considérés comme complètement couplés. Dans le cas paramétrique, cela se résume à remplacer la fonction \mathcal{F} de l'Équation (3.12) par celle résultant de la formule des moments phase-amplitudes de l'Équation (2.86) dans le cas ($m = 0, p = 0, q = 2$), ce qui donne :

$$\mathcal{F}(x) = \frac{1}{x}((1-x)\log(1-x) + x). \quad (3.43)$$

Des exemples de réalisation des signaux $s_i^{(t)}$ et de la modification apportée $\tilde{s}_i^{(t)}(n)$ pour $t = 1$ et respectivement $i = 2$ et $i = 3$ sont fournis dans les Figures 3.7.(a) et 3.7.(b).

Les Figures 3.7.(c)-(h) représentent un exemple de matrice de CFs pour le cas non régularisé $pPLV^{nonpar}(0,0)$ et pour 5 mesures de CFs du Tableau 3.1. Finalement, dans cette situation, seule la pPLV non paramétrique $pPLV^{nonpar}(\lambda, \gamma)$ permet d'obtenir un profil d'activation de CFs satisfaisant. Cependant, on remarquera que les valeurs de CFs sont ici assez faibles comparées aux situations précédentes. Pour les autres méthodes, les valeurs obtenues sont nulles. En effet, on a ici des signaux quasi orthogonaux. Des mesures de covariance entre les signaux originaux ou analytiques ne permettent donc pas de révéler ce type de couplage. Cela met encore une fois en lumière les limites de l'utilisation des méthodes paramétriques pour ce genre de transformations complexes. Notons qu'ici la sur-modulation est très forte, cependant, même dans le cas de sur-modulation avec h plus faible (exemple $h = 1.3$), les résultats restent très similaires à ceux de la Figure 3.7.

Le tableau 3.7 présente les deux scores (**SumNeg** et **Contrast**) évaluant la performance d'estimation de 6 mesures de pPLV sur 100 réalisations de signaux obtenus via le modèle de Roessler avec sur-modulation. Les scores pour les méthodes paramétriques et utilisant la corrélation partielle ne

sont pas indiqués car les valeurs de CFs obtenues dans ce cas sont nulles. Seule la $pPLV^{param}$ à des performances satisfaisantes.

Conclusion : Pour chacune des situations explorées dans cette section, seule la pPLV non paramétrique régularisée $pPLV^{nonpar}(\lambda, \gamma)$ a permis dans tous les cas de mesurer de façon pertinente la synchronie des phases entre les signaux modélisés.

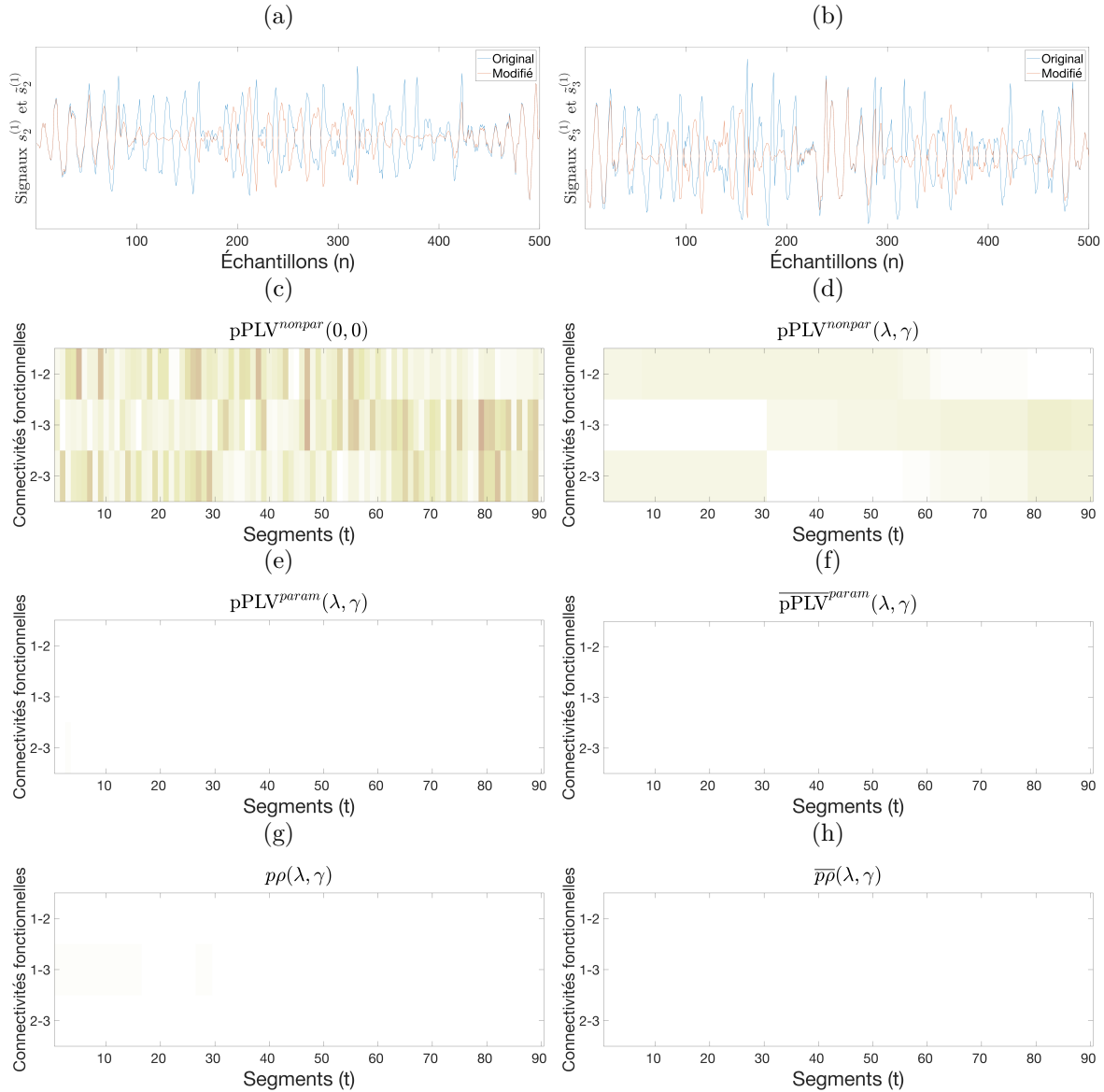


FIGURE 3.7 – (a), (b), Exemples de signaux provenant du modèle de Roessler et de sa transformation avec une sur-modulation pour $t = 1$ et $i = 2$ et 3 . - Matrices de CFs obtenues respectivement à partir des méthodes (c) $pPLV^{nonpar}(0,0)$, (d) $pPLV^{nonpar}(\lambda, \gamma)$, (e) $pPLV^{param}(\lambda, \gamma)$, (f) $\overline{pPLV}^{param}(\lambda, \gamma)$, (g) $p\rho(\lambda, \gamma)$ et (h) $\overline{p\rho}(\lambda, \gamma)$.

	$pPLV^{nonpar}(0,0)$	$pPLV^{nonpar}$	$pPLV^{param}$	\overline{pPLV}^{param}	$p\rho$	$\overline{p\rho}$
SumNeg	10	0.9	-	-	-	-
Contrast	-0.21	0.03	-	-	-	-

Tableau 3.7 – Performance de la méthode $pPLV^{nonpar}(0,0)$ (cas non régularisé) et des 5 mesures de CFs étudiées avec régularisations (λ, γ) sur le modèle de Roessler avec sur modulation.

3.4 Application sur des données sEEG

On considère des signaux sEEG réels provenant de 5 patients différents [89], [46]. Le nombre d'électrodes intracrâniennes varie de 102 à 138. Afin d'éviter une trop forte corrélation spatiale entre les signaux, un quart des électrodes est sélectionné (1 électrode sur 4 par tige stéréotaxique). Les 3 premiers patients sont associés à 4 crises, et 3 crises pour les deux derniers. Les crises ont des durées hétérogènes. Pour chacune d'elle, le calcul dynamique des CFs est réalisé via un fenêtrage rectangulaire de 2 secondes, et un pas d'une seconde, formant T segments.

À partir de ces données, on propose une étude numérique afin de comparer trois mesures de CF : la PLV empirique (Définition 9), la pPLV non paramétrique $\text{pPLV}^{\text{nonpar}}(0, 0)$ définie dans l'Équation 3.22 et enfin la pPLV non paramétrique régularisée $\text{pPLV}^{\text{nonpar}}(\lambda, \gamma)$ définie dans l'Équation 3.22. Nous ne prenons pas en compte la pPLV paramétrique puisque les limites de cette méthode qui ont été étudiées dans la section précédente sont trop importantes sur le modèle de Roessler. Le but de notre expérience est de montrer quelle mesure produit la plus forte proportion de CF cohérente sur l'ensemble des crises de chaque patient. En effet, on suppose que la dynamique des CFs pour chaque patient est comparable d'une crise à l'autre. Pour chaque crise d'un patient, on fait alors l'hypothèse que les mêmes CFs s'activent mais sur des périodes temporelles différentes. Pour chaque crise, les 30 CFs qui ont la plus forte activation sont sélectionnées. Cela correspond à trouver un seuil sur les valeurs de CF tel que seulement 30 CFs s'activent au moins une fois sur l'ensemble des segments. On cherche alors un ensemble \mathcal{S}_{tresh} et un seuil α_{tresh} tel que :

$$\mathcal{S}_{tresh} = \{i \mid \max(x_{i,:}) > \alpha_{tresh}\}, \quad \text{card}(\mathcal{S}_{tresh}) = 30. \quad (3.44)$$

Avec $\mathbf{X} \in \mathbb{R}^{L \times T}$ la matrice de FC d'une crise et $\text{card}(\bullet)$ la fonction qui compte le nombre d'éléments dans un ensemble. Dans le cas où 4 crises ont été enregistrées pour un patient, les CFs sont divisées en 4 catégories : les CFs communes à toutes les crises, celles qui apparaissent dans 3 crises, celles qui apparaissent dans 2 crises et celles qui sont spécifiques à une crise particulière. De la même manière, si 3 crises ont été enregistrées pour un patient, les CFs sont divisées en 3 catégories : les CFs communes à toutes les crises, celles qui apparaissent dans 2 crises et celles qui sont spécifiques à une crise particulière. Les proportions de CFs dans chaque catégorie pour les 3 mesures de CFs considérées sont présentées dans la Figure 3.8. Pour chaque patient, un barreau indique la proportion de CFs dans chaque catégorie et chaque méthode (à gauche). Ces valeurs sont indiquées dans un tableau (à droite).

Pour la pPLV non-paramétrique régularisée, une grille de paramètre (λ, γ) est utilisée avec $\lambda = [0.1, 0.2, 0.3, 0.4, 0.5]$ et $\gamma = [1, 2, 3, 5, 10]$. Pour chaque patient, seul le résultat de la combinaison (λ, γ) fournissant la plus faible proportion de CFs spécifiques à une crise est retenue.

Pour chaque patient, on observe que la pPLV non régularisée $\text{pPLV}^{\text{nonpar}}(0, 0)$ fournit en moyenne des résultats moins robustes sur l'ensemble des crises que la PLV empirique. En effet, le nombre de CFs spécifique augmente dans chaque cas. Malgré tout, la proportion de CFs commune à l'ensemble des crises n'est pas impactée ou augmente légèrement. Dans tous les cas, la pPLV régularisée $\text{pPLV}^{\text{nonpar}}(\lambda, \gamma)$ permet d'estimer des CFs plus robustes que les deux autres estimateurs. En effet, on observe que la proportion de CFs communes à toutes les crises augmente en utilisant la pPLV régularisée. Aussi, le nombre de CFs spécifique à chaque crise diminue grâce à cette méthode.

Afin d'illustrer les CFs obtenues à partir des trois mesures que sont la PLV, la pPLV sans régularisation $\text{pPLV}^{\text{nonpar}}(0, 0)$, la pPLV avec régularisation $\text{pPLV}^{\text{nonpar}}(\lambda, \gamma)$, on représente la matrice de CFs obtenue pour la première crise du patient 1 de la Figure 3.9. On remarque que moins de CFs sont exprimées en utilisant la pPLV que la PLV. On constate alors que la pPLV supprime de nombreuses connections grâce à la connaissance des autres signaux. Cependant, il est difficile à travers la matrice de pPLV d'observer les CFs les plus pertinentes. On remarque que la régularisation apporte une lecture claire des CFs importantes pour décrire la crise d'épilepsie.

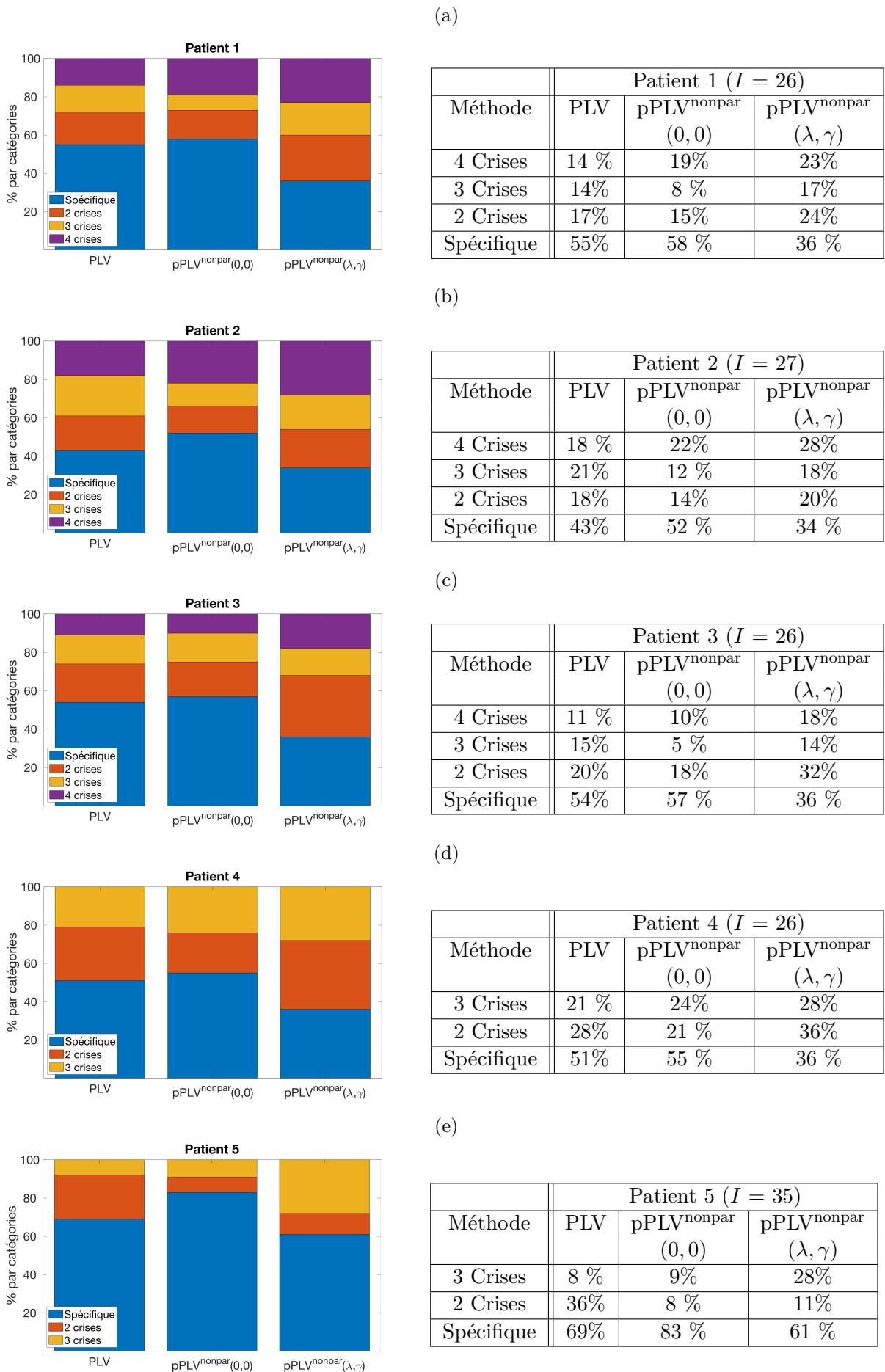


FIGURE 3.8 – Proportion des CFs les plus actives communes à 2 crises, 3 crises, (4 crises) ou spécifique à une crise pour les patients 1 à 5 respectivement figure (a)-(e).

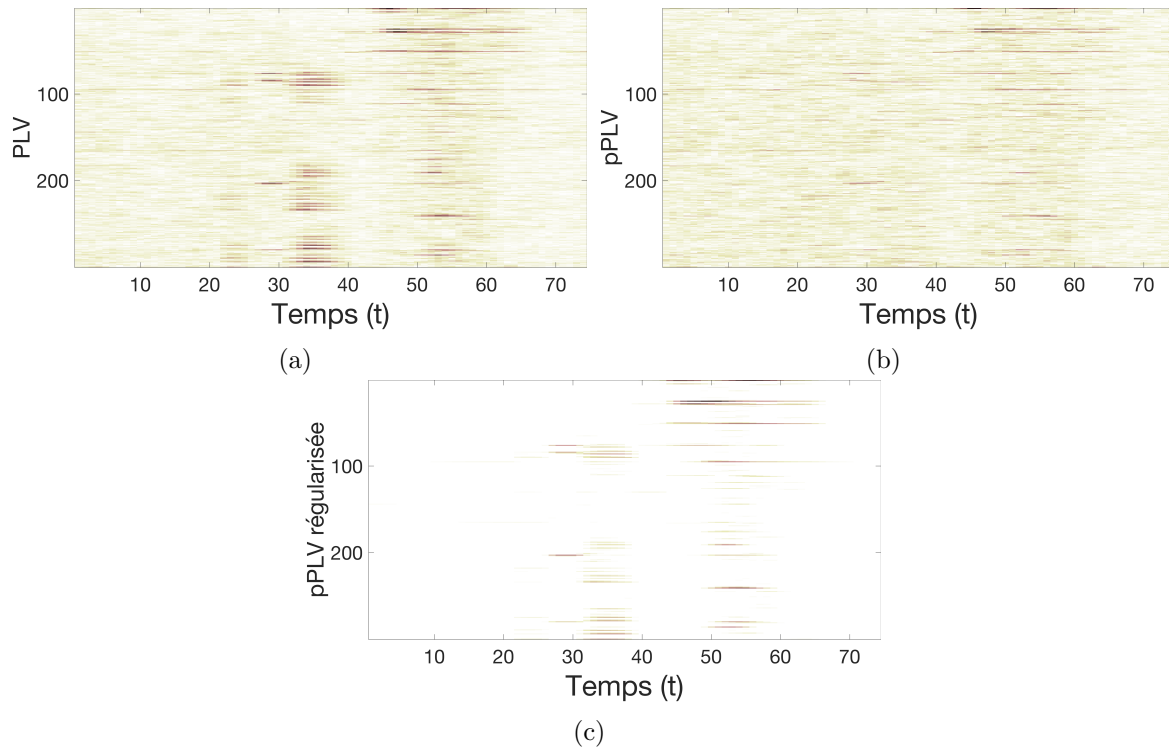


FIGURE 3.9 – Matrices de CFs d’un enregistrement iEEG pendant une crise, en utilisant : (a) la PLV, (b) la pPLV non paramétrique sans régularisation $pPLV(0,0)$, (c) la pPLV non paramétrique avec régularisation $pPLV(\lambda, \gamma)$.

3.5 Conclusion

Ce chapitre introduit l’utilisation de pPLV partielle et régularisée pour calculer les CFs de données en neurophysiologie. On propose deux méthodes pour calculer la pPLV. La première fait l’hypothèse que les signaux étudiés ont une distribution gaussienne multivariée. La seconde mesure de pPLV non paramétrique se base sur la matrice de synchronie paire à paire empirique. Les régularisations sont ajoutées à l’aide du lasso graphique qui fait aussi l’hypothèse que les signaux sont gaussiens. On impose ainsi aux mesures de pPLV d’être parcimonieuses et similaires d’un segment temporel à un autre, cela se traduit par un problème de lasso graphique variant dans le temps. La résolution de ce problème utilisant la méthode de direction alternée des multiplicateurs (ADMM) est détaillée dans ce chapitre. Finalement, nous comparons les différentes mesures de pPLV avec la corrélation classique sur un modèle de signaux synchrones réalistes, générés par un oscillateur de Roessler. Les résultats de cette étude montrent qu’une seule mesure de pPLV régularisée est satisfaisante, dans le sens où elle n’est jamais impactée par l’amplitude des signaux modélisés : il s’agit de la pPLV non paramétrique. Sur un jeu de données réelles, la pPLV régularisée, la pPLV non régularisée et la PLV classique sont comparées. On montre ainsi que la pPLV régularisée infère des CFs qui ont tendance à être plus cohérentes sur l’ensemble des crises d’un même patient.

Cependant, bien que ce chapitre montre la pertinence de la pPLV régularisée non paramétrique, il s’agit d’un travail exploratoire qui soulève de nombreuses questions, car pour le moment, cette mesure peut difficilement être utilisée sur des signaux réels pour une application clinique. Les problématiques de cette nouvelle mesure sont les suivantes :

- Les signaux neurophysiologiques, lorsqu’un grand nombre d’électrodes est utilisé, ont tendance à être spatialement corrélés. Ainsi, l’étude sur les signaux sEEG réels de ce chapitre a été réalisée sur un sous-ensemble d’électrodes, permettant d’obtenir des résultats intéressants. Au contraire, si on utilise toutes les électrodes, la majorité des similitudes obtenues révèlent uniquement des adjacences spatiales entre les électrodes alors que de nombreuses CFs qui semblent pertinentes pour caractériser la

crise d'épilepsie disparaissent. Devant ce constat deux approches sont envisageables. La première est une modification de la définition de pPLV, permettant de ne pas partialiser un signal avec les signaux des électrodes trop proches spatialement. La seconde approche est de calculer des mesures de PLV régularisées simples, sans partialisation préalable des signaux, il s'agit alors d'un problème beaucoup plus complexe à implémenter car non convexe.

- Comme le montre les expériences de la Section 3.3.3, l'hypothèse de signaux gaussiens semble trop réductrice notamment pour prendre en compte des effets de modulation. Tout cela questionne l'utilisation du lasso graphique, qui fait aussi l'hypothèse que les signaux sont gaussiens, afin de calculer la pPLV régularisée non paramétrique. Comme montré dans les travaux de [11], des alternatives plus robustes au lasso graphique existent. On pourrait alors remplacer le problème d'optimisation de l'Équation (3.31) par la critère CLIME [43] :

$$\begin{aligned} \operatorname{argmin}_{\Theta \in \mathbb{S}_{++}^p} \|\Theta\|_1 & \tag{3.45} \\ \text{t.q. } \|\hat{\mathbf{R}}\Theta - \mathbf{I}\|_\infty & \leq \lambda \end{aligned}$$

Afin de pouvoir appliquer ce problème à une mesure dynamique de CFs, il faudrait étendre le critère de l'Équation (3.45) aux mesures évoluant dans le temps pour appliquer une régularisation temporelle.

- Le problème du choix des paramètres de régularisation est inhérent à ce genre de méthode. On propose dans notre expérience une méthode, impliquant de calculer les pPLV régularisées selon une grille de paramètres (λ, γ) , et de choisir le couple de paramètres permettant d'obtenir la plus faible proportion de CFs spécifiques à chaque crise. Il s'agit d'une proposition intéressante, mais on pourrait aussi sélectionner le couple de paramètres permettant d'obtenir la plus forte proportion de CFs appartenant à l'ensemble des crises. Une comparaison de ces deux procédures pourrait être réalisée, ainsi qu'une implémentation efficace des mesures de pPLV le long de la grille de paramètres (λ, γ) .

- L'algorithme ADMM, bien que fonctionnel, n'est souvent pas le meilleur choix d'implémentation en terme de vitesse de convergence. Une comparaison des performances obtenues pour inférer des pPLV régularisées avec un algorithme de type Forward-Backward pourrait être envisagée [186].

- Finalement, nous avons montré que seule la pPLV non paramétrique permettait de mesurer réellement le couplage de phase pour chaque expérience réalisée. La mesure de pPLV paramétrique n'est pas partialisée uniquement selon la phase des autres signaux, mais intègre aussi leurs amplitudes. Malheureusement, une formule analytique de pPLV paramétrique, dans le cas gaussien, qui partialise uniquement selon la phase des autres signaux semble impossible [147].

Dans les chapitres suivants, la PLV simple sera utilisée au lieu de la pPLV régularisée. Ce choix est guidé par les considérations ci-dessus. De plus, la comparaison des méthodes de décomposition de graphes dynamiques qui seront présentées dans les sections suivantes, sera plus simple si les mesures de CFs utilisées ne produisent pas déjà des résultats parcimonieux et structurés dans le temps.

Annexe

3.A Détails sur l'implémentation de la version dynamique du lasso Graphique

3.A.1 ADMM

Une méthode numérique efficace afin de résoudre le problème d'optimisation de l'Équation (3.31) est la méthode de direction alternée des multiplicateurs (ADMM) [90]. En effet, le critère est trop complexe pour pouvoir directement appliquer un algorithme de descente de gradient ou de Newton. C'est pour cela qu'une méthode de division de la variable Θ comme l'ADMM est appliquée. Cette méthode consiste donc à découper le problème de l'Équation (3.31) en plusieurs sous problèmes plus simples à résoudre. On introduit les variables $\mathbf{B} = \{\mathbf{B}^{(1)}, \dots, \mathbf{B}^{(t)}\}$, $\mathbf{C} = \{\mathbf{C}^{(1)}, \dots, \mathbf{C}^{(t)}\}$ et $\mathbf{D} = \{\mathbf{D}^{(1)}, \dots, \mathbf{D}^{(t)}\}$ avec $\mathbf{B}^{(t)} \in \mathbb{R}^{I \times I}$, $\mathbf{C}^{(t)} \in \mathbb{R}^{I \times I}$ et $\mathbf{D}^{(t)} \in \mathbb{R}^{I \times I} \forall t \in \{1, \dots, T\}$. Ces variables peuvent être vues comme des copies des matrices $\Theta^{(t)}$, permettant de décomposer le problème de l'Équation (3.31), et produisant le critère de l'ADMM qui s'écrit :

$$\begin{aligned} \underset{\Theta \in \mathbb{S}_{++}^p, \mathbf{B}, \mathbf{C}, \mathbf{D}}{\operatorname{argmin}} \quad & \sum_{t=1}^T \left[-\ln(\det(\Theta^{(t)})) + \operatorname{Tr}(\Theta^{(t)} \hat{\mathbf{K}}^{(t)}) + \lambda \|\mathbf{B}^{(t)}\|_{od,1} \right] + \gamma \sum_{\tau=1}^{T-1} g(\mathbf{D}^{(\tau+1)} - \mathbf{C}^{(\tau)}), \\ \text{t.q.} \quad & \mathbf{B}^{(t)} = \Theta^{(t)}, \quad \mathbf{C}^{(t)} = \Theta^{(t)}, \quad \mathbf{D}^{(t)} = \Theta^{(t)}. \end{aligned} \quad (3.46)$$

Comme l'utilisation seule du lagrangien afin de résoudre ce problème d'optimisation sous contrainte est soumis à des points-selles [37], on introduit à la place le lagrangien augmenté. En introduisant le paramètre de pénalité de la méthode ADMM $\delta > 0$, et les variables duales normalisées $(\mathbf{U}_B^{(t)}, \mathbf{U}_C^{(t)}, \mathbf{U}_D^{(t)})$ pour tout $t \in \{1, \dots, T\}$, le lagrangien augmenté de l'Équation (3.46) s'écrit [90] :

$$\begin{aligned} \mathcal{L}_\delta(\Theta, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{U}_B, \mathbf{U}_C, \mathbf{U}_D) &= \sum_{t=1}^T l(\Theta^{(t)}) + \lambda \|\mathbf{B}^{(t)}\|_{od,1} + \gamma \sum_{\tau=1}^{T-1} g(\mathbf{D}^{(\tau+1)} - \mathbf{C}^{(\tau)}) \\ &+ \frac{\delta}{2} \left(\sum_{t=1}^T h(\Theta^{(t)}, \mathbf{B}^{(t)}, \mathbf{U}_B^{(t)}) + \sum_{\tau=1}^{T-1} h(\Theta^{(\tau)}, \mathbf{C}^{(\tau)}, \mathbf{U}_C^{(\tau)}) + h(\Theta^{(\tau+1)}, \mathbf{D}^{(\tau+1)}, \mathbf{U}_D^{(\tau+1)}) \right), \end{aligned} \quad (3.47)$$

avec :

$$l(\Theta^{(t)}) = -\ln(\det(\Theta^{(t)})) + \operatorname{Tr}(\Theta^{(t)} \hat{\mathbf{K}}^{(t)}), \quad (3.48)$$

$$h(\mathbf{X}, \mathbf{Y}, \mathbf{U}_Y) = \|\mathbf{X} - \mathbf{Y} + \mathbf{U}_Y\|_F^2 + \|\mathbf{U}_Y\|_F^2. \quad (3.49)$$

On notera que pour $t = 1$, $\mathbf{D}^{(1)}$ et $\mathbf{U}_D^{(1)}$ ne sont jamais impliquées dans l'implémentation et correspondent à des matrices nulles, de même $\mathbf{C}^{(T)}$ et $\mathbf{U}_C^{(T)}$ quand $t = T$. On remarquera aussi que si les variables duales normalisées sont des matrices nulles $\forall t \in \{1, \dots, t\}$, alors l'Équation (3.47) se réduit à un lagrangien classique. L'algorithme de l'ADMM consiste à alterner les itérations suivantes (indexées par

la variable (κ) :

$$(I). \quad \Theta_{(\kappa+1)}^{(t)} = \underset{\Theta^{(t)} \in \mathbb{S}_{++}^p}{\operatorname{argmin}} \mathcal{L}_\delta(\Theta^{(t)}, \mathbf{B}_{(\kappa)}^{(t)}, \mathbf{C}_{(\kappa)}^{(t)}, \mathbf{D}_{(\kappa)}^{(t)}, \mathbf{U}_{B,(\kappa)}^{(t)}, \mathbf{U}_{C,(\kappa)}^{(t)}, \mathbf{U}_{D,(\kappa)}^{(t)}), \quad (3.50)$$

$$(II). \quad \{\mathbf{B}_{(\kappa+1)}^{(t)}, \mathbf{C}_{(\kappa+1)}^{(t)}, \mathbf{D}_{(\kappa+1)}^{(t)}\} = \underset{\mathbf{B}^{(t)}, \mathbf{C}^{(t)}, \mathbf{D}^{(t)}}{\operatorname{argmin}} \mathcal{L}_\delta(\Theta_{(\kappa+1)}^{(t)}, \mathbf{B}^{(t)}, \mathbf{C}^{(t)}, \mathbf{D}^{(t)}, \mathbf{U}_{B,(\kappa)}^{(t)}, \mathbf{U}_{C,(\kappa)}^{(t)}, \mathbf{U}_{D,(\kappa)}^{(t)}), \quad (3.51)$$

$$(III). \quad \begin{cases} \mathbf{U}_{B,(\kappa+1)}^{(t)} = \mathbf{U}_{B,(\kappa)}^{(t)} + (\Theta_{(\kappa+1)}^{(t)} - \mathbf{B}_{(\kappa+1)}^{(t)}), \\ \mathbf{U}_{C,(\kappa+1)}^{(\tau)} = \mathbf{U}_{C,(\kappa)}^{(\tau)} + (\Theta_{(\kappa+1)}^{(\tau)} - \mathbf{C}_{(\kappa+1)}^{(\tau)}), \\ \mathbf{U}_{D,(\kappa+1)}^{(\tau+1)} = \mathbf{U}_{D,(\kappa)}^{(\tau+1)} + (\Theta_{(\kappa+1)}^{(\tau+1)} - \mathbf{D}_{(\kappa+1)}^{(\tau+1)}), \end{cases} \quad \forall t \in \{1, \dots, T\}, \quad \forall \tau \in \{1, \dots, T-1\}. \quad (3.52)$$

Les étapes (I) et (II) constituent des problèmes de minimisation classique. L'étape (III) est propre aux méthodes primales-duales comme l'ADMM ou l'algorithme de Douglas-Rachford [37], [83]. Afin de préciser les formulations des étapes (I) et (II), on introduit l'opérateur proximal [161], [25], qui se note $\operatorname{prox}_{f(\bullet)}(y)$ pour toutes fonctions convexes $f(\bullet)$, et s'écrit :

$$\operatorname{prox}_{f(\bullet)}(y) = \underset{x}{\operatorname{argmin}} f(x) + \frac{1}{2} \|x - y\|_F^2. \quad (3.53)$$

L'opérateur proximal intervient dans de nombreux algorithmes d'optimisation. Une liste de solutions analytiques pour différentes fonctions $f(\bullet)$ est fournie sur le site [1]. Alors, on peut remarquer que $\forall t \in \{1, \dots, T\}$ et $\forall \tau \in \{1, \dots, T-1\}$:

$$(I). \quad \Theta_{(\kappa+1)}^{(t)} = \operatorname{prox}_{\frac{\varpi_N}{3\delta}(-\ln(\det(\bullet)) + \operatorname{Tr}(\hat{\mathbf{K}}^{(t)}\bullet))} \left(\frac{\mathbf{F}_{(\kappa)}^{(t)} + \mathbf{F}_{(\kappa)}^{(t)*}}{2} \right),$$

$$(II) \quad \begin{cases} (a). \quad \mathbf{B}_{(\kappa+1)}^{(t)} = \operatorname{prox}_{\frac{\lambda}{\delta}\|\bullet\|_{\text{od},1}}(\Theta_{(\kappa+1)}^{(t)} + \mathbf{U}_{B,(\kappa)}^{(t)}), \\ (b). \quad [\mathbf{D}_{(\kappa+1)}^{(\tau+1)}, \mathbf{C}_{(\kappa+1)}^{(\tau)}]^t = \operatorname{prox}_{\frac{\gamma}{\delta}g([\mathbf{I}, -\mathbf{I}]\bullet)}([\Theta_{(\kappa+1)}^{(\tau+1)} + \mathbf{U}_{D,(\kappa)}^{(\tau+1)}, \Theta_{(\kappa+1)}^{(\tau)} + \mathbf{U}_{C,(\kappa)}^{(\tau)}]^t), \end{cases} \quad (3.54)$$

Avec $\mathbf{I} \in \mathbb{R}^{I \times I}$ la matrice identité. Pour le problème (I), on précise que ϖ_t correspond au nombre de variables duales impliquées (on a $\varpi_t = 2$ pour $t = \{1, T\}$ et sinon $\varpi_t = 3$). On a aussi :

$$\mathbf{F}_{(\kappa)}^{(t)} = \mathbf{B}_{(\kappa)}^{(t)} + \mathbf{C}_{(\kappa)}^{(t)} + \mathbf{D}_{(\kappa)}^{(t)} - \mathbf{U}_{B,(\kappa)}^{(t)} - \mathbf{U}_{C,(\kappa)}^{(t)} - \mathbf{U}_{D,(\kappa)}^{(t)}. \quad (3.55)$$

Pour le problème (II).(b), on précise que la matrice notée $[\mathbf{X}, \mathbf{Y}]$ correspond à la concaténation horizontale des matrices \mathbf{X} et \mathbf{Y} . Notons que la fonction $g(\bullet)$ dans l'Équation (3.54) peut aussi être écrite à partir d'une seule matrice concaténée. En effet on a :

$$g(\mathbf{X} - \mathbf{Y}) = g([\mathbf{I}, -\mathbf{I}][\mathbf{X}, \mathbf{Y}]^t). \quad (3.56)$$

Ainsi la résolution du problème de l'Équation (3.31) implique de trouver le résultat des trois opérateurs proximaux de l'Équation (3.54)

3.A.2 Etape (I), Mise à jour de la matrice Θ

En notant $\mathbf{Y} = \frac{\mathbf{F}_{(\kappa)}^{(t)} + \mathbf{F}_{(\kappa)}^{(t)*}}{2}$, on a :

$$\Theta_{(\kappa+1)}^{(t)} = \operatorname{prox}_{\frac{\varpi_N}{3\delta}(-\ln(\det(\bullet)) + \operatorname{Tr}(\hat{\mathbf{K}}^{(t)}\bullet))}(\mathbf{Y}), \quad (3.57)$$

$$= \underset{\mathbf{X}}{\operatorname{argmin}} -\ln(\det(\mathbf{X})) + \operatorname{Tr}(\hat{\mathbf{K}}^{(t)}\mathbf{X}) + \frac{3\delta}{2\varpi_N} \|\mathbf{X} - \mathbf{Y}\|_F^2. \quad (3.58)$$

Une solution analytique de ce problème existe : trouver quand s'annule le gradient du critère de l'Équation (3.58) revient à résoudre l'équation suivante :

$$\frac{3\delta}{\varpi_t} \mathbf{X} - \mathbf{X}^{-1} = \frac{3\delta}{\varpi_t} \mathbf{Y} - \hat{\mathbf{K}}^{(t)}. \quad (3.59)$$

En posant $\mathbf{V}\mathbf{\Lambda}\mathbf{V}^t$ et $\mathbf{W}\mathbf{\Sigma}\mathbf{W}^t$ les décompositions en valeurs propres respectives de \mathbf{X} et $\frac{3\delta}{\varpi_t} \mathbf{Y} - \hat{\mathbf{K}}^{(t)}$, on obtient :

$$\mathbf{V} \left(\frac{3\delta}{\varpi_t} \mathbf{\Lambda} - \mathbf{\Lambda}^{-1} \right) \mathbf{V}^t = \mathbf{W} \mathbf{\Sigma} \mathbf{W}^t, \quad (3.60)$$

Une solution consiste donc à fixer $\mathbf{V} = \mathbf{W}$ et à trouver les zéros positifs des I équations du second ordre suivantes :

$$\begin{aligned} \frac{3\delta}{\varpi_t} \mathbf{\Lambda} - \mathbf{\Lambda}^{-1} &= \mathbf{\Sigma}, \\ \frac{3\delta}{\varpi_t} \mathbf{\Lambda}^2 - \mathbf{\Sigma} \mathbf{\Lambda} - \mathbf{I} &= 0. \end{aligned} \quad (3.61)$$

Finalement on obtient la forme analytique, solution de l'Équation (3.58) :

$$\Theta_{(\kappa+1)}^{(t)} = \frac{\varpi_N}{6\delta} \mathbf{W} \left(\mathbf{\Sigma} + \sqrt{\mathbf{\Sigma}^2 + \frac{12\delta}{\varpi_N} \mathbf{I}} \right) \mathbf{W}^t. \quad (3.62)$$

On peut observer comment la fonction de l'entropie relative de l'Équation (3.27) permet de s'assurer que la matrice $\Theta_{(\kappa+1)}^{(t)}$ appartient au cône des matrices semie-définies positives. En effet, les matrices \mathbf{Y} et $\hat{\mathbf{K}}^{(t)}$ étant symétriques, alors selon le théorème spectral, les valeurs propres contenues dans la matrice diagonale $\mathbf{\Sigma}$ sont supérieures ou égales à 0. Les valeurs propres de la matrice $\Theta_{(\kappa+1)}^{(t)}$ seront alors nécessairement supérieures ou égales à 0 selon l'Équation (3.62).

3.A.3 Etape (II).(a), Mise à jour de la matrice \mathbf{B}

En notant maintenant $\mathbf{Y} = \Theta_{(\kappa+1)}^{(t)} + \mathbf{U}_{B,(\kappa)}^{(t)}$, on a :

$$\mathbf{B}_{(\kappa+1)}^{(t)} = \text{prox}_{\frac{\lambda}{\delta} \|\bullet\|_{od,1}}(\mathbf{Y}). \quad (3.63)$$

Comme la fonction $\|\bullet\|_{od,1}$ correspond à la somme du module des éléments non diagonaux, la règle de séparabilité par bloc de l'opérateur proximal s'applique [161], on a $\forall i \neq j$:

$$\dagger \left(\mathbf{B}_{(\kappa+1)}^{(t)} \right)_{ij} = \text{prox}_{\frac{\lambda}{\delta} \|\bullet\|_{od,1}}(y_{ij}), \quad (3.64)$$

$$= \underset{x_{ij}}{\text{argmin}} \quad |x_{ij}| + \frac{\delta}{2\lambda} \|x_{ij} - y_{ij}\|_F^2. \quad (3.65)$$

La solution de ce problème est bien connue quand $x_{ij} \in \mathbb{R}$ [161]. Dans ce cas la fonction $f(x) = |x|$ n'est pas différentiable pour $x = 0$. A la place on considère les sous-différentiels de la fonction f .

Definition 11 : Sous-différentiel

Le sous-différentiel d'une fonction convexe f en $\mathbf{x} \in \mathbb{R}^N$, noté $\partial f(\mathbf{x})$ [161] est l'ensemble des vecteurs \mathbf{p} tels que :

$$\partial f(\mathbf{x}) = \{ \mathbf{p} \mid \mathbf{p}^t (\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x}), \quad \forall \mathbf{y} \in \mathbb{R}^N \}. \quad (3.66)$$

Notons que si la fonction f est dérivable en \mathbf{x} alors $\partial f(\mathbf{x}) = \nabla f(\mathbf{x})$.

Alors pour la fonction $f(x) = |x|$, $\partial f(0)$ correspond à l'ensemble des pentes p , tel que la droite $d = py$ $\forall y \in \mathbb{R}$ soit inférieure ou égale en tous points à la courbe $|x|$. Ainsi :

$$\partial |x_{ij}| \in \begin{cases} 1, & \text{si } x > 0, \\ -1, & \text{si } x < 0, \\ [-1, 1], & \text{si } x = 0. \end{cases} \quad (3.67)$$

On illustre des exemples de sous-gradient de $\partial f(0)$ dans la Figure 3.A.1.

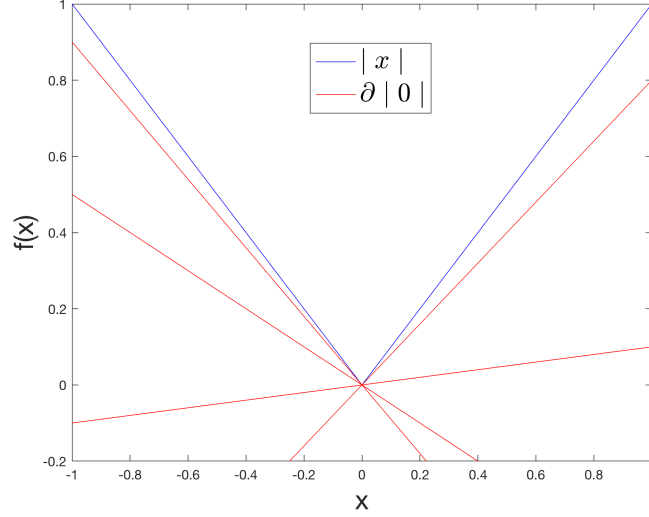


FIGURE 3.A.1 – Exemple de sous-gradient (en rouge) représenté par des droites de pente p pour la fonction $f(x) = |x|$ en $x = 0$.

Un minimum est atteint pour l'Équation (3.65) si 0 appartient au sous gradient du critère à minimiser, soit :

$$0 \in \partial |x_{ij}| + \frac{\delta}{\lambda}(x_{ij} - y_{ij}), \quad (3.68)$$

$$y_{ij} \in x_{ij} + \frac{\lambda}{\delta} \partial |x_{ij}|. \quad (3.69)$$

Si $y_{ij} \geq \frac{\lambda}{\delta}$ la solution est $x_{ij} = y_{ij} - \frac{\lambda}{\delta}$. De même, la solution est simple à trouver si $y_{ij} < -\frac{\lambda}{\delta}$. Enfin, si $-\frac{\lambda}{\delta} < y_{ij} < \frac{\lambda}{\delta}$ alors naturellement $y_{ij} \in \frac{\lambda}{\delta}[-1, 1]$ ce qui implique que $y_{ij} \in 0 + \frac{\lambda}{\delta} \partial |0|$, on a donc $x_{ij} = 0$ comme solution. Pour résumer, dans le cas réel, la solution de l'Équation (3.65) correspond à la fonction seuillage doux tel que :

$$\text{SD}_{\frac{\lambda}{\delta}}(y_{ij}) = \begin{cases} y_{ij} - \lambda, & \text{si } y_{ij} > \frac{\lambda}{\delta}, \\ y_{ij} + \lambda, & \text{si } y_{ij} < -\frac{\lambda}{\delta}, \\ 0, & \text{sinon.} \end{cases} \quad (3.70)$$

Cette fonction peut s'écrire sous la forme simplifiée :

$$\text{SD}_{\frac{\lambda}{\delta}}(y_{ij}) = \max \left(|y_{ij}| - \frac{\lambda}{\delta}, 0 \right) \text{sign}(y_{ij}), \quad (3.71)$$

avec $\max(a, b)$, la fonction qui vaut a si $a \geq b$, sinon b . $\text{sign}(a) = 1$ si a positif, $\text{sign}(a) = -1$ si a est négatif et $\text{sign}(a) = 0$ si $a = 0$.

Cependant, pour les calculs de pPLV, on travaille avec des matrices complexes. On précise alors la solution de l'Équation (3.65) quand $x_{ij} \in \mathbb{C}$. Soit $x \in \mathbb{C}$ avec $x = x_R + \text{i}x_I$ les parties réelles et imaginaires de la variable x . La fonction $f(x) = |x|$ ne respectant pas la condition de Cauchy-Riemann, cette fonction n'est pas holomorphe et n'est pas dérivable selon le sens usuel de la dérivée complexe

[123]. Le calcul de Wirtinger [166] (ou $\mathbb{C}\mathbb{R}$ -calcul) à été introduit afin de pouvoir généraliser la notion de dérivée à des fonctions non holomorphes à support dans \mathbb{C} . L'idée est d'étudier à la place de la fonction $f(x)$, la même fonction réécrite selon deux variables $g(x_R, x_I)$ ou $h(x, x^*)$ avec :

$$g(x_R, x_I) = \sqrt{x_R^2 + x_I^2}, \quad h(x, x^*) = \sqrt{xx^*}. \quad (3.72)$$

La fonction $g(x_R, x_I)$ est particulièrement intéressante dans notre cas car elle permet de traiter l'ensemble \mathbb{C} comme un isomorphisme de \mathbb{R}^2 . Plus précisément, la fonction $g(x_R, x_I)$ correspond à la norme du vecteur $\mathbf{x} = [x_R, x_I]$. La sous-différentielle de la norme est connue et correspond à :

$$\partial g(x_R, x_I) = \partial \|\mathbf{x}\|_F = \begin{cases} \frac{\mathbf{x}}{\|\mathbf{x}\|_F}, & \text{si } \|\mathbf{x}\|_F > 0, \\ \{\mathbf{p} \mid \|\mathbf{p}\|_F \leq 1\}, & \text{si } \|\mathbf{x}\|_F = 0. \end{cases} \quad (3.73)$$

On peut remarquer qu'il s'agit d'une simple généralisation du sous gradient de la fonction valeur absolue de l'Équation (3.67) en 2 dimensions. Alors, on peut directement traduire ce résultat en fonction du nombre complexe $x = x_R + \mathrm{i}x_I$ (ce dernier écrit en coordonnées polaires, correspond à $x = ae^{\mathrm{i}\phi}$) pour calculer le sous-gradient de $|x|$ [101] :

$$\partial |x| = \begin{cases} e^{\mathrm{i}\phi}, & \text{si } |x| > 0, \\ \{p \mid |p| \leq 1\}, & \text{si } |x| = 0. \end{cases} \quad (3.74)$$

Minimiser le problème de l'Équation (3.65) dans \mathbb{C} , revient alors de nouveau à trouver si 0 appartient au sous gradient du critère à minimiser. Cela correspond à résoudre l'Équation (3.68) sauf que désormais les sous-gradients de $|x|$ sont définis par l'Équation (3.74). Une solution à ce problème est connue [155], [187] et consiste à utiliser la fonction seuillage doux complexe, notée $\mathrm{ST}_\lambda(x)$ et qui s'écrit, pour tout nombre imaginaire $x = ae^{\mathrm{i}\phi}$:

$$\mathrm{ST}_\lambda(x) = \max(a - \lambda, 0)e^{\mathrm{i}\phi}. \quad (3.75)$$

Alors le résultat de l'Équation (3.63) est :

$$\begin{cases} \left(\mathbf{B}_{(\kappa+1)}^{(t)}\right)_{ij} = \mathrm{ST}_{\frac{\lambda}{\delta}}(\mathbf{y}_{ij}), & \text{si } \forall i \neq j, \\ \left(\mathbf{B}_{(\kappa+1)}^{(t)}\right)_{ii} = \mathbf{y}_{ii}, & \forall i. \end{cases} \quad (3.76)$$

3.A.4 Etape (II).(b), Mise à jour des matrices C et D

En notant maintenant $\mathbf{Y}_1 = \Theta_{(\kappa+1)}^{(\tau+1)} + \mathbf{U}_{D,(\kappa)}^{(\tau+1)}$ et $\mathbf{Y}_2 = \Theta_{(\kappa+1)}^{(\tau)} + \mathbf{U}_{C,(\kappa)}^{(\tau)}$ on a :

$$[\mathbf{D}_{(\kappa+1)}^{(\tau+1)}, \mathbf{C}_{(\kappa+1)}^{(\tau)}]^t = \mathrm{prox}_{\frac{\gamma}{\delta}g([\mathbf{I}, -\mathbf{I}]\bullet)}([\mathbf{Y}_1, \mathbf{Y}_2]^t). \quad (3.77)$$

On note que l'opérateur proximal de la fonction $g(\bullet)$, précisé par l'Équation (3.32) est connu. Cependant, ici la composition avec la fonction $[\mathbf{I}, -\mathbf{I}]\bullet$ empêche d'utiliser directement ce résultat. Il existe une propriété sur les opérateurs proximaux [161], [1], permettant de simplifier le problème de l'Équation (3.77). En considérant toutes fonctions $f(\bullet) = g(\mathbf{V}\bullet + \mathbf{W})$, avec \mathbf{V} et \mathbf{W} matrices quelconques sous condition que $\mathbf{V}\mathbf{V}^t = \alpha\mathbf{I}$ ($\alpha \in \mathbb{R}$), alors :

$$\mathrm{prox}_{f(\bullet)}(\mathbf{Y}) = \mathbf{Y} + \frac{1}{\alpha}\mathbf{V}^t(\mathrm{prox}_{\alpha g(\bullet)}(\mathbf{V}\mathbf{Y} + \mathbf{W}) - \mathbf{V}\mathbf{Y} - \mathbf{W}). \quad (3.78)$$

À partir de cette propriété et en notant $\mathbf{I}_g = [\mathbf{I}, -\mathbf{I}]$, on peut écrire :

$$\begin{aligned} \mathrm{prox}_{\frac{\gamma}{\delta}g([\mathbf{I}, -\mathbf{I}]\bullet)}([\mathbf{Y}_1, \mathbf{Y}_2]^t) &= [\mathbf{Y}_1, \mathbf{Y}_2]^t - \frac{1}{2}\mathbf{I}_g^t\mathbf{I}_g[\mathbf{Y}_1, \mathbf{Y}_2]^t + \frac{1}{2}\mathbf{I}_g^t\mathrm{prox}_{\frac{2\gamma}{\delta}g(\bullet)}(\mathbf{I}_g[\mathbf{Y}_1, \mathbf{Y}_2]^t), \\ &= \frac{1}{2}\left([\mathbf{Y}_1 + \mathbf{Y}_2, \mathbf{Y}_1 + \mathbf{Y}_2]^t + \mathbf{I}_g^t\mathrm{prox}_{\frac{2\gamma}{\delta}g(\bullet)}(\mathbf{Y}_1 - \mathbf{Y}_2)\right). \end{aligned} \quad (3.79)$$

Il existe maintenant une solution explicite de l'opérateur proximal qui est applicable à chaque colonne de la matrice d'entrée $\mathbf{y}_{:i}$, il s'agit d'une fonction de seuillage doux par bloc, notée STB [90] :

$$\left(\text{prox}_{\gamma g(\bullet)}(\mathbf{Y})\right)_{:i} = (\text{STB}_{\gamma}(\mathbf{Y}))_{:i} = \max\left(1 - \frac{\gamma}{\|\mathbf{y}_{:i}\|_F^2}, 0\right) \mathbf{y}_{:i}. \quad (3.80)$$

Cette fonction s'applique de la même manière que \mathbf{Y} soit une matrice réelle ou complexe. On peut donc actualiser les matrices $\mathbf{C}_{(\kappa+1)}^{(\tau)}$ et $\mathbf{D}_{(\kappa+1)}^{(\tau)}$ en appliquant les opérations suivantes :

$$\mathbf{D}_{(\kappa+1)}^{(\tau+1)} = \frac{1}{2} \left(\mathbf{Y}_1 + \mathbf{Y}_2 + \text{STB}_{\frac{2\gamma}{\delta}}(\mathbf{Y}_1 - \mathbf{Y}_2) \right), \quad (3.81)$$

$$\mathbf{C}_{(\kappa+1)}^{(\tau)} = \frac{1}{2} \left(\mathbf{Y}_1 + \mathbf{Y}_2 - \text{STB}_{\frac{2\gamma}{\delta}}(\mathbf{Y}_1 - \mathbf{Y}_2) \right). \quad (3.82)$$

3.A.5 Critère d'arrêt

Les conditions suffisantes d'optimalité de l'ADMM dans notre situation sont :

$$\mathbf{ResP} = \sum_{t=1}^T \mathbf{B}^{(t)} - \Theta^{(t)} + \sum_{\tau=1}^{T-1} \mathbf{C}^{(\tau)} + \mathbf{D}^{(\tau+1)} - \Theta^{(\tau)} - \Theta^{(\tau+1)} = 0, \quad (3.83)$$

$$\sum_{t=1}^T \nabla l(\Theta^{(t)}) + \delta \mathbf{U}_B^{(t)} + \sum_{\tau=1}^{T-1} (\mathbf{U}_C^{(\tau)} + \mathbf{U}_D^{(\tau+1)}) = 0, \quad (3.84)$$

$$0 \in \sum_{t=1}^T \partial \lambda \|\mathbf{B}^{(t)}\|_1 + \delta \mathbf{U}_B^{(t)} + \sum_{\tau=1}^{T-1} \nabla \gamma g(\mathbf{D}^{(\tau+1)} - \mathbf{C}^{(\tau)}) + \delta (\mathbf{U}_C^{(\tau)} + \mathbf{U}_D^{(\tau+1)}). \quad (3.85)$$

Avec $\nabla \bullet$ et $\partial \bullet$ respectivement les gradients et les sous différentiels, utilisés selon que la fonction soit dérivable en tous points ou non. La condition de l'Équation (3.83) introduit une matrice de résidus \mathbf{ResP} qui est nulle si toutes les contraintes du problème de l'Équation (3.46) sont respectées. Les conditions des Équations (3.84) et (3.85) précisent que les dérivées partielles selon les matrices $\Theta^{(t)}$ et $\mathbf{B}^{(t)}$, $\mathbf{C}^{(\tau)}$, $\mathbf{D}^{(\tau)}$ ($\forall t \in \{1, \dots, T\}$, $\forall \tau \in \{1, \dots, T-1\}$) doivent respectivement être nulles. La fonction $l(\bullet)$ est définie dans l'Équation (3.48).

On pourra facilement remarquer que pour chaque itération $\kappa + 1$, la condition d'optimalité de l'Équation (3.85) est toujours valide après avoir mis à jour les variables $\mathbf{B}_{(\kappa+1)}$, $\mathbf{C}_{(\kappa+1)}$, $\mathbf{D}_{(\kappa+1)}$. Comme la variable $\Theta_{(\kappa+1)}$ minimise $\mathcal{L}_{\delta}(\Theta, \mathbf{B}_{(\kappa)}, \mathbf{C}_{(\kappa)}, \mathbf{D}_{(\kappa)}, \mathbf{U}_{B,(\kappa)}, \mathbf{U}_{C,(\kappa)}, \mathbf{U}_{D,(\kappa)})$ on a :

$$\begin{aligned} & \sum_{t=1}^T \nabla l(\Theta_{(\kappa+1)}^{(t)}) + \delta \mathbf{U}_{B,(\kappa+1)}^{(t)} + \sum_{\tau=1}^{T-1} (\mathbf{U}_{C,(\kappa+1)}^{(\tau)} + \mathbf{U}_{D,(\kappa+1)}^{(\tau+1)}) \\ &= \delta \sum_{t=1}^T \mathbf{B}_{(\kappa+1)}^{(t)} - \mathbf{B}_{(\kappa)}^{(t)} + \delta \sum_{\tau=1}^{T-1} \mathbf{C}_{(\kappa+1)}^{(\tau)} - \mathbf{C}_{(\kappa)}^{(\tau)} + \mathbf{D}_{(\kappa+1)}^{(\tau+1)} - \mathbf{D}_{(\kappa)}^{(\tau+1)}. \end{aligned} \quad (3.86)$$

Ce qui implique, en plus de la matrice de résidus \mathbf{ResP} , qu'une autre matrice de résidus peut être introduite notée \mathbf{ResS} qui s'écrit :

$$\mathbf{ResS}_{(\kappa+1)} = \delta \sum_{t=1}^T \mathbf{B}_{(\kappa+1)}^{(t)} - \mathbf{B}_{(\kappa)}^{(t)} + \delta \sum_{\tau=1}^{T-1} \mathbf{C}_{(\kappa+1)}^{(\tau)} - \mathbf{C}_{(\kappa)}^{(\tau)} + \mathbf{D}_{(\kappa+1)}^{(\tau+1)} - \mathbf{D}_{(\kappa)}^{(\tau+1)}. \quad (3.87)$$

\mathbf{ResS} doit correspondre à une matrice nulle lorsque une solution au problème de l'Équation (3.46) est trouvée. De plus, comme démontré dans [37], il existe une borne inférieure sur la fonction de coût du problème d'optimisation de l'Équation (3.46), dont la valeur minimum est notée min_* (associée à la matrice $\Theta_*^{(t)}$), telle que :

$$\sum_{t=1}^T \left[l(\Theta_{(\kappa)}^{(t)}) + \lambda \|\mathbf{B}_{(\kappa)}^{(t)}\|_{od,1} \right] + \gamma \sum_{\tau=1}^{T-1} g\left(\mathbf{D}_{(\kappa)}^{(\tau+1)} - \mathbf{C}_{(\kappa)}^{(\tau)}\right) - \min_* \leq \epsilon_a \|\mathbf{ResR}_{(\kappa)}\|_F + \epsilon_b \|\mathbf{ResS}_{(\kappa)}\|_F, \quad (3.88)$$

$$\epsilon_a^2 = \sum_{t=1}^T \|\mathbf{U}_{B,(\kappa)}^{(t)}\|_F^2 + \sum_{\tau=1}^{T-1} \|\mathbf{U}_{C,(\kappa)}^{(\tau)}\|_F^2 + \|\mathbf{U}_{D,(\kappa)}^{(\tau)}\|_F^2, \quad (3.89)$$

$$\epsilon_b^2 = \sum_{i=0}^N \|\Theta_{(\kappa)}^{(t)} - \Theta_*^{(t)}\|_F^2. \quad (3.90)$$

Cette borne suggère que les normes des deux matrices de résidus doivent être faibles, soit un critère d'arrêt qui satisfait les deux conditions suivantes :

$$\|\mathbf{ResR}_{(\kappa)}\|_F \leq \epsilon_R, \quad (3.91)$$

$$\|\mathbf{ResS}_{(\kappa)}\|_F \leq \epsilon_S. \quad (3.92)$$

Pour sélectionner les variables ϵ_R , et ϵ_S on suit les recommandations de [37], ce qui donne :

$$\epsilon_R = \epsilon_{abs} \sqrt{TI^2 + 2(I^2(T-1))} + \epsilon_{rel} \max(\epsilon_c, \epsilon_d), \quad (3.93)$$

$$\epsilon_S = \epsilon_{abs} \sqrt{TI^2 + 2(I^2(T-1))} + \epsilon_{rel} \epsilon_a, \quad (3.94)$$

$$\epsilon_c^2 = \sum_{t=1}^T \|\mathbf{B}_{(\kappa)}^{(t)}\|_F^2 + \sum_{\tau=1}^{T-1} \|\mathbf{C}_{(\kappa)}^{(\tau)}\|_F^2 + \|\mathbf{D}_{(\kappa)}^{(\tau)}\|_F^2, \quad (3.95)$$

$$\epsilon_d^2 = \sum_{t=1}^T \|\Theta_{(\kappa)}^{(t)}\|_F^2 + \sum_{\tau=1}^{T-1} \|\Theta_{(\kappa)}^{(\tau)}\|_F^2 + \|\Theta_{(\kappa)}^{(\tau)}\|_F^2. \quad (3.96)$$

ϵ_{abs} et ϵ_{rel} sont des variables choisies en fonction du problème, on fixe dans notre cas $\epsilon_{rel} = \epsilon_{abs} = 10^{-4}$ pour toutes les applications considérées dans ce chapitre.

3.A.6 Mise à jour du paramètre de l'ADMM

La sélection du paramètre δ est essentielle pour que l'algorithme ADMM ait un comportement satisfaisant. En effet, l'algorithme est assuré de converger quelle que soit la valeur de δ . Cependant, la vitesse de convergence peut devenir extrêmement lente si la variable δ est mal sélectionnée. Afin d'avoir une implémentation moins dépendante de ce paramètre, on propose d'appliquer à chaque itération une mise à jour de δ , proposée dans [37], telle que :

$$\delta_{(\kappa+1)} = \begin{cases} \zeta_1 \delta_{(\kappa)} & , \text{ si } \|\mathbf{ResR}_{(\kappa)}\|_F \geq \zeta_2 \|\mathbf{ResS}_{(\kappa)}\|_F, \\ \frac{\delta_{(\kappa)}}{\zeta_1} & , \text{ si } \|\mathbf{ResS}_{(\kappa)}\|_F \geq \zeta_2 \|\mathbf{ResR}_{(\kappa)}\|_F, \\ \delta_{(\kappa)} & \text{ sinon.} \end{cases} \quad (3.97)$$

Avec $\zeta_1 > 0$ et $\zeta_2 > 0$, on utilise $\zeta_1 = 2$ et $\zeta_2 = 10$ [37]. L'idée, derrière ces mises à jour, est une adaptation afin de diminuer les normes des deux matrices de résidus en même temps. Une valeur de δ trop forte aura tendance à marquer peu de différences entre les itérations, et trop faible à ne pas assez prendre en compte les effets des régularisations.

Chapitre 4

Contexte : Décomposition matricielle et tensorielle pour l'extraction d'informations structurées

Les mesures dynamiques de CFs forment un signal multidimensionnel qui évolue en fonction du temps, noté par une matrice $\mathbf{X} \in \mathbb{R}^{L \times T}$, où L est le nombre de CFs et T le nombre d'échantillons temporels. Des techniques de factorisation de matrices, qui sont des méthodes d'apprentissage non supervisées, peuvent être utilisées afin d'extraire les informations structurées contenues dans ces matrices. Ce sont ces méthodes qui seront employées afin d'analyser le graphe multiplex de CF obtenu. Une introduction à l'apprentissage par décomposition matricielle est proposée dans la Section 4.1.

Cette représentation matricielle est cependant restrictive lorsque les données se déclinent selon d'autres modes, comme par exemple, la fréquence ou différentes occurrences d'une crise d'un même patient. Dans ce cas, au lieu de vivre sur deux dimensions, les données sont représentées en D dimensions et peuvent être représentées sous forme d'un tenseur. Une brève présentation sur les tenseurs est fournie en Section 4.2. Enfin, en Section 4.3 une liste des décompositions tensorielles usuelles est proposée, il s'agit de l'extension naturelle aux tenseurs des méthodes de décompositions de matrices.

4.1 Apprentissage par décomposition matricielle

4.1.1 La décomposition matricielle

Definition 12 : Décomposition matricielle

Une décomposition matricielle est la factorisation d'une matrice $\mathbf{X} \in \mathbb{R}^{L \times T}$ en produit de deux matrices $\mathbf{F} \in \mathbb{R}^{L \times K}$ et $\mathbf{V} \in \mathbb{R}^{T \times K}$. Généralement on a $K \ll \min(L, T)$ si l'on souhaite faire de la compression ou de la réduction d'informations. La multiplication des matrices \mathbf{F} et \mathbf{V} forment une approximation de la matrice \mathbf{X} . On a donc :

$$\mathbf{X} \approx \mathbf{F}\mathbf{V}^t. \quad (4.1)$$

Remarque : Dans le contexte d'une matrice de CFs \mathbf{X} , les K composantes de la matrice \mathbf{F} forment **les facteurs** de CFs. Idéalement, ces facteurs contiennent des groupes de CFs qui caractérisent un état clinique de la crise. Les K composantes de la matrice \mathbf{V} sont **les profils d'activation** temporelle de chaque facteur de CFs. Afin d'estimer des composantes \mathbf{F} et \mathbf{V} qui contiennent des informations utiles pour analyser le jeu de données, il est souvent nécessaire de faire des a priori sur la forme qu'elles doivent avoir. Cette phase de modélisation du problème est essentielle, il en découle alors un problème d'optimisation de la forme :

$$\underset{\mathbf{F}, \mathbf{V}}{\operatorname{argmin}} \operatorname{Att}(\mathbf{X}, \mathbf{F}, \mathbf{V}) + \lambda \operatorname{Reg}(\mathbf{F}, \mathbf{V}), \quad (4.2)$$

$$\text{s.t.} \quad \operatorname{Cont}(\mathbf{F}, \mathbf{V}). \quad (4.3)$$

Où $\text{Att}(\bullet)$ est la **fonction d'attache aux données** mesurant la similarité entre la matrice \mathbf{X} et la matrice reconstituée par les composantes \mathbf{F} et \mathbf{V} , cette différence est appelée résidu. λ est le paramètre associé à la **fonction de régularisation** $\text{Reg}(\mathbf{F}, \mathbf{V})$ et $\text{Cont}(\mathbf{F}, \mathbf{V})$ est une **contrainte**.

Une fonction d'attache aux données est généralement choisie en fonction de l'hypothèse que l'on fait sur les résidus. Dans le cadre de la décomposition de données de CF, la fonction d'attache aux données utilisées est :

$$\text{Att}(\mathbf{X}, \mathbf{F}, \mathbf{V}) = || \mathbf{X} - \mathbf{FV}^t ||_F^2 . \quad (4.4)$$

Il s'agit de l'erreur quadratique moyenne dans le cas où l'on fait l'hypothèse que les résidus suivent une loi gaussienne. Au lieu de la norme de Frobenius des résidus, d'autres fonctions existent si l'on considère que les résidus suivent une loi exponentielle [208], ou des distributions plus complexes [50].

L'objectif d'une fonction de régularisation est de pénaliser des solutions jugées non pertinentes, et se base sur un a priori que l'on a sur la structure des composantes. La somme des fonctions de régularisation et d'attache aux données forme la **fonction de coût** qui doit être minimisée. La contrainte est une notion plus forte que celle de la régularisation. Alors seuls les résultats respectant la contrainte imposée sur les composantes \mathbf{F} et \mathbf{V} sont admissibles. Les fonctions de régularisation et les contraintes sont généralement choisies pour : (a) produire des résultats cohérents par rapport à l'application qui sont interprétables (cette notion qualitative sera définie et illustrée en Section 4.1.3); (b) Surmonter des problèmes algorithmiques pour calculer une solution du critère de l'Équation (4.2).

Le problème de l'Équation (4.2) est la plupart du temps non convexe même lorsque $K = 1$. Cela vient de la forme multiplicative \mathbf{FV}^t . Cela implique que la solution trouvée à partir de n'importe quel algorithme n'est pas nécessairement le minimum global de l'Équation (4.2). Ainsi, on ne cherche pas forcément le minimum global de la fonction, mais un résultat satisfaisant. Pour atteindre ce résultat, on peut initialiser la recherche de notre minimum par des matrices que l'on pense proches d'une solution satisfaisante, ou relancer plusieurs fois l'algorithme avec des initialisations différentes afin de garder le résultat qui minimise l'Équation (4.2). Finalement, précisons que si on fixe la matrice \mathbf{F} ou \mathbf{V} , le problème de l'Équation (4.2) se réduit à un problème de régression convexe si la fonction d'attache aux données, les fonctions de régularisations et les contraintes sont convexes. Ainsi, un algorithme permettant de résoudre une décomposition matricielle consiste souvent à alterner la régression au sens des moindres carrés de \mathbf{F} et \mathbf{V} après avoir fixé l'autre matrice [128], [103], [141].

4.1.2 Motivations pour le choix des régularisations et contraintes

Des régularisations et contraintes pertinentes pour notre application sont définies. Nous les étudions pour chacune des matrices \mathbf{F} et \mathbf{V} séparément.

Le mode des connectivités fonctionnelles

Definition 13 : Régularisation parcimonieuse

La régularisation parcimonieuse appliquée à la matrice \mathbf{F} correspond à [184], [94] :

$$\text{Reg}(\mathbf{F}) = || \mathbf{F} ||_1, \quad (4.5)$$

Justification : Le mode des CFs vit souvent dans un espace de grande dimension, cela constitue sa principale limitation. En effet, le nombre de CFs est quadratique du nombre d'électrodes. Ainsi 100 électrodes impliquent plus de 5000 mesures de CFs. De plus, pour certaines étapes d'une crise, on observe qu'un grand nombre de CFs peuvent être passivement impliquées dans le réseau cortical. Afin de réaliser une sélection semi-automatique des CFs les plus actives, nous proposons une régularisation parcimonieuse de la matrice \mathbf{F} . C'est la même régularisation que celle utilisée dans la partie III. On précise qu'il s'agit d'une relaxation de la fonction qui compte le nombre d'éléments non-nuls dans la matrice \mathbf{F} . L'avantage de cette relaxation est qu'il s'agit d'une fonction convexe. Cependant, des artefacts peuvent plus facilement apparaître en utilisant cette définition relaxée de la parcimonie. \square

Le mode temporel

Definition 14 : Contrainte elastic-net

La contrainte elastic-net [210] permet de rendre parcimonieux les profils d'activation tout en imposant qu'ils appartiennent à la boule unité. Cette contrainte s'écrit :

$$\mathbf{Cont}(\mathbf{v}_{:k}) : \alpha_1 \|\mathbf{v}_{:k}\|_1 + \|\mathbf{v}_{:k}\|_F^2 \leq 1, \quad \forall k \in \{1, \dots, K\}. \quad (4.6)$$

Avec α_1 paramètre modulant le niveau de parcimonie imposé sur les profils d'activation.

Justification : On cherche ici à favoriser l'estimation de graphes qui sont caractéristiques d'une étape particulière d'une crise, c'est-à-dire qui ne restent pas actifs durant l'intégralité de l'enregistrement. On propose alors d'ajouter une régularisation parcimonieuse sur les profils d'activation \mathbf{V} , et ainsi privilégier des graphes de CFs qui ont des profils d'activation localisés dans le temps. Notons que l'ajout simultané de régularisations parcimonieuses sur les matrices \mathbf{V} et \mathbf{F} peut produire des résultats peu pertinents [15], ou avec une forte colinéarité entre les composantes [76]. Alors, on contraint les profils d'activation temporelle à appartenir à la boule unité (la contrainte d'appartenance à la sphère unité n'étant pas convexe). Cette contrainte s'écrit :

$$\mathbf{Cont}(\mathbf{v}_{:k}) : \|\mathbf{v}_{:k}\|_F^2 \leq 1, \quad \forall k \in \{1, \dots, K\}. \quad (4.7)$$

Cette contrainte force l'énergie contenue dans la matrice \mathbf{X} d'être principalement représentée par la matrice \mathbf{F} , limitant les ambiguïtés d'échelles entre les deux matrices \mathbf{F} et \mathbf{V} . La contrainte elastic-net correspond simplement à additionner à cette contrainte la fonction permettant au profil d'activation d'être parcimonieux. \square

Definition 15 : Contrainte fused-lasso

Une extension connue d'elastic-net permettant d'avantager des profils d'activation compactes qui indiquent des étapes de la crise est dénommée fused-lasso [185], et s'écrit :

$$\mathbf{Cont}(\mathbf{v}_{:k}) : \alpha_1 \|\mathbf{v}_{:k}\|_1 + \alpha_2 \|\mathbf{\Gamma} \mathbf{v}_{:k}\|_1 + \|\mathbf{v}_{:k}\|_F^2 \leq 1, \quad \forall k \in \{1, \dots, K\}. \quad (4.8)$$

Avec α_1 paramètre modulant le niveau de parcimonie imposé sur les profils d'activation, α_2 indiquant le niveau de compacité des profils d'activation et $\mathbf{\Gamma} \in \mathbb{R}^{(T-1) \times T}$ qui s'écrit :

$$\mathbf{\Gamma} = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -1 & 1 \end{pmatrix}, \quad (4.9)$$

Justification : On suppose qu'il existe naturellement une cohérence temporelle entre deux graphes de CFs successifs. La fonction usuelle afin d'empêcher des changements trop abrupts dans les profils d'activation entre deux instants consécutifs est la régularisation de Thikonov, qui s'écrit :

$$\mathbf{Reg}(\mathbf{V}) = \sum_{t=2}^T \sum_{k=1}^K (v_{tk} - v_{(t-1)k})^2 = \|\mathbf{\Gamma} \mathbf{V}\|_F^2, \quad (4.10)$$

Il s'agit principalement d'une fonction permettant de lisser les profils d'activation privilégiant des valeurs d'activation consécutives similaires. Cependant, si l'on suppose désormais uniquement une forte cohérence temporelle des profils d'activation pour chaque étape d'une crise, des changements abrupts peuvent alors subvenir entre deux étapes consécutives. Alors, la régularisation par les variations totales est plus adaptée. Cette régularisation s'écrit :

$$\mathbf{Reg}(\mathbf{V}) = \|\mathbf{\Gamma} \mathbf{V}\|_1, \quad (4.11)$$

Alors, l'addition de la contrainte elastic-net à celle des variations totales permet de favoriser des activations parcimonieuses et compactes, qui pourront être vues comme des étapes de la crise. \square

Les deux modes simultanément

Definition 16 : Contrainte de non négativité

Cette contrainte impose aux facteurs de CFs et aux profils d'activation de contenir uniquement des valeurs positives ou nulles et s'écrit :

$$\text{Cont}(\mathbf{V}, \mathbf{F}) : f_{lk} \geq 0, \quad v_{tk} \geq 0, \quad \forall l \in 1, \dots, L, \quad \forall t \in 1, \dots, T, \quad \forall k \in 1, \dots, K. \quad (4.12)$$

Pour simplifier la notation, on notera par la suite $\mathbf{F} \geq 0$ et $\mathbf{V} \geq 0$ pour préciser que tous les éléments des matrices \mathbf{F} et \mathbf{V} doivent être non négatifs.

Justification : Comme les CFs sont souvent des mesures non-négatives (comprises entre 0 et 1), il est naturel d'imposer aux matrices \mathbf{F} et \mathbf{V} d'être aussi non-négatives. Il s'agit, comme la régularisation parcimonieuse, d'une contrainte aidant à la visualisation des résultats. Un exemple d'application de la non-négativité pour obtenir des résultats plus facilement interprétables est fourni en Section 4.3. \square

4.1.3 Exemples usuels de décomposition matricielle

Décomposition en valeurs singulières (SVD)

On considère la décomposition de l'Équation (4.2) lorsque aucune régularisation ni contrainte n'est appliquée.

$$\underset{\mathbf{F}, \mathbf{V}}{\text{argmin}} \|\mathbf{X} - \mathbf{F}\mathbf{V}^t\|_F^2. \quad (4.13)$$

Il s'agit alors de trouver la meilleure approximation de rang K minimisant la norme de Frobenius. Ce problème a une infinité de solutions. Cependant il existe une décomposition spécifique bien connue, appelée décomposition en valeurs singulières telle que [195] :

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^t, \quad (4.14)$$

Avec comme contraintes que $\mathbf{U} \in \mathbb{R}^{L \times L}$ et $\mathbf{V} \in \mathbb{R}^{T \times T}$ sont des matrices unitaires (appelées respectivement matrices singulières gauche et droite) et $\mathbf{\Sigma} \in \mathbb{R}^{L \times T}$ est une matrice diagonale dont ses éléments sont positifs et décroissants (en considérant $L \leq T$ on a $\sigma_{11} \geq \sigma_{22} \geq \dots \geq \sigma_{LL} \geq 0$). Les vecteurs \mathbf{u}_l et \mathbf{v}_l sont les vecteurs singuliers associés à la valeur singulière σ_l . Précisons que cette décomposition est unique si toutes les paires de valeurs singulières sont différentes ($\sigma_{ii} \neq \sigma_{jj} \forall (i, j)$), ce qui est souvent le cas sur les jeux de données réelles qui sont bruitées.

On note $\mathbf{U}^{(K)} \in \mathbb{R}^{L \times K}$, $\mathbf{V}^{(K)} \in \mathbb{R}^{T \times K}$ et $\mathbf{\Sigma}^{(K)} \in \mathbb{R}^{K \times K}$ les matrices contenant les K premiers éléments singuliers de \mathbf{U} , \mathbf{V} et $\mathbf{\Sigma}$. Ainsi, la matrice $\mathbf{U}^{(K)}\mathbf{\Sigma}^{(K)}\mathbf{V}^{(K)t}$ est la meilleure approximation de rang K de la matrice \mathbf{X} . En posant $\mathbf{F} = \mathbf{U}^{(K)}\mathbf{\Sigma}^{(K)}$ et $\mathbf{V} = \mathbf{V}^{(K)}$ on trouve une solution à l'Équation (4.13). La décomposition en valeurs singulières est additive, on parle aussi de décomposition gloutonne, cela s'explique car les composantes singulières peuvent être obtenues successivement. En effet, il est possible d'obtenir les $K^{\text{ième}}$ composantes en résolvant le problème d'optimisation suivant :

$$\underset{\mathbf{u}, \sigma, \mathbf{v}}{\text{argmin}} \|\mathbf{X} - \mathbf{U}^{(K-1)}\mathbf{\Sigma}^{(K-1)}\mathbf{V}^{(K-1)t} - \mathbf{u}\sigma\mathbf{v}^t\|_F^2, \quad (4.15)$$

$$\text{s.t.} \quad \mathbf{u}^t\mathbf{u} = 1, \quad \mathbf{v}^t\mathbf{v} = 1. \quad (4.16)$$

En d'autres termes, les composantes K forment la meilleure approximation de rang 1 de la matrice \mathbf{X} après lui avoir soustrait sa meilleure approximation de rang $K - 1$.

On illustre l'application de la SVD sur un modèle de matrice de CFs dans la Figure 4.1. Le modèle utilisé correspond à trois périodes d'activation de CFs. Toutes les CFs qui s'activent dans la première période s'activent aussi dans la seconde. Les CFs activées dans la troisième période sont toutes différentes de celles activées dans les deux premières. On remarque que le premier facteur de CFs contient les valeurs en moyenne les plus élevées. Au contraire, la quatrième composante ne semble quasiment pas contenir d'information. En effet, les trois premières composantes semblent suffisantes pour approximer le jeu de données. Cela illustre le caractère compressif de cette décomposition [172].

Effectivement, la SVD est souvent utilisée pour trouver des plans de projections afin de visualiser les données, principe de l'analyse en composantes principales [195].

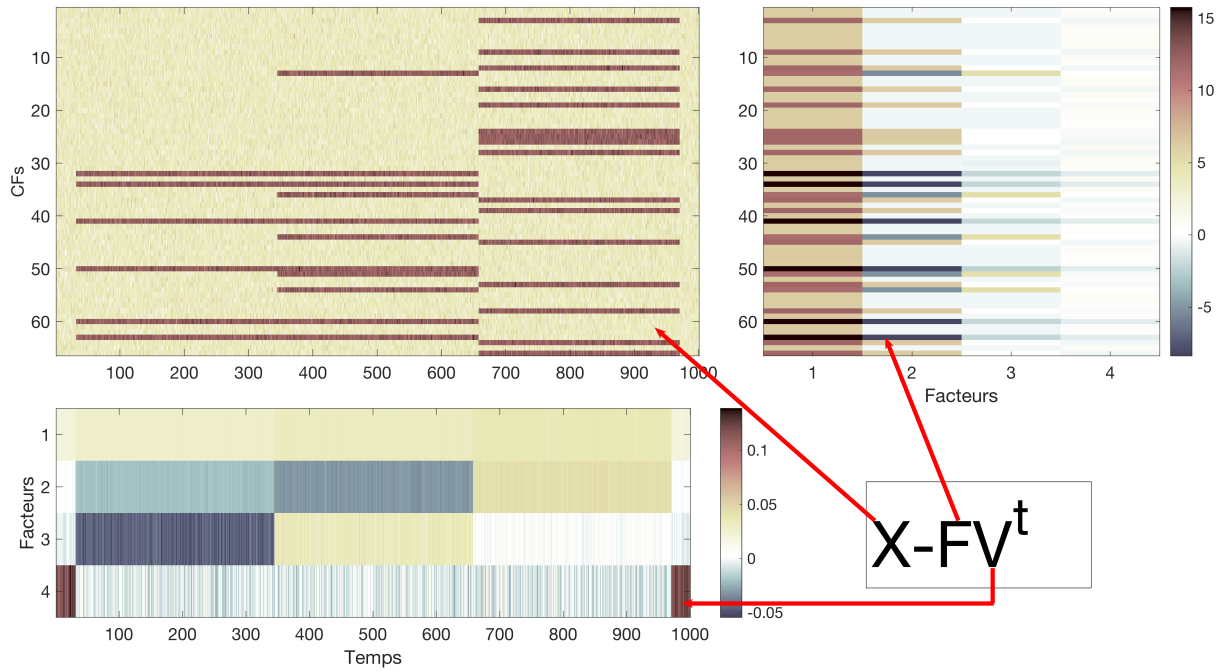


FIGURE 4.1 – Illustration des 4 premières composantes obtenues via la décomposition en valeur singulière (SVD).

Factorisation en matrices non-négatives (NMF)

Le caractère compressif de la SVD est souhaitable dans de nombreuses applications [195]. Cependant, il n'y a aucune assurance que les composantes obtenues porteront des informations neuro-physiologiques distinctes et facilement interprétables visuellement. Ainsi, lorsque l'on travaille avec un jeu de données à valeurs non négatives, une factorisation en matrices non-négatives (NMF) [128], [133] favorise l'interprétabilité des composantes obtenues [205]. Dans ce manuscrit cette notion se définit comme suit :

Definition 17 : Composante interprétable

Composante contenant une information structurée associée à une réalité neuro-physiologique qui puisse être expliquée cliniquement.

On cherche alors les matrices \mathbf{F} et \mathbf{V} qui minimisent :

$$\operatorname{argmin}_{\mathbf{F} \geq 0, \mathbf{V} \geq 0} \|\mathbf{X} - \mathbf{FV}^t\|_F^2. \quad (4.17)$$

On précise que la définition de la NMF se généralise pour toutes fonctions d'attaché aux données $\mathbf{Att}(\bullet)$. L'ajout de la contrainte de non-négativité implique que les composantes ne sont plus orthogonales [65]. Les composantes ne sont donc ici plus additives comme pour la SVD et pour deux valeurs de K différentes ; l'ensemble des composantes de \mathbf{F} et \mathbf{V} peut alors être modifié. Notons que des extensions naturelles de la NMF existent, consistant à rajouter des régularisations et/ou contraintes diverses (comme la parcimonie [97], [119] ou une contrainte de lissage [131]).

On illustre l'application de la NMF sur le modèle de matrice de CFs dans la Figure 4.2. On observe la perte du caractère compressif de la SVD, dans le sens où l'information semble plus répartie entre les composantes. En revanche, on remarque qu'au moins pour les trois premières composantes, le découpage temporel des profils d'activation est clair et indique précisément les trois étapes d'activation des CFs. Par conséquent, les facteurs de CFs sont eux aussi très lisibles et indiquent l'ensemble des CFs actives pour chaque étape.

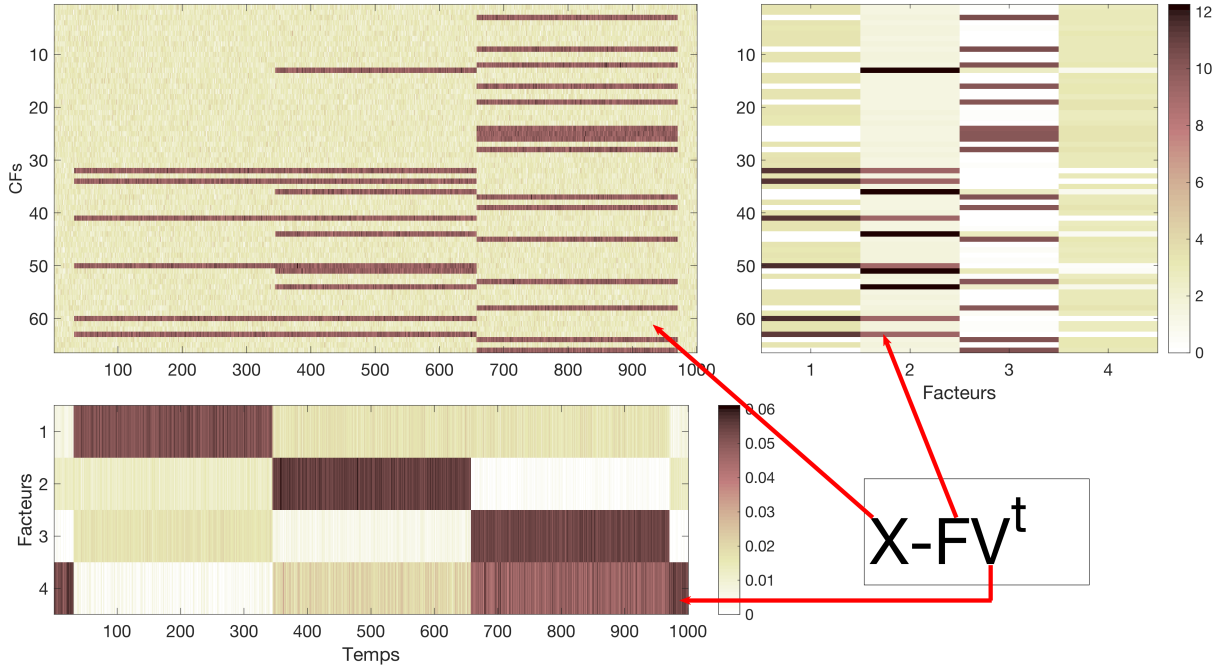


FIGURE 4.2 – Illustration de la décomposition en matrices non-négatives (NMF).

Partitionner avec k-means

k -means est l'un des algorithmes les plus connus pour faire du partitionnement de données aussi appelé clustering non supervisé [103]. Nous apportons un soin particulier à sa présentation car cet algorithme sera utilisé dans la suite du manuscrit. De plus, il s'agit d'un exemple intéressant de décomposition matricielle avec pour objectif de produire des composantes facilement interprétables.

L'objectif de k -means, appliqué sur les colonnes de \mathbf{X} , est de trouver P classes de CFs. La colonne l de \mathbf{X} appartient à la classe p (notée $l \in C_p$) si le centroïde le plus proche qui lui est associé est θ_p . Il s'agit donc de trouver P centroïdes $\theta_p \in \mathbb{R}^T$ tel que le critère suivant soit minimisé :

$$\arg \min_{\theta_1, \theta_2, \dots, \theta_P} \sum_{p=1}^P \sum_{l \in C_p} \|\mathbf{x}_l - \theta_p\|_F^2. \quad (4.18)$$

L'algorithme de Lloyd [103] est couramment utilisé pour résoudre ce problème. Notons cependant que ce critère peut-être modifié pour correspondre à un problème de décomposition matricielle. On appelle $\mathbf{S} \in \mathbb{R}^{L \times P}$ la matrice indicatrice, elle est construite de telle sorte que :

$$s_{lp} = \begin{cases} 1 & \text{si } l \in C_p, \\ 0 & \text{sinon.} \end{cases} \quad (4.19)$$

On peut donc écrire la fonction de coût de l'Équation (4.18) [24], [64] ainsi :

$$\begin{aligned} \sum_{p=1}^P \sum_{l \in C_p} \|\mathbf{x}_l - \theta_p\|_F^2 &= \sum_{p=1}^P \sum_l s_{lp} \|\mathbf{x}_l - \theta_p\|_F^2, \\ &= \|\mathbf{X} - \mathbf{S}\Theta\|_F^2. \end{aligned} \quad (4.20)$$

Alors, on observe bien une décomposition matricielle dont le nombre de facteur est $K = P$. Il s'agit alors du problème de l'Équation (4.2), en posant $\mathbf{F} = \mathbf{S}$ et $\mathbf{V} = \Theta$, et en imposant à \mathbf{F} d'avoir une structure de matrice indicatrice de l'Équation (4.19).

Afin de préciser les contraintes appliquées sur \mathbf{F} , on remarque que :

$$\|\mathbf{X} - \mathbf{S}\Theta\|_F^2 = \text{Tr}(\mathbf{X}\mathbf{X}^t) - 2\text{Tr}(\mathbf{S}\Theta\mathbf{X}^t) + \text{Tr}(\mathbf{S}\Theta\Theta^t\mathbf{S}^t), \quad (4.21)$$

$$= \text{Tr}(\mathbf{X}\mathbf{X}^t) - \text{Tr}(\mathbf{S}\Theta\mathbf{X}^t). \quad (4.22)$$

Le passage de l'Équation (4.21) à l'Équation (4.22) s'explique car aucune contrainte n'est appliquée sur Θ , ainsi on sait que :

$$\Theta = (\mathbf{S}^t \mathbf{S})^{-1} \mathbf{S}^t \mathbf{X} = \mathbf{S}^\dagger \mathbf{X}. \quad (4.23)$$

Notons que ce résultat à une signification physique. Comme les colonnes de \mathbf{S} sont orthogonales, on obtient bien l'une des étapes de l'algorithme de Lloyd [103] correspondant au à la moyenne des éléments de chaque classe :

$$\theta_{p.} = \frac{\sum_{l=1}^L s_{lp} \mathbf{x}_l}{\sum_{l=1}^L s_{lp}} = \frac{1}{c_p} \sum_{l \in C_p} x_l, \quad (4.24)$$

avec c_p le cardinal de la classe C_p . Ainsi minimiser l'Équation (4.20) est équivalent à maximiser :

$$\begin{aligned} Tr(\mathbf{S} \Theta \mathbf{X}^t) &= Tr(\mathbf{S} (\mathbf{S}^t \mathbf{S})^{-1} \mathbf{S} \mathbf{X} \mathbf{X}^t), \\ &= Tr((\mathbf{S}^t \mathbf{S})^{-\frac{1}{2}} \mathbf{S}^t \mathbf{X} \mathbf{X}^t \mathbf{S} (\mathbf{S}^t \mathbf{S})^{-\frac{1}{2}}), \\ &= \|\mathbf{A}^t \mathbf{X}\|_F^2. \end{aligned} \quad (4.25)$$

où $\mathbf{A} \in \mathbb{R}^{L \times K}$ correspond à la matrice indicatrice normalisée, $\mathbf{A} = \mathbf{S} (\mathbf{S}^t \mathbf{S})^{-\frac{1}{2}}$, telle que :

$$a_{lp} = \begin{cases} \frac{1}{\sqrt{c_p}} & \text{si } l \in C_p, \\ 0 & \text{sinon.} \end{cases} \quad (4.26)$$

En remarquant que les colonnes de \mathbf{A} forment une base canonique d'un sous espace de dimension P de $\mathbb{R}^{L \times L}$, l'algorithme des k -means est équivalent à trouver la meilleure base canonique qui maximise [197] :

$$\operatorname{argmax}_{\mathbf{A}} \quad \|\mathbf{A}^t \mathbf{X}\|_F^2, \quad (4.27)$$

$$\text{s.t.} \quad \mathbf{a}_{.p} \text{ vecteur de base canonique.} \quad (4.28)$$

Cette contrainte de base canonique impose à la matrice \mathbf{A} d'être :

- Non-négative $a_{lp} \geq 0$
- Parcimonieuse car \mathbf{A} est une base standard de vecteurs
- Les colonnes sont orthonormales $\mathbf{A}^t \mathbf{A} = \mathbf{I}$, et donc appartiennent à la sphère unité $\|\mathbf{A}_{.p}\|_F = 1$.

Alors, la matrice \mathbf{F} qui correspond simplement à remplacer par des 1 les valeurs non nulles de \mathbf{A} (ainsi les colonnes n'appartiennent plus à la sphère unité) est une matrice fortement contrainte, à la fois orthogonale, non-négative et parcimonieuse.

On illustre l'application de k -means sur le modèle de matrice de CF dans la Figure 4.2. Le découpage obtenu est facilement lisible, on retrouve dans les facteurs de CFs les différents modes d'activation de la matrice \mathbf{X} . La matrice \mathbf{V} indique la durée d'activation de chacun de ces facteurs. En comparant avec la décomposition obtenue en utilisant la SVD définie dans l'Équation (4.13), ce résultat illustre bien comment l'apport de diverses contraintes permet de rendre une décomposition matricielle plus facilement interprétables.

Conclusion : Nous avons présenté dans cette section le principe de la décomposition matricielle, en soulignant l'importance de l'ajout de régularisations pour avoir une décomposition pertinente. Nous étendons maintenant le principe des décompositions matricielles aux tenseurs.

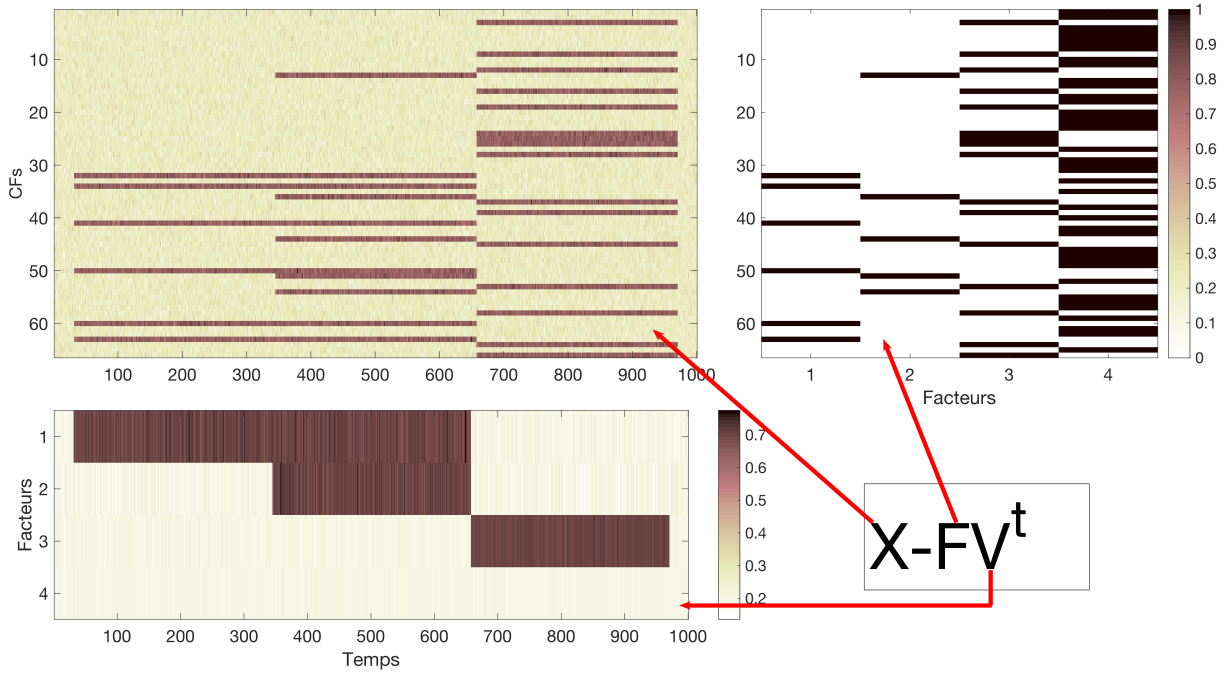


FIGURE 4.3 – Illustration de la décomposition matricielle liée à k -means.

4.2 Le concept de tenseur

Dans cette partie, nous présentons la notion de tenseur et les concepts utiles pour comprendre la suite du manuscrit.

4.2.1 Le tenseur

Un tenseur composé de M modes (noté \mathcal{X}) est préférablement appelé vecteur si $M = 1$ (noté dans ce cas \mathbf{x}) et matrice si $M = 2$ (notée dans ce cas \mathbf{X}). Ainsi un tenseur généralise le concept de vecteur et de matrice et permet par exemple d'écrire les variations des CFs en fonction du temps et d'autres modalités (fréquences, essais ...). On suggère [54] et [122] comme des excellentes introductions aux tenseurs et leurs décompositions, et aussi [176] pour plus de détails techniques. Dans la suite du manuscrit on s'intéresse principalement aux tenseurs de 3-modes, noté $\mathcal{X} \in \mathbb{R}^{L \times T \times S}$. Sans perdre en généralité, on précise que L est associé au mode des CFs, T au mode temporel et S aux différentes crises d'un même patient. Afin d'illustrer certains concepts, nous utiliserons le tenseur provenant de [122], avec $L = 3$, $T = 4$ et $S = 2$ et qui s'écrit :

$$\mathbf{X}_{::1} = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix}, \quad \mathbf{X}_{::2} = \begin{pmatrix} 13 & 16 & 19 & 22 \\ 14 & 17 & 20 & 23 \\ 15 & 18 & 21 & 24 \end{pmatrix}. \quad (4.29)$$

4.2.2 Matricer

Definition 18 : Matricer

Matricer ou encore appelé aplatir ou déployer un tenseur, est une procédure consistant à réarranger les éléments de \mathcal{X} afin de les placer dans une matrice.

Ce concept est important car il permet de transformer les opérations sur les tenseurs en calcul matriciel. Comme la notation formelle du matricage [122] n'aide ni à la compréhension, ni à l'implémentation de ce concept, nous l'illustrons directement. Pour un tenseur de 3-modes $\mathcal{X} \in \mathbb{R}^{L \times T \times S}$, il existe trois différents matricages : celui du mode des CFs noté $\mathbf{X}_{(L)} \in \mathbb{R}^{L \times TS}$, du mode temporel noté

$\mathbf{X}_{(T)} \in \mathbb{R}^{T \times LS}$ et du mode des crises noté $\mathbf{X}_{(S)} \in \mathbb{R}^{S \times LT}$. Dans le cas de notre exemple ils valent respectivement :

$$\mathbf{X}_{(L)} = \begin{pmatrix} 1 & 4 & 7 & 10 & 13 & 16 & 19 & 22 \\ 2 & 5 & 8 & 11 & 14 & 17 & 20 & 23 \\ 3 & 6 & 9 & 12 & 15 & 18 & 21 & 24 \end{pmatrix}, \quad (4.30)$$

$$\mathbf{X}_{(T)} = \begin{pmatrix} 1 & 2 & 3 & 13 & 14 & 15 \\ 4 & 5 & 6 & 16 & 17 & 18 \\ 7 & 8 & 9 & 19 & 20 & 21 \\ 10 & 11 & 12 & 22 & 23 & 24 \end{pmatrix}, \quad (4.31)$$

$$\mathbf{X}_{(S)} = \begin{pmatrix} 1 & 2 & 3 & \dots & 11 & 12 \\ 13 & 14 & 15 & \dots & 23 & 24 \end{pmatrix}. \quad (4.32)$$

On notera que pour le mode des CFs, $\mathbf{X}_{(L)}$ est la concaténation des tranches $\mathbf{X}_{:,s}$ du tenseur \mathcal{X} . Pour le mode temporel, $\mathbf{X}_{(T)}$ est la concaténation des tranches $\mathbf{X}_{:,s}^t$ du tenseur \mathcal{X} . On précise que ces ordonnancements sont des conventions et qu'il en existe d'autres. Cependant, elles ont beaucoup d'impacts sur les notations de certaines équations qui seront présentées ci-dessous. On utilise ici les conventions proposées dans [122] qui sont globalement les plus utilisées.

4.2.3 Les produits tensoriels

Tout comme on utilise couramment le produit matriciel, les tenseurs sont associés à deux produits : le produit vectoriel et le produit de contraction. Ils sont particulièrement intéressants car ils généralisent le produit matriciel usuel. Comme dans cette thèse on s'intéresse à l'écriture matricielle des tenseurs, il est nécessaire d'introduire les produits de Kronecker et de Kathri-Rao.

Le produit de contraction et le produit de Kronecker

Le produit de contraction permet de multiplier les tenseurs entre eux. Réalisé entre un tenseur de M modes et un tenseur de M' modes, il produit un tenseur composé de $(M + M' - 2)$ modes. Il existe un maximum de $\max(M, M')$ produits de contractions entre eux. Ainsi les deux produits de contractions possibles entre une matrice $\mathbf{X} \in \mathbb{R}^{L \times T}$ et $\mathbf{Y} \in \mathbb{R}^{L \times T}$, noté \bullet_1 et \bullet_2 sont :

$$\mathbf{X} \bullet_1 \mathbf{Y}^t = \mathbf{Y}^t \mathbf{X}, \quad (4.33) \quad \mathbf{X} \bullet_2 \mathbf{Y} = \mathbf{X} \mathbf{Y}^t (= \mathbf{Y}^t \bullet_1 \mathbf{X}). \quad (4.34)$$

Notons que l'on retrouve le produit matriciel usuel. Le produit de contraction est possible uniquement si la dimension du mode n de \mathbf{X} est par convention équivalente à la dimension du second mode de l'élément auquel il est multiplié.

La notation formelle de ce produit entre la mode temporel de \mathcal{X} et la matrice $\mathbf{V} \in \mathbb{R}^{T \times K}$, s'écrit :

$$(\mathcal{X} \bullet_2 \mathbf{V}^t)_{lks} = \sum_{t=1}^T x_{lts} v_{tk}. \quad (4.35)$$

Plus généralement, on peut considérer trois matrices $\mathbf{U} \in \mathbb{R}^{L \times K_L}$, $\mathbf{V} \in \mathbb{R}^{T \times K_T}$ et $\mathbf{W} \in \mathbb{R}^{S \times K_S}$, alors il existe un tenseur $\mathcal{G} \in \mathbb{R}^{K_L \times K_T \times K_S}$ tel que :

$$\mathcal{G} = \mathcal{X} \bullet_1 \mathbf{U}^t \bullet_2 \mathbf{V}^t \bullet_3 \mathbf{W}^t, \quad (4.36)$$

avec

$$g_{k_l k_t k_s} = \sum_{l=1}^L \sum_{t=1}^T \sum_{s=1}^S x_{lts} u_{lk_l} v_{tk_t} w_{sk_s}. \quad (4.37)$$

L'égalité de l'Équation (4.36) peut aussi s'inverser en :

$$\mathcal{X} = \mathcal{G} \bullet_1 \mathbf{U} \bullet_2 \mathbf{V} \bullet_3 \mathbf{W}. \quad (4.38)$$

Il s'agit d'une décomposition générale et couramment utilisée d'un tenseur. Alors, afin de matricer la formule de l'Équation (4.38), il est nécessaire d'introduire le produit de Kronecker.

Definition 19 : Produit de Kronecker (et application)

Le produit de Kronecker [190], [137] est noté $\mathbf{C} = \mathbf{A} \otimes \mathbf{B}$ avec $\mathbf{A} \in \mathbb{R}^{L \times T}$, $\mathbf{B} \in \mathbb{R}^{K \times N}$ et $\mathbf{C} \in \mathbb{R}^{LK \times TN}$, il s'écrit :

$$\mathbf{C} = \begin{pmatrix} a_{11}\mathbf{B} & \dots & a_{1T}\mathbf{B} \\ \vdots & & \vdots \\ a_{L1}\mathbf{B} & \dots & a_{LT}\mathbf{B} \end{pmatrix}, \quad (4.39)$$

Il permet d'écrire les trois versions matricées de l'Équation (4.38).

$$\mathbf{X}_{(L)} = \mathbf{U}\mathbf{G}_{(L)}(\mathbf{W} \otimes \mathbf{V})^t, \quad (4.40)$$

$$\mathbf{X}_{(T)} = \mathbf{V}\mathbf{G}_{(T)}(\mathbf{W} \otimes \mathbf{U})^t, \quad (4.41)$$

$$\mathbf{X}_{(S)} = \mathbf{W}\mathbf{G}_{(S)}(\mathbf{V} \otimes \mathbf{U})^t. \quad (4.42)$$

On notera que l'ordre dans lequel les matrices \mathbf{U} , \mathbf{V} et \mathbf{W} apparaissent dans les formules de matricage de l'Équation (4.40) est causé par les conventions de matricage utilisées [122].

Le produit vectoriel et le produit de Khatri-Rhao

Le produit vectoriel, appliqué entre un tenseur de M modes et un tenseur de M' modes, produit un tenseur composé de $(M + M')$ modes. Appliqué entre deux vecteurs $\mathbf{u} \in \mathbb{R}^L$, $\mathbf{v} \in \mathbb{R}^T$, il forme une matrice de rang 1 qui se note :

$$(\mathbf{u} \times \mathbf{v})_{lt} = \mathbf{u}_l \mathbf{v}_t. \quad (4.43)$$

De la même manière, un tenseur de 3-modes et de rang 1 [54] peut être créé en considérant en plus le vecteur $\mathbf{w} \in \mathbb{R}^S$, il s'écrit :

$$(\mathbf{u} \times \mathbf{v} \times \mathbf{w})_{lts} = \mathbf{u}_l \mathbf{v}_t \mathbf{w}_s. \quad (4.44)$$

De la même manière qu'il est possible de décomposer une matrice en somme de matrice de rang 1, un tenseur $\mathcal{X} \in \mathbb{R}^{L \times T \times S}$ de rang K est décomposable en une somme de tenseurs de rang 1 [54], cette décomposition peut s'écrire :

$$\mathcal{X} = \sum_{k=0}^K \mathbf{u}_{:k} \times \mathbf{v}_{:k} \times \mathbf{w}_{:k}, \quad (4.45)$$

avec $\mathbf{U} \in \mathbb{R}^{L \times K}$, $\mathbf{V} \in \mathbb{R}^{T \times K}$ et $\mathbf{W} \in \mathbb{R}^{S \times K}$. Il s'agit à nouveau d'une décomposition générale et couramment utilisée d'un tenseur. Il est alors important de pouvoir matricer la formule de l'Équation (4.45). Dans un premier temps, on peut remarquer que l'Équation (4.45) est équivalente à :

$$\mathcal{X} = \mathcal{I} \bullet_1 \mathbf{U} \bullet_2 \mathbf{V} \bullet_3 \mathbf{W}. \quad (4.46)$$

Avec $\mathcal{I} \in \mathbb{R}^{K \times K \times K}$ un tenseur hyperdiagonal tel que $i_{kkk} = 1 \forall k$, sinon 0. Dans ce cas, le matricage de l'Équation (4.46) est évident, il suffit d'utiliser les formules de l'Équation (4.40) car on retrouve une décomposition similaire à l'Équation (4.36). Cependant, au lieu d'utiliser le produit de Kronecker, le produit de Khatri-Rao [137] permet de profiter de l'abondance de valeurs nulles dans le tenseur \mathcal{I} afin d'éliminer de nombreux calculs inutiles.

Definition 20 : Produit de Khatri-Rao (et application)

Le produit de Khatri-Rao est noté $\mathbf{E} = \mathbf{A} \odot \mathbf{D}$, avec $\mathbf{D} \in \mathbb{R}^{K \times T}$ et $\mathbf{E} \in \mathbb{R}^{LK \times T}$ et s'écrit :

$$\mathbf{E} = \begin{pmatrix} a_{11}\mathbf{d}_{:1} & \dots & a_{1T}\mathbf{d}_{:T} \\ \vdots & & \vdots \\ a_{L1}\mathbf{d}_{:1} & \dots & a_{LT}\mathbf{d}_{:T} \end{pmatrix}. \quad (4.47)$$

Il permet d'écrire les trois formules de matricage des Équations (4.45) ou (4.46) :

$$\mathbf{X}_{(L)} = \mathbf{U}(\mathbf{W} \odot \mathbf{V})^t, \quad (4.48)$$

$$\mathbf{X}_{(T)} = \mathbf{V}(\mathbf{W} \odot \mathbf{U})^t, \quad (4.49)$$

$$\mathbf{X}_{(S)} = \mathbf{W}(\mathbf{V} \odot \mathbf{U})^t. \quad (4.50)$$

Notons que des versions plus complexes du produit de Khatri-Rao existent [176], et qui généralisent des situations où la matrice \mathcal{I} ne serait pas hyperdiagonale, mais avec une structure particulière contenant de nombreux zéros (hyperdiagonale par bloc par exemple).

4.2.4 Vectoriser**Definition 21 : Vectoriser**

Vectoriser un tenseur \mathcal{X} consiste à réarranger ses éléments dans un vecteur.

Appliquée au tenseur de l'Équation (4.29), cette opération se note :

$$\text{vec}(\mathcal{X}) = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ \vdots \\ 24 \end{pmatrix}. \quad (4.51)$$

Cette vectorisation permet de simplifier de nombreux calculs. En effet, cette opération ne modifie pas le résultat d'une norme de Frobenius (ou de n'importe quelle opération terme à terme), on a :

$$\|\mathcal{X}\|_F^2 = \|\text{vec}(\mathcal{X})\|_F^2. \quad (4.52)$$

Toujours dans l'objectif de simplifier certains calculs, il existe deux formules de compatibilité entre la vectorisation et d'une part le produit de Kronecker et d'autre part le produit de Khatri-Rao :

Definition 22 : Formules de compatibilités de la vectorisation

La formule de compatibilité entre la vectorisation et le produit de Kronecker appliquée à trois matrices \mathbf{A} , \mathbf{B} , \mathbf{M} se lit :

$$\text{vec}(\mathbf{AMB}^t)^t = \text{vec}(\mathbf{M})^t(\mathbf{A} \otimes \mathbf{B})^t. \quad (4.53)$$

La formule de compatibilité entre la vectorisation et le produit de Khatri-Rao est un cas particulier de la formule de l'Équation (4.53) quand \mathbf{M} est une matrice diagonale (que l'on note \mathbf{D}). Elle se note :

$$\text{vec}(\mathbf{ADB}^t)^t = \text{diag}(\mathbf{D})^t(\mathbf{A} \odot \mathbf{B})^t, \quad (4.54)$$

avec $\text{diag}(\mathbf{D})$ le vecteur contenant les éléments diagonaux de la matrice \mathbf{D}

Conclusion : Cette section présente l'ensemble des éléments utiles pour comprendre les tenseurs, les notions de matricage et de vectorisation, ainsi qu'une justification sur l'utilisation des produits de Kronecker et de Khatri-Rao. Ces notions sont la base des décompositions tensorielles usuelles présentées dans la section suivante.

4.3 Les décompositions tensorielles usuelles

Il existe un grand nombre de décompositions tensorielles, une liste est fournie dans [122]. Cependant, seulement deux sont couramment appliquées : la décomposition canonique polyadique (CP) et la décomposition en valeurs singulières d'ordre supérieur (HOSVD). Il est intéressant de noter qu'elles se revendiquent toutes deux comme une extension, pour les tenseurs, de la SVD (Équation (4.14)). En effet, la SVD est à la fois la meilleure décomposition en somme de matrice de rang 1 (selon la norme de Frobenius), et aussi une décomposition avec contrainte d'orthogonalité sur les matrices singulières. Cependant, il est impossible d'obtenir une décomposition qui conserve ces deux propriétés pour les tenseurs.

4.3.1 La décomposition Canonique Polyadique (CP)

La décomposition canonique polyadique (CP) aussi nommée PARAFAC, CANDECOMP, CAND, [122] est certainement la plus utilisée des décompositions tensorielles. Il s'agit d'une extension de la SVD dans le sens où elle permet d'approximer le tenseur \mathcal{X} en une somme de tenseurs de rang 1. En rappelant la transformation de la notation d'une somme de tenseur de rang 1 (Équation (4.45)) en sa version matricielle (Équation (4.40)), il s'agit de trouver les matrices $\mathbf{U} \in \mathbb{R}^{L \times K}$, $\mathbf{V} \in \mathbb{R}^{T \times K}$ et $\mathbf{W} \in \mathbb{R}^{S \times K}$ qui minimisent :

$$\underset{\mathbf{U}, \mathbf{V}, \mathbf{W}}{\operatorname{argmin}} \|\mathbf{X}_{(L)} - \mathbf{U}(\mathbf{W} \odot \mathbf{V})^t\|_F^2. \quad (4.55)$$

La décomposition est illustrée dans la Figure 4.4. Notons que la valeur minimale de K permettant de reconstituer le tenseur \mathcal{X} correspond au rang de ce tenseur. Il s'agit d'un sujet théorique très discuté dans la littérature, car contrairement au cas des matrices, trouver le rang d'un tenseur est NP-hard. Si le rang d'un tenseur est si difficile à trouver, c'est que les composantes obtenues ne sont pas additives et ne peuvent donc pas être obtenues séquentiellement. Pour chaque valeur de K fixée, toutes les composantes de chaque matrice minimisant l'Équation (4.55) doivent être de nouveau calculées. Ainsi, pour déterminer le rang de \mathcal{X} , il faut incrémenter la valeur de K et calculer sa décomposition CP jusqu'à obtenir une reconstitution exacte (notons qu'il existe des théorèmes fournissant une borne inférieure et supérieure de ce rang [176]). Il existe de nombreux algorithmes permettant de réaliser cette décomposition [177], [55], le plus intuitif étant l'alternance des moindres carrés [130]. Il s'agit d'alterner trois étapes calculant une régression des moindres carrés de chaque matrice après avoir fixé les deux autres. Son implémentation est fournie dans l'Algorithme (2).

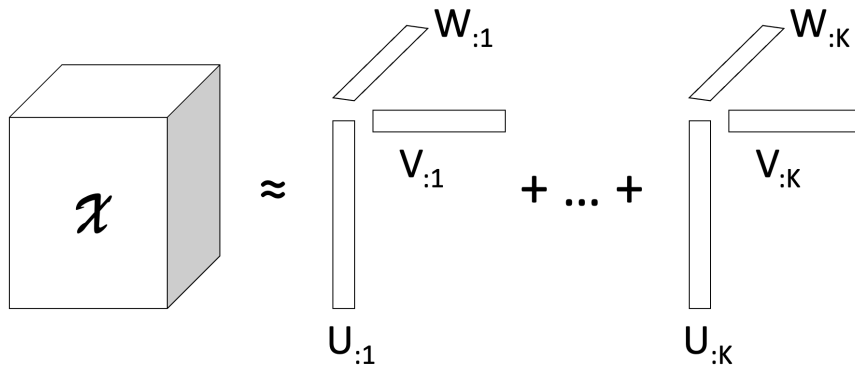


FIGURE 4.4 – Illustration de la décomposition canonique polyadique (CP).

Comme montré dans [132], le problème de l'Équation (4.55) peut être mal posé et ne pas admettre de solutions. Il est alors recommandé d'ajouter des contraintes de non-négativités sur les matrices \mathbf{U} , \mathbf{V} et \mathbf{W} , menant à une décomposition canonique polyadique non négative (nnCP) [119] qui admettra toujours une solution. Dans ce cas, on peut utiliser l'implémentation de [49].

Finalement, on notera que des extensions de la CP existent, non pas comme somme de tenseurs de rang 1, mais en somme de tenseurs avec une structure plus complexe. Ces décompositions sont appelées "Block term décomposition" (BTD) [61], [177].

Algorithm 2 Estimation \mathbf{U} , \mathbf{V} et \mathbf{W} via décomposition CP.

Require: $\mathcal{X} \in \mathbb{R}^{L \times T \times S}$, le paramètre de réduction K , l'incrément de tolérance $\epsilon > 0$, i_{max} , une première supposition pour $\mathbf{U}_0, \mathbf{V}_0, \mathbf{W}_0$

$i = 0$

while $i < i_{max}$ or $\| \mathbf{U}_i(\mathbf{W}_i \odot \mathbf{V}_i)^t \|_F^2 - \| \mathbf{U}_{i+1}(\mathbf{W}_{i+1} \odot \mathbf{V}_{i+1})^t \|_F^2 > \epsilon$ **do**

$\mathbf{U}_{i+1} = (\mathbf{W}_i \odot \mathbf{V}_i)^\dagger \mathbf{X}_{(L)}^t$

$\mathbf{V}_{i+1} = (\mathbf{W}_i \odot \mathbf{U}_{i+1})^\dagger \mathbf{X}_{(T)}^t$

$\mathbf{W}_{i+1} = (\mathbf{V}_{i+1} \odot \mathbf{U}_{i+1})^\dagger \mathbf{X}_{(S)}^t$

$i = i+1$

end while

4.3.2 La décomposition en valeurs singulières d'ordre supérieure (HOSVD)

Contrairement à la SVD, la décomposition CP n'a aucune raison d'estimer des matrices \mathbf{U} , \mathbf{V} et \mathbf{W} unitaires. La décomposition en valeurs singulières d'ordre supérieure (HOSVD) [62] est une autre extension possible de la SVD, qui cette fois, cherche les sous-espaces caractéristiques de chaque mode. La HOSVD exacte consiste à trouver trois matrices unitaires $\mathbf{U} \in \mathbb{R}^{L \times L}$, $\mathbf{V} \in \mathbb{R}^{T \times T}$ et $\mathbf{W} \in \mathbb{R}^{S \times S}$ solutions de l'équation suivante :

$$\mathbf{X}_{(L)} = \mathbf{U} \mathbf{G}_{(L)} (\mathbf{W} \otimes \mathbf{V})^t. \quad (4.56)$$

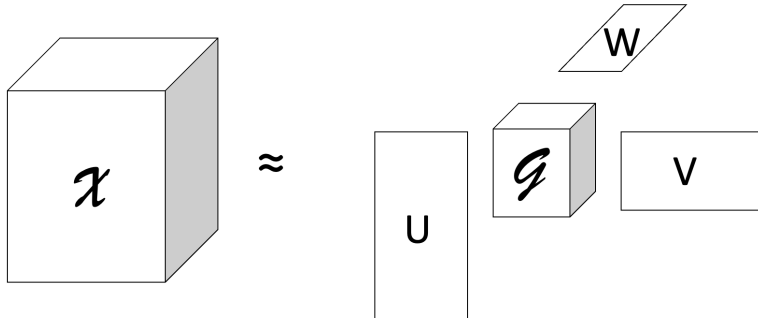


FIGURE 4.5 – Illustration de la décomposition en valeurs singulières d'ordre supérieure (HOSVD).

Contrairement à la SVD, le tenseur \mathcal{G} nommé tenseur corps, écrit ici sous forme de la matrice $\mathbf{G}_{(L)} \in \mathbb{R}^{L \times TS}$ n'est pas hyperdiagonal mais dense en général. Pour calculer cette décomposition, on utilise l'algorithme de [62]. L'algorithme de la HOSVD résout la décomposition de l'Équation 4.56) (présenté dans l'Algorithme 3 avec $\mathbf{U}^{(K)} = \text{left-}K\text{-SVD}(\mathbf{X})$ l'algorithme qui calcule les K premières composantes singulières gauches).

Algorithm 3 Estimer \mathbf{U} , \mathbf{V} et \mathbf{W} via HOSVD.

Require: $\mathcal{X} \in \mathbb{R}^{L \times T \times S}$

$$\mathbf{U} = \text{left_}L\text{-SVD}(\mathbf{X}_{(L)})$$

$$\mathbf{V} = \text{left_}T\text{-SVD}(\mathbf{X}_{(T)})$$

$$\mathbf{W} = \text{left_}S\text{-SVD}(\mathbf{X}_{(S)})$$

Afin de trouver des sous-espaces de rangs plus faibles, on cherche les matrices $\mathbf{U}^{(K_L)} \in \mathbb{R}^{L \times K_L}$, $\mathbf{V}^{(K_T)} \in \mathbb{R}^{T \times K_T}$ et $\mathbf{W}^{(K_S)} \in \mathbb{R}^{S \times K_S}$ de manière à ce qu'elles soient solutions du problème d'optimisation suivant :

$$\begin{aligned} & \underset{\mathbf{U}^{(K_L)}, \mathbf{V}^{(K_T)}, \mathbf{W}^{(K_S)}}{\operatorname{argmin}} \quad \|\mathbf{X}_{(L)} - \mathbf{U}^{(K_L)} \mathbf{G}_{(L)} (\mathbf{W}^{(K_S)} \otimes \mathbf{V}^{(K_T)})^t\|_F^2, \\ & \text{t.q.} \quad \mathbf{U}^{(K_L)t} \mathbf{U}^{(K_L)} = \mathbf{I}_{K_L}, \quad \mathbf{V}^{(K_T)t} \mathbf{V}^{(K_T)} = \mathbf{I}_{K_T}, \quad \mathbf{W}^{(K_S)t} \mathbf{W}^{(K_S)} = \mathbf{I}_{K_S}. \end{aligned} \quad (4.57)$$

Cette approximation est illustrée dans la Figure 4.5. Notons que $\mathbf{G}_{(L)} \in \mathbb{R}^{K_L \times K_T \times K_S}$ reste dense. Contrairement à la SVD, les matrices de bas rang recherchées ne résultent pas simplement de la troncature de la solution \mathbf{U} (resp. \mathbf{V} , \mathbf{W}) en ses K_L (resp. K_T , K_S) premières colonnes (même si cette procédure produit souvent une approximation satisfaisante). À la place, nous devons recourir à un algorithme d'optimisation itératif comme le "Higher Order Orthogonal Iteration of Tensors" (HOOI). Cette procédure, proposée dans [207, 63] que l'on décrit dans l'Algorithme 4, correspond à une alternance des moindres carrés. Les travaux de [203] montrent que dans la plupart des cas, HOOI atteint la solution optimale de l'Équation (4.57), avec de bonnes performances de convergence.

Algorithm 4 Estime $\mathbf{U}^{(K_L)}$, $\mathbf{V}^{(K_T)}$ et $\mathbf{W}^{(K_S)}$ via HOOI.

Require: $\mathcal{X} \in \mathbb{R}^{L \times T \times S}$, les paramètres de réduction (K_L, K_T, K_S) , l'incrément de tolérance $\epsilon > 0$,

i_{max} .

$i = 0$

$$[\mathbf{U}_0, \mathbf{V}_0, \mathbf{W}_0] = \text{HOSVD}(\mathcal{X})$$

while $i < i_{max}$ or $\|\mathbf{G}_{(L)i}\|_F^2 - \|\mathbf{G}_{(L)i-1}\|_F^2 > \epsilon$ **do**

$$\mathbf{U}_{i+1}^{(K_L)} = \text{left_}K_L\text{-SVD}(\mathbf{X}_{(L)}(\mathbf{W}_i^{(K_S)} \otimes \mathbf{V}_i^{(K_T)}))$$

$$\mathbf{V}_{i+1}^{(K_T)} = \text{left_}K_T\text{-SVD}(\mathbf{X}_{(T)}(\mathbf{U}_{i+1}^{(K_L)} \otimes \mathbf{W}_i^{(K_S)}))$$

$$\mathbf{W}_{i+1}^{(K_S)} = \text{left_}K_S\text{-SVD}(\mathbf{X}_{(S)}(\mathbf{V}_{i+1}^{(K_T)} \otimes \mathbf{U}_{i+1}^{(K_L)}))$$

$$\mathbf{G}_{(L)i+1} = \mathbf{U}_{i+1}^{(K_L)t} \mathbf{X}_{(L)} (\mathbf{W}_{i+1}^{(K_S)} \otimes \mathbf{V}_{i+1}^{(K_T)})$$

$i = i+1$

end while

4.3.3 Parafac2

Lorsque l'on considère un tenseur de 3-modes, le nombre de colonnes de chacune de ses tranches doit rester constant. Or, dans le contexte de l'épilepsie, l'hétérogénéité temporelle des différentes crises d'un même patient est courante. Cela se traduit par des crises associées à un nombre différent d'échantillons temporels, et donc il est impossible de représenter toutes les crises d'un patient grâce au tenseur \mathcal{X} . Dans ce cas on ne parle plus de tenseurs, et on note par $\mathbf{X}\{s\} \in \mathbb{R}^{L \times T(s)}$ la $s^{\text{ième}}$ crise d'un patient. La décomposition Parafac2 [117] est une extension de la CP, lorsque cette hétérogénéité sur la dimension d'un mode existe. Elle consiste à trouver une matrice de facteurs de CFs $\mathbf{U} \in \mathbb{R}^{L \times K}$ et S profils d'activation (un pour chaque crise), notés $\mathbf{V}\{s\} \in \mathbb{R}^{T(s) \times K}$, $\forall s \in \{1, \dots, S\}$, ainsi qu'une matrice $\mathbf{W} \in \mathbb{R}^{S \times K}$ qui minimisent :

$$\begin{aligned} & \underset{\mathbf{U}, \mathbf{V}\{1\}, \dots, \mathbf{V}\{S\}, \mathbf{W}}{\operatorname{argmin}} \quad \sum_{s=1}^S \|\mathbf{X}\{s\} - \mathbf{U} \mathbf{D}(\mathbf{w}_s) \mathbf{V}\{s\}^t\|_F^2, \\ & \text{t.q.} \quad \mathbf{V}\{s\}^t \mathbf{V}\{s\} = \mathbf{C} \quad \forall s \in \{1, \dots, S\}. \end{aligned} \quad (4.58)$$

avec $D(\mathbf{w})$ la matrice diagonale dont les éléments sont le vecteur \mathbf{w} et \mathbf{C} une matrice de covariance. La décomposition est illustrée dans la Figure 4.6.

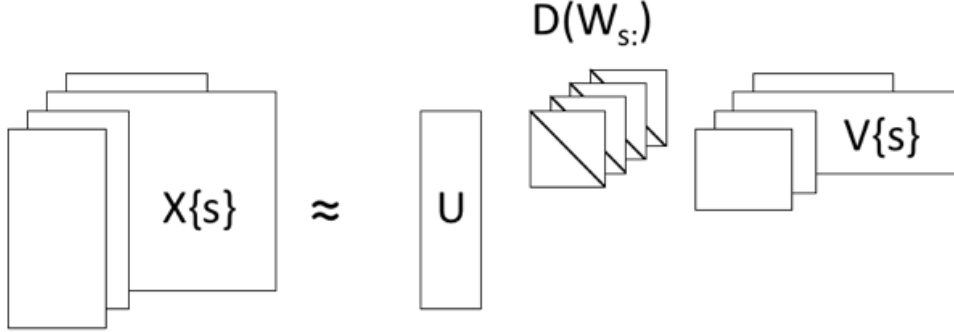


FIGURE 4.6 – Illustration de Parafac2.

La décomposition Parafac2 implique de contraindre chaque matrice $\mathbf{V}\{s\}$ à avoir la même covariance. Cette contrainte est importante pour satisfaire l'unicité de la solution [117], en effet, on a :

$$\mathbf{U}D(\mathbf{w}_{s:})\mathbf{V}\{s\}^t = \mathbf{U}\mathbf{A}D(\mathbf{b}_{s:})D(\mathbf{b}_{s:})^{-1}\mathbf{A}^{-1}D(\mathbf{w}_{s:})\mathbf{V}\{s\}^t = \mathbf{U}'D(\mathbf{w}'_{s:})\mathbf{V}'\{s\}^t, \quad (4.59)$$

avec $\mathbf{A} \in \mathbb{R}^{K \times K}$ et $\mathbf{B} \in \mathbb{R}^{S \times K}$, des matrices inversibles quelconques, et $\mathbf{U}' = \mathbf{U}\mathbf{A}$, $\mathbf{W}' = \mathbf{B}$ et $\mathbf{V}'\{s\} = \mathbf{V}\{s\}D(\mathbf{w}_{s:})\mathbf{A}^{-1}D(\mathbf{b}_{s:})^{-1}$, $\forall s \in \{1, \dots, S\}$. Alors, si \mathbf{U} , \mathbf{W} et $\mathbf{V}\{s\}$ sont solutions du problème de l'Équation (4.58), les matrices \mathbf{U}' , \mathbf{W}' et $\mathbf{V}'\{s\}$ le sont aussi.

De plus, cette contrainte permet une implémentation simple et élégante de l'algorithme minimisant l'Équation (4.58). En effet, afin de garantir le respect de la contrainte de Parafac2 (Équation (4.58)), on pose $\mathbf{V}\{s\} = \mathbf{Q}\mathbf{P}\{s\}$, avec $\mathbf{P}\{s\}$ une matrice orthonormale par colonne et \mathbf{Q} une matrice telle que $\mathbf{Q}^t\mathbf{Q} = \mathbf{C}$ (dans ce cas on a forcément $\mathbf{V}\{s\}\mathbf{V}\{s\}^t = \mathbf{C} \forall s \in \{1, \dots, S\}$). On peut déduire les matrices $\mathbf{P}\{s\}$ après avoir fixé les matrices \mathbf{Q} , \mathbf{W} et \mathbf{U} en minimisant :

$$\begin{aligned} & \underset{\mathbf{P}\{1\}, \dots, \mathbf{P}\{S\}}{\operatorname{argmin}} \sum_{s=1}^S \|\mathbf{X}(s) - \mathbf{U}D(\mathbf{w}_{s:})\mathbf{Q}^t\mathbf{P}\{s\}^t\|_F^2, \\ & \text{t.q.} \quad \mathbf{P}\{s\}^t\mathbf{P}\{s\} = \mathbf{I} \quad \forall s \in \{1, \dots, S\}. \end{aligned} \quad (4.60)$$

Il existe une solution à ce problème en posant $\tilde{\mathbf{U}}\{s\}\tilde{\mathbf{\Sigma}}\{s\}\tilde{\mathbf{V}}\{s\}$ la SVD de $\mathbf{X}\{s\}^t\mathbf{U}D(\mathbf{w}_{s:})\mathbf{Q}^t$ (Équation (4.14)). Alors $\mathbf{P}\{s\} = \tilde{\mathbf{U}}\{s\}\tilde{\mathbf{V}}\{s\}^t$. Enfin, pour trouver les matrices \mathbf{Q} , \mathbf{W} et \mathbf{U} en fixant les matrices $\mathbf{P}\{s\} \forall s \in \{1, \dots, S\}$, il faut minimiser :

$$\underset{\mathbf{U}, \mathbf{W}, \mathbf{C}}{\operatorname{argmin}} \sum_{s=1}^S \|\mathbf{X}\{s\}\mathbf{P}\{s\}^t - \mathbf{U}D(\mathbf{w}_{s:})\mathbf{Q}\|_F^2. \quad (4.61)$$

En posant maintenant $\tilde{\mathbf{X}}_{:,s} = \mathbf{X}\{s\}\mathbf{P}\{s\}^t$, on remarque que $\tilde{\mathbf{X}} \in \mathbb{R}^{L \times K \times S}$ a une structure de tenseur, et que l'Équation (4.61) correspond à une décomposition CP que l'on peut réécrire :

$$\underset{\mathbf{U}, \mathbf{W}, \mathbf{C}}{\operatorname{argmin}} \sum_{s=1}^S \|\tilde{\mathbf{X}}_{(L)} - \mathbf{U}(\mathbf{W} \odot \mathbf{Q})^t\|_F^2. \quad (4.62)$$

Ainsi, Parafac2, présenté dans l'Algorithme 5, consiste à alterner le calcul d'une SVD et d'une décomposition CP.

Algorithm 5 Estimation de \mathbf{U} , \mathbf{W} et $\mathbf{V}\{s\}$ via Parafac2.

Require: $\mathcal{X} \in \mathbb{R}^{L \times T \times S}$, le paramètre de réduction K , l'incrément de tolérance $\epsilon > 0$, i_{max} , une première supposition pour $\mathbf{U}_0, \mathbf{Q}_0, \mathbf{W}_0$

$i = 0$

while $i < i_{max}$ or $\sum_{s=1}^S \|\mathbf{U}_i \mathbf{D}(\mathbf{w}_{s:})_i \mathbf{V}\{s\}_i\| - \sum_{s=1}^S \|\mathbf{U}_{i+1} \mathbf{D}(\mathbf{w}_{s:})_{i+1} \mathbf{V}\{s\}_{i+1}\|$ **do**

 Calculer $\tilde{\mathbf{U}}\{s\} \tilde{\mathbf{\Sigma}}\{s\} \tilde{\mathbf{V}}\{s\}$ la SVD de $\mathbf{X}\{s\}_i^t \mathbf{U}_i \mathbf{D}(\mathbf{w}_{s:})_i \mathbf{Q}_i^t$

$\mathbf{P}\{s\}_{i+1} = \tilde{\mathbf{U}}(s) \tilde{\mathbf{V}}\{s\}^t$

$\tilde{\mathbf{X}}_{::s} = \mathbf{X}\{s\} \mathbf{P}\{s\}_{i+1}^t$

 Calculer $\mathbf{U}_{i+1}, \mathbf{Q}_{i+1}, \mathbf{W}_{i+1}$ via décomposition CP de $\tilde{\mathcal{X}}$ avec l'Algorithme 2

$\mathbf{V}\{s\}_{i+1} = \mathbf{Q}_{i+1} \mathbf{P}\{s\}_{i+1}$

$i = i+1$

end while

Notons que cette procédure a comme désavantage de difficilement supporter l'ajout de contraintes, notamment sur les profils d'activation, car le découpage $\mathbf{V}\{s\} = \mathbf{Q}\mathbf{P}\{s\}$ avec $\mathbf{P}\{s\}$ orthonormal par colonne devient souvent impossible.

4.4 Conclusion

Ce chapitre présente des décompositions de matrices et de tenseurs qui sont à la base des méthodes utilisées dans la suite de cette thèse. Notons que dans le chapitre 5, nous nous intéresserons au cas où chaque crise d'un même patient est écrite avec le même nombre d'échantillons temporels, permettant de former un jeu de données présentable sous forme d'un tenseur. Au contraire, le chapitre 6 s'intéresse à des crises plus complexes où les durées de chaque crise sont très variables. Dans ce cas, elles ne peuvent plus être exprimées sous la forme d'un tenseur.

Chapitre 5

Partitionnement de réseaux dynamiques via une décomposition tensorielle parcimonieuse

On propose dans cette section, une méthode afin de partitionner un graphe dynamique de CFs en un ensemble de sous-graphes contenant une information neuro-physiologique pertinente. On rappelle que le graphe dynamique obtenu est une représentation imparfaite de la dynamique corticale du cerveau durant une crise d'épilepsie. Comme les algorithmes de partitionnement sont connus pour être sensibles aux bruits, on propose de considérer plusieurs crises d'un même patient. En effet, un des atouts des enregistrements sEEG est la possibilité d'enregistrer plusieurs crises. Comme il est généralement admis que la dynamique des CFs est comparable d'une crise à l'autre pour un même patient, l'analyse conjointe de ces enregistrements devrait faciliter l'identification de facteurs de CFs communs à toutes les crises et caractéristiques du problème de santé du patient. Malgré la prise en compte de plusieurs crises, nous réalisons toujours un partitionnement sur des données bruitées et de grandes dimensions. L'algorithme de partitionnement utilisé est k-means, bien que les limites de ses performances soient connues dans ces conditions [103].

Pour faire face à ces limites, on développe une méthode de réduction dimensionnelle qui extrait les caractéristiques pertinentes des CFs communes à toutes les crises sous forme de facteurs caractéristiques. Puisque les données correspondent à des CFs évoluant dans le temps et pour différentes crises, elles peuvent être commodément représentées comme un tenseur de 3-modes. La réduction dimensionnelle proposée est une décomposition tensorielle originale et spécifique au problème, qui impose des contraintes structurelles à la solution, comme la parcimonie par exemple. La méthode que nous développons ici peut alors être interprétée comme une décomposition de graphes dynamiques en signature structurelle (sous-graphe de CFs) et temporelle (profils d'activation temporelles). Cette procédure pourrait être utilisée dans d'autres contextes impliquant des graphes dynamiques ou multiplexes comme l'étude des interactions sociales [79, 92] ou les réseaux de transport [93]. Notons que considérer les graphes dynamiques (ou multiplexes) comme des tenseurs n'est pas nouveau, des travaux ont déjà exploré cette idée, par exemple en échantillonnage [157] pour l'inférence de graphes dynamiques [174] ou pour la détection de communautés [173].

La contribution de ce chapitre est de développer un nouvel algorithme de réduction de la dimension applicable à des données tensorielles. Des contraintes structurelles sont imposées pour obtenir une solution pertinente. L'algorithme proposé répond aussi à la nécessité d'extraire des facteurs de CFs communes à toutes les crises. Sur cette base, nous proposons une chaîne de traitement permettant d'inférer et d'analyser des graphes dynamiques provenant d'enregistrements sEEG. Pour évaluer la capacité de notre méthode à partitionner le réseau dynamique, on la compare à d'autres approches de réduction de dimension (y compris d'autres décompositions de tenseurs). Enfin, cette décomposition de graphes dynamiques est appliquée sur des enregistrements sEEG réels, dans l'objectif d'extraire les caractéristiques principales des CFs communes à l'ensemble des crises d'un patient.

L'organisation de ce chapitre est la suivante : la Section 5.1 justifie l'utilisation d'une décomposition tensorielle spécifique pour faire du partitionnement de graphe dynamique. Les Sections 5.2 et 5.3 détaillent l'implémentation de la décomposition tensorielle, respectivement sans et avec contraintes structurelles. La Section 5.4 souligne les performances et les limites de la méthode proposée. Enfin en Section 5.5, nous appliquons la méthode sur des données réelles montrant sa capacité à inférer un graphe dynamique caractérisant l'évolution de la CF lors d'une crise.

5.1 Partitionner un réseau dynamique

5.1.1 Le réseau dynamique de CFs

Comme présenté en Section 1.2.1, le calcul des CFs dans le temps forme un réseau dynamique, réalisation bruitée de l'état cortical du cerveau durant une crise d'épilepsie. Ce réseau est noté \mathcal{G}_t , $t \in \{1, \dots, T\}$ avec T le nombre de mesures temporelles de CFs. Au lieu de travailler sur ce graphe dynamique, les partitions que nous recherchons sont en fait extraites de son *line-graph*. On rappelle qu'un graphe-ligne (ou graphe adjoint) noté $\mathcal{L}\mathcal{G}_t$, correspond au graphe dont les nœuds sont les liens de \mathcal{G}_t et de plus, des liens existent si les deux arêtes associées au graphe \mathcal{G}_t partagent un nœud commun. Le signal sur graphe correspond maintenant aux éléments diagonaux supérieurs de la matrice d'adjacence de \mathcal{G}_t , contenant les L valeurs de CFs formant une série temporelle L -dimensionnelle de T échantillons. C'est une manière de motiver la construction de la matrice de CFs $\mathbf{X} \in \mathbb{R}^{L \times T}$. On justifie l'utilisation du graphe-ligne, car elle permet de représenter les mesures de CFs sous forme de matrice de CFs. Le partitionnement de cette matrice permet alors de regrouper les CFs qui s'activent de manière synchrone dans le temps. En effet, cela est différent du partitionnement de nœuds, qui est une tâche plus habituelle des méthodes pour traiter un graphe \mathcal{G}_t (que ce soit pour la détection de communautés dans des réseaux complexes [71, 70], ou l'inférence de groupes dans les modèles stochastiques par blocs de graphe [3, 35, 111]). Cela favorise l'estimation de sous-graphes complets, ce qui n'est pas souhaitable pour notre application. L'utilisation du graphe-ligne $\mathcal{L}\mathcal{G}_t$ implique donc que la partition de la matrice de CFs \mathbf{X} correspond directement à une décomposition des arêtes du réseau dynamique \mathcal{G}_t . Idéalement, la recherche d'une partitions en N ensembles de CFs à partir de la matrice de CFs \mathbf{X} devrait produire $N - 1$ graphes contenant les liens qui s'activent de manière synchrone dans le temps, ainsi qu'un graphe restant rassemblant les CFs non synchronisées.

On illustre cette décomposition sur un exemple simple représenté dans la Figure 5.1. À partir des signaux affichés dans la Figure 5.1.(a), nous estimons la matrice de CFs \mathbf{X} montré dans la Figure 5.1.(b). L'utilisation de k-means sur les lignes de \mathbf{X} sert à identifier $N = 3$ partitions. Les trois sous-graphes résultants sont représentés dans la Figure 5.1.(c), où les deux premiers correspondent au modèle d'activation et le troisième contient les CFs non synchronisées.

Cependant, comme S différentes crises d'un même patient sont observées, les matrice de CFs peuvent être empilées dans un tenseur de 3-modes $\mathcal{X} \in \mathbb{R}^{L \times T \times S}$. Tout cela demande de développer une méthode pour effectuer cette décomposition du réseau dynamique sur la structure du tenseur, dont l'objectif est d'obtenir des sous-graphes de CFs pertinents qui s'activent de manière synchrone à travers les différentes crises.

5.1.2 Partitionner un tenseur avec k-means

Il existe de nombreuses approches pour faire du partitionnement de CFs :

- la détection de communautés est utilisée dans [115] pour discriminer des CFs caractéristiques d'états temporels d'une crise d'épilepsie,
- le partitionnement spectral [95] permet de combiner des électrodes formant des graphes complets,
- la décomposition en matrices non négatives [85] permet de réaliser un partitionnement doux des CFs,

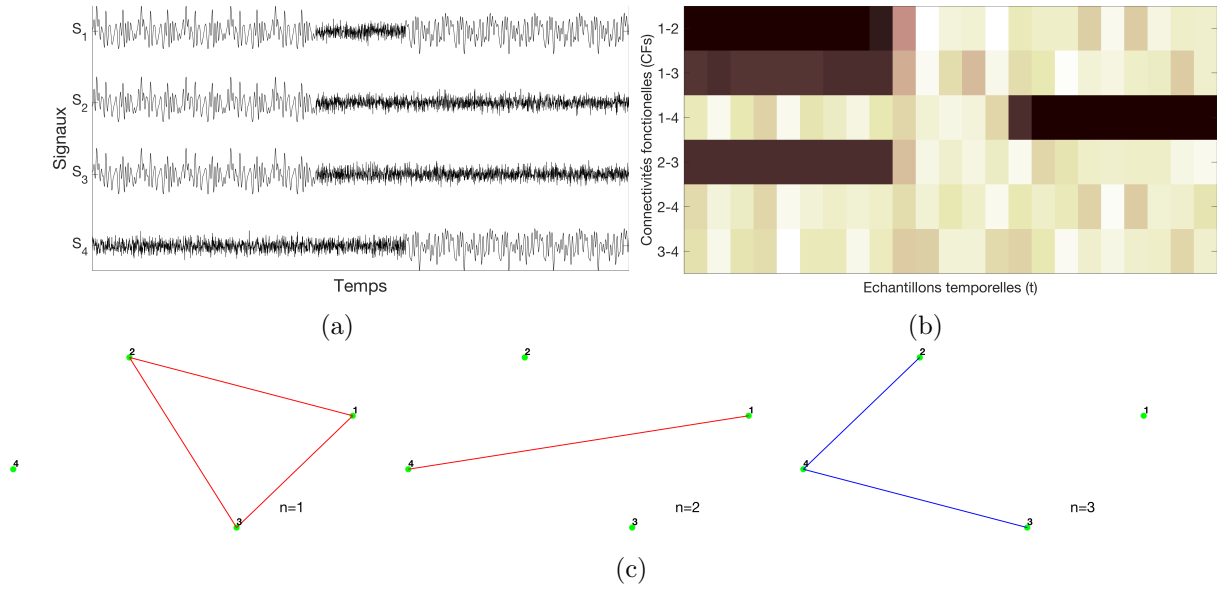


FIGURE 5.1 – (a) 4 signaux sEEG simulés, (b) Matrice de CFs $\mathbf{X} \in \mathbb{R}^{L \times T}$ entre les 6 paires de signaux calculées par PLV (voir la Définition 9), (c) Sous-graphe obtenu après application de k-means sur la matrice de CFs avec $N = 3$ partitions. Sur la représentation, les nœuds correspondent aux électrodes 1 à 4, et les liens correspondent à la solution du partitionnement. Les deux graphes rouges sont associés à des liens qui s’activent de manière synchrone dans le temps, le graphe bleu contient les liens restants et non synchronisés.

— enfin l’algorithme k-means a été souvent utilisé pour partitionner des CFs ou des états temporels [85, 135, 80].

On utilise ici comme méthode de partitionnement l’algorithme k-means [103] car il est simple à utiliser et reste performant face à d’autres méthodes plus récentes [175] comme dbscan [68] ou le partitionnement spectral [72]. De plus, comme discuté plus tard, ses propriétés et limites ont été étudiées en détail.

Une extension directe de k-means pour le tenseur \mathcal{X} revient à chercher P matrices $\Theta_{p::}$, qui sont les centroïdes des P partitions qui caractérisent le mieux les données. La tranche l , notée $\mathbf{X}_{l::}$, appartient à la partition p (notée $l \in C_p$) si son centroïde le plus proche est $\Theta_{p::}$. En s’inspirant de la formulation usuelle de k-means pour les matrices (Équation (4.18)), on cherche la partition $C_p \forall p \in \{1, \dots, P\}$ et les centroïdes minimisant :

$$\operatorname{argmin}_{C_p, \Theta_{1::}, \Theta_{2::}, \dots, \Theta_{P::}} \sum_{p=1}^P \sum_{l \in C_p} \|\mathbf{X}_{l::} - \Theta_{p::}\|_F^2. \quad (5.1)$$

En utilisant la vectorisation (voir en Section 4.2.4), on peut proposer une réécriture du critère de l’Équation (5.1) :

$$\operatorname{argmin}_{C_p, \Theta_{1::}, \Theta_{2::}, \dots, \Theta_{P::}} \sum_{p=1}^P \sum_{l \in C_p} \|\operatorname{vec}(\mathbf{X}_{l::}) - \operatorname{vec}(\Theta_{p::})\|_F^2. \quad (5.2)$$

En remarquant que le matricage de \mathcal{X} sur le premier mode correspond à $\mathbf{X}_{(L)l:} = \operatorname{vec}(\mathbf{X}_{l::})$, on montre que ce problème est équivalent à appliquer k-means sur les lignes de $\mathbf{X}_{(L)}$, soit la concaténation de toutes les matrices de CFs. Il s’agit alors d’une manière intuitive de traiter le mode des crises [44]. On note pour simplifier $\theta_p = \operatorname{vec}(\Theta_{p::})$, avec $\theta_p \in \mathbb{R}^{TS}$ les centroïdes qui sont maintenant des vecteurs. Alors, la partition du premier mode du tenseur avec k-means revient simplement à résoudre le problème d’optimisation suivant :

$$\operatorname{argmin}_{C_p, \theta_{1:}, \theta_{2:}, \dots, \theta_{P:}} \sum_{p=1}^P \sum_{l \in C_p} \|\mathbf{X}_{(L)l:} - \theta_{p:}\|_F^2. \quad (5.3)$$

Cependant, comme $\mathbf{X}_{(L)}$ est dans un espace de grande dimension ($TS \gg 0$), et comme k-means est connu pour être un problème non convexe et NP-hard, ses performances peuvent se dégrader

considérablement sur des matrices de grandes tailles [103]. De plus, les mesures de CFs sont souvent bruitées. Tout cela induit l'existence de minima locaux qui ne correspondent pas nécessairement aux solutions de partitions pertinentes, et le nombre de répétitions avec différentes conditions initiales nécessaires pour garantir un minimum local satisfaisant, compromet l'utilisation de la formulation de l'Équation (5.3) pour traiter notre application.

5.1.3 La réduction de dimension

Comme montré en Section 4.1.3, k-means peut être modifié afin de correspondre à un problème de factorisation matricielle. Plus précisément, on peut considérer le problème d'optimisation associé à k-means comme trouver une matrice \mathbf{A} qui maximise (voir Section 4.1.3) :

$$\begin{aligned} \operatorname{argmax}_{\mathbf{A}} \quad & \|\mathbf{A}^t \mathbf{X}_{(L)}\|_F^2 \quad n \\ \text{t.q.} \quad & \mathbf{a}_{\cdot p} \text{ vecteur de base canonique.} \end{aligned} \quad (5.4)$$

Afin de contourner la "malédiction de la dimensionnalité", nous proposons d'appliquer k-means à une matrice de K facteurs $\mathbf{F} \in \mathbb{R}^{L \times K}$ contenant les facteurs de CFs pertinents de $\mathbf{X}_{(L)}$. Le problème de l'Équation (5.4), après avoir réduit la dimension de $\mathbf{X}_{(L)}$, devient alors :

$$\begin{aligned} \operatorname{argmax}_{\mathbf{A}} \quad & \|\mathbf{A}^t \mathbf{F}\|_F^2, \\ \text{t.q.} \quad & \mathbf{a}_{\cdot p} \text{ vecteur de base canonique.} \end{aligned} \quad (5.5)$$

Bien que les solutions optimales des Équations (5.4) et (5.5) devraient partager la même structure, il est à noter qu'en général elles ne sont pas égales. En effet, les facteurs de CFs \mathbf{F} doivent contenir uniquement les motifs de CFs les plus caractéristiques du jeu de données, et guider l'algorithme pour obtenir une partition plus robuste. De plus, si la réduction dimensionnelle impose à la matrice \mathbf{F} de partager des contraintes similaires à la matrice \mathbf{A} (qui est une matrice non négative, parcimonieuse et orthogonale), alors l'application de k-means devrait être facilitée [64].

La réduction de dimension la plus utilisée avant d'appliquer k-means est la SVD [154, 33, 195]. Le tandem SVD avec k-means a déjà été employé pour des applications réseaux [154, 96], en imagerie médicale [163] pour l'analyse de CFs [85], et étudié en profondeur pour les données d'expression des gènes [33]. En respectant les notations de la Section 4.1.3, utiliser la SVD comme méthode de réduction dimensionnelle conjointement à k-means correspond à résoudre l'Équation (5.5) quand $\mathbf{F}_{SVD} = \mathbf{U}^{(K)} \mathbf{\Lambda}^{(K)}$. Des résultats théoriques récents montrent que sous certaines conditions, \mathbf{F}_{SVD} est une bonne approximation de $\mathbf{X}_{(L)}$ afin d'obtenir la meilleure partition solution de l'Équation (5.4) [69]. Cependant, cette approche matricielle empêche l'utilisation de contraintes structurelles sur le mode temporel et le mode des crises séparément (puisqu'elles sont concaténées en un seul mode). Une amélioration importante consiste alors à considérer la décomposition de chaque mode du tenseur \mathcal{X} pour estimer une matrice de facteurs \mathbf{F} .

On peut proposer l'utilisation de la décomposition CP, qui, en utilisant les notations de la Section 4.3.1, permet de réduire le tenseur \mathcal{X} en une matrice $\mathbf{F}_{nnCP} = \mathbf{U}$. Cependant, cette décomposition n'est pas la plus pertinente lorsque l'on souhaite des facteurs communs à toutes les crises car elle peut produire des composantes spécifiques à chaque crise. Il existe alors une autre option, l'utilisation de la HOSVD appliquée conjointement à k-means. En utilisant les notations de la Section 4.3.2, cela correspond à résoudre l'Équation (5.5) quand $\mathbf{F}_{HOOI} = \mathbf{U}^{(K_L)} \mathbf{G}_{(L)}$. En fixant $K_S = 1$, on peut ainsi obtenir les facteurs de CFs les plus actifs sur l'ensemble des crises. Malgré tout, cette décomposition souffre d'un manque de parcimonie, offrant des résultats difficiles à analyser car elle ne distingue pas les CFs les plus critiques.

5.1.4 Décomposition tensorielle proposée

Notre objectif est de proposer une décomposition capable d'isoler les facteurs dynamiques de CFs qui apparaissent dans toutes les crises d'un patient. Ainsi on obtient, par application de la HOSVD

avec $K_S = 1$, et $K_L = K_T = K$ (voir la Section 4.3.2), l'approximation suivante :

$$\mathbf{X}_{(L)} \approx \mathbf{F}(\mathbf{w} \otimes \mathbf{V})^t. \quad (5.6)$$

Le vecteur $\mathbf{w} \in \mathbb{R}^{S \times 1}$ mesure le taux de contribution de chaque crise dans l'inférence des facteurs de CFs et des profils d'activation obtenus. Par exemple, un vecteur constant ($\mathbf{w}_s = \frac{1}{\sqrt{S}} \forall s \in 1, \dots, S$) signifie que la simple moyenne des matrices de CFs de chaque crise, permet d'obtenir les facteurs de CFs les plus pertinents. Au contraire si les crises ne partagent aucune caractéristique commune, le vecteur \mathbf{w} identifie la crise s telle que $\|\mathbf{X}_{::s}\|_F^2$ soit maximal. Comme la matrice de CFs générique est une combinaison linéaire de toutes les crises, il est bon de regarder de près le vecteur $\mathbf{w} \in \mathbb{R}^S$ afin d'observer l'hétérogénéité des crises.

Le choix $K_L = K_T = K$ est considéré dans un souci de simplicité afin que les matrices $\mathbf{F} \in \mathbb{R}^{L \times K}$ et $\mathbf{V} \in \mathbb{R}^{T \times K}$ aient le même nombre de facteurs, cela permet de faciliter la résolution de l'Équation (5.6). En effet, différencier ces deux paramètres rendrait le modèle plus complexe sans permettre d'obtenir des résultats forcément meilleurs. On propose de contraindre à la fois les facteurs de CFs et les profils d'activation temporelle à être parcimonieux. L'importance de la parcimonie sur les profils de CFs est justifiée en Section 4.1.2 où il s'agit de la régularisation de la Définition 13. De même, l'importance de la parcimonie sur les profils d'activation y est aussi justifiée (correspondant à la régularisation de la Définition 13). Ici, cette contrainte est encore plus importante car, suite au moyennage pondéré réalisé par le vecteur \mathbf{w} , elle permet d'éliminer les périodes temporelles où il n'y a pas d'activation commune des facteurs de CFs. Au total, la décomposition que nous proposons est la suivante.

Proposition 11

Soit \mathcal{X} un tenseur contenant les matrices de CFs de plusieurs crises d'un même patient, alors la décomposition tensorielle permettant d'obtenir des facteurs de CFs \mathbf{F} communs à chaque crise correspond au problème d'optimisation suivant :

$$\begin{aligned} \underset{\mathbf{F}, \mathbf{V}, \mathbf{w}}{\operatorname{argmin}} \quad & \|\mathbf{X}_{(L)} - \mathbf{F}(\mathbf{w} \otimes \mathbf{V})^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1, \\ \text{t.q.} \quad & \lambda_2 \|\mathbf{v}_{:k}\|_1 + \|\mathbf{v}_{:k}\|_F^2 \leq 1, \\ \text{et} \quad & \mathbf{w}^t \mathbf{w} = 1. \end{aligned} \quad (5.7)$$

où les hyperparamètres λ_1 et λ_2 permettent de régler le compromis entre la précision et la parcimonie de l'approximation. On appelle cette décomposition fournie dans l'Équation (5.7) décomposition modale parcimonieuse d'ordre supérieur (qui sera dénotée par le sigle HOSMR pour High Order Sparse Mode Decomposition).

L'importance de la contrainte elastic-Net est justifiée en Section 4.1.2. En effet, on a observé [76] des décompositions peu pertinentes si des contraintes de parcimonie seules sont appliquées à la fois à la matrice \mathbf{F} et \mathbf{V} , cette observation est corroborée par [15]. On préfère ici l'utilisation de la contrainte elastic-Net, plutôt que Fused-lasso, car le moyennage des profils d'activation, via le vecteur \mathbf{w} , détériore les localisations temporelles des différentes étapes d'une crise.

Après avoir réalisé l'approximation de l'Équation (5.7), \mathbf{F} est mis à l'échelle de telle sorte que $\|\mathbf{w}\|_F^2 = \|\mathbf{v}_{:k}\|_F^2 = 1, \forall k$ afin de transférer toute l'énergie aux facteurs de CFs. Une solution de l'Équation (5.7) produit une matrice de CFs de faible dimension $\mathbf{F}_{\text{HOSMR}}$ que l'on insère dans l'Équation (5.5) afin de réaliser le partitionnement avec k-means.

5.2 Une modification de HOOI pour résoudre le problème de la HOSMR sans contrainte

La HOSMR fournie dans l'Équation (5.7), peut être reliée à des décompositions tensorielles déjà existantes. La Section 5.2.1 propose une brève liste de ces décompositions avant d'explicitier le lien entre la HOSMR et la HOSVD (présenté Section 4.3.2). Ce lien est important car nous proposons pour résoudre le problème de la HOSMR de nous inspirer d'une modification de l'algorithme HOOI. En

effet, on montre en Section 5.2.2 qu’une modification de l’algorithme HOOI dans le cas particulier $K_L = K_T = K$ et $K_S = 1$ est possible. Une étape de cette modification fait intervenir un problème de factorisation matricielle. Nous verrons que des contraintes peuvent facilement être ajoutées à cette factorisation matricielle, donnant une méthode simple pour résoudre le problème de l’Équation (5.7). Finalement, on montre en Section 5.A que notre modification de l’algorithme HOOI, en plus de livrer une approche originale pour résoudre la fonction de coût de la HOSVD approximée par l’Équation (4.57), a une complexité moindre que l’algorithme HOOI, et converge plus rapidement que celui-ci sur des modèles de réseaux dynamique de CFs.

5.2.1 Lien entre la HOSMR et la HOSVD

On peut relier la fonction de coût de l’Équation (5.7) à des décompositions tensorielles déjà existantes. La décomposition proposée est proche d’une *Block term decomposition* [61] en une composante, et avec des contraintes supplémentaires de parcimonie appliquées aux deux premiers modes. Il s’agit aussi d’une variante parcimonieuse de la HOSVD [146, 207, 126], où la contrainte d’estimer des matrices unitaires est relaxée (on pourra parler de décomposition de Tucker dans ce cas). En effet, en notant par f la fonction de coût du problème de la HOSVD de l’Équation (4.57), quand $K_S = 1$, $K_L = K_T = K$, on obtient :

$$f(\mathbf{U}, \mathbf{V}, \mathbf{w}) = \|\mathbf{X}_{(L)} - \mathbf{U}\mathbf{G}_{(L)}(\mathbf{w} \otimes \mathbf{V})^t\|_F^2. \quad (5.8)$$

Afin d’estimer, pour chaque facteur de CFs, des profils d’activation temporelle distincts, on applique une régularisation de lasso sur la matrice de facteur $\mathbf{U}\mathbf{G}_{(L)}$, nous conduisant à la fonction de coût régularisée suivante :

$$f^{reg}(\mathbf{U}, \mathbf{V}, \mathbf{w}) = f(\mathbf{U}, \mathbf{V}, \mathbf{w}) + \lambda_1 \|\mathbf{U}\mathbf{G}_{(L)}\|_1. \quad (5.9)$$

Afin de simplifier la fonction de coût, et en gardant à l’esprit que notre intérêt réside dans la matrice de facteur, et non dans l’analyse du tenseur corps ou de la matrice \mathbf{U} , on pose $\mathbf{F} = \mathbf{U}\mathbf{G}_{(L)} \in \mathbb{R}^{L \times K}$. Ainsi, nous obtenons la fonction de coût de la HOSMR équation de l’Équation (5.7), qui s’écrit :

$$g(\mathbf{U}, \mathbf{V}, \mathbf{w}) = \|\mathbf{X}_{(L)} - \mathbf{F}(\mathbf{w} \otimes \mathbf{V})^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1. \quad (5.10)$$

Des implémentations pour réaliser des décompositions tensorielles parcimonieuses existent déjà. Par exemple, les auteurs de [146] proposent un algorithme pour réaliser une décomposition de Tucker parcimonieuse. Cependant, son implémentation permet seulement d’ajouter de simples régularisations de lasso sur les modes que l’on souhaite parcimonieux, ce qui, comme on l’a déjà mentionné, ne semble pas approprié lorsque la parcimonie est attendue dans plusieurs modes [15]. Les auteurs de [6] utilisent une méthode de puissance itérée, connue pour ne pas fournir une solution optimale [54], et de produire des composantes gloutonnes qui manquent d’interprétabilité. Finalement, comme proposé dans [207] et [6], l’orthogonalité et la parcimonie peuvent être imposées conjointement. Cependant, la solution gloutonne correspondante, comme les solutions d’une méthode de puissance itérée, sera difficile à interpréter. On justifie alors le besoin de proposer une implémentation originale du problème de la HOSMR de l’Équation (5.7). On profite ainsi du lien entre la HOSMR et la HOSVD, pour présenter une approche provenant d’une modification de l’algorithme HOOI.

5.2.2 Modification de l’algorithme HOOI

L’algorithme HOOI (Algorithme 3) de la Section 4.3.2, minimise la fonction de coût de l’Équation (5.8), avec \mathbf{U} , \mathbf{V} et \mathbf{w} contraints d’être des matrices semi-orthogonales et avec $\mathbf{G}_{(L)} = \mathbf{U}^t \mathbf{X}_{(L)} (\mathbf{w} \otimes \mathbf{V})$. A chaque itération i de cet algorithme, les matrices \mathbf{U}_i et \mathbf{V}_i sont calculées en extrayant les valeurs singulières de deux matrices différentes. En notant que la matrice de CFs générique, associée au vecteur \mathbf{w} est une matrice, l’idée derrière la modification est d’extraire les composantes singulières gauches et droites de cette matrice pour obtenir \mathbf{U}_i et \mathbf{V}_i . On souligne que cette astuce est limitée au cas $K_S = 1$. Quand ce n’est pas le cas ($K_S = 2$ par exemple), le tenseur \mathcal{X} est compressé en deux matrices de

CFs qui forment un tenseur dont les matriciations sur le mode- L et le mode- T correspondent à deux matrices de dimensions différentes. Cela rend alors impossible d'extraire \mathbf{U}_i et \mathbf{V}_i à partir de la même factorisation, ce qui permet de réduire deux des étapes de la HOOI à une seule étape. La modification de la HOOI proposée (appelée mHOOI) alterne les 2 étapes suivantes :

Étape (A) Factorisation de la matrice de CFs générique

À l'itération $(i + 1)$, on suppose \mathbf{w}_i connu. Ce vecteur permet de calculer une matrice de CFs générique, en appliquant le produit de contraction entre celui-ci et le tenseur \mathcal{X} . On peut alors noter en appliquant la formule de matricage de l'Équation (4.40) que :

$$(\mathcal{X} \bullet_3 \mathbf{w}_i)_{(L)} = \mathbf{X}_{(L)}(\mathbf{w}_i \otimes \mathbf{I}_T), \quad (5.11)$$

avec $\mathbf{X}_{(L)}(\mathbf{w}_i \otimes \mathbf{I}_T) \in \mathbb{R}^{L \times T}$ la matrice générique. On peut écrire une variante de la fonction de coût de l'Équation (4.57) appliquée au matricage du troisième mode du tenseur \mathcal{X} (les deux fonctions de coût sont équivalentes) :

$$f(\mathbf{U}, \mathbf{V}, \mathbf{w}_i) = \|\mathbf{X}_{(S)} - \mathbf{w}_i \mathbf{G}_{(S)}(\mathbf{V} \otimes \mathbf{U})^t\|_F^2, \quad (5.12)$$

On rappelle que $\mathbf{G}_{(S)} = \mathbf{w}_i^t \mathbf{X}_{(S)}(\mathbf{V} \otimes \mathbf{U})$, on a donc :

$$\begin{aligned} f(\mathbf{U}, \mathbf{V}, \mathbf{w}_i) &= -Tr(\mathbf{w}_i^t \mathbf{X}_{(S)}(\mathbf{V} \otimes \mathbf{U})(\mathbf{V} \otimes \mathbf{U})^t \mathbf{X}_{(S)}^t \mathbf{w}_i) + Ct, \\ &= -\|\mathbf{w}_i^t \mathbf{X}_{(S)}(\mathbf{V} \otimes \mathbf{U})\|_F^2 + Ct, \end{aligned} \quad (5.13)$$

Avec Ct une constante (ici égale à $Ct = Tr(\mathbf{X}_{(S)} \mathbf{X}_{(S)}^t)$). En remarquant que $\mathbf{w}_i^t \mathbf{X}_{(S)}$ correspond à la vectorisation de la matrice générique :

$$\mathbf{w}_i^t \mathbf{X}_{(S)} = \text{vec}(\mathbf{X}_L(\mathbf{w}_i \otimes \mathbf{I}_T))^t, \quad (5.14)$$

avec $\mathbf{I}_T \in \mathbb{R}^{T \times T}$ matrice identité, et en rappelant la formule de compatibilité entre la vectorisation et le produit de Kronecker de l'Équation (4.53), on obtient :

$$\begin{aligned} f(\mathbf{U}, \mathbf{V}, \mathbf{w}_i) &= -\|\text{vec}(\mathbf{X}_L(\mathbf{w}_i \otimes \mathbf{I}_T))^t(\mathbf{V} \otimes \mathbf{U})\|_F^2 + Ct, \\ &= -\|\text{vec}(\mathbf{U}^t \mathbf{X}_L(\mathbf{w}_i \otimes \mathbf{I}_T)^t \mathbf{V})\|_F^2 + Ct, \\ &= -\|\mathbf{U}^t \mathbf{X}_L(\mathbf{w}_i \otimes \mathbf{I}_T)^t \mathbf{V}\|_F^2 + Ct. \end{aligned} \quad (5.15)$$

Alors, la minimisation de la fonction f selon les variables \mathbf{U} et \mathbf{V} sous contrainte de semi-orthogonalité correspond à :

$$\begin{aligned} \underset{\mathbf{U}, \mathbf{V}}{\text{argmax}} \quad &\|\mathbf{U}^t \mathbf{X}_{(L)}(\mathbf{w}_i \otimes \mathbf{I}_T) \mathbf{V}\|_F^2, \\ \text{t.q.} \quad &\mathbf{U}^t \mathbf{U} = \mathbf{I}_L, \quad \mathbf{V}^t \mathbf{V} = \mathbf{I}_T. \end{aligned} \quad (5.16)$$

La solution optimale du problème de l'Équation (5.16) est donnée par les approximations de rang faible \mathbf{U}_{i+1} et \mathbf{V}_{i+1} , obtenues comme les K premières composantes de la SVD de la matrice de CFs générique :

$$\mathbf{X}_{(L)}(\mathbf{w}_i \otimes \mathbf{I}_T) = \mathbf{U}_{i+1} \mathbf{\Lambda}_{i+1} \mathbf{V}_{i+1}^t. \quad (5.17)$$

Précisons que dans ce cas, comme il s'agit de la solution optimale, la fonction de coût f de l'Équation (5.8) diminue au fil des itérations :

$$f(\mathbf{U}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_i) \leq f(\mathbf{U}_i, \mathbf{V}_i, \mathbf{w}_i). \quad (5.18)$$

Étape (B) Mise à jour de la contribution de chaque crise

Afin de mettre à jour le vecteur \mathbf{w}_{i+1} , connaissant les matrices \mathbf{U}_{i+1} , et \mathbf{V}_{i+1} , on utilise l'écriture de la fonction de coût de l'Équation (5.13), qui se note dans ce cas :

$$f(\mathbf{U}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_i) = - \|\mathbf{w}^t \mathbf{X}_{(S)}(\mathbf{V}_{i+1} \otimes \mathbf{U}_{i+1})\|_F^2 + Ct. \quad (5.19)$$

La minimisation de f selon la variable \mathbf{w} avec la contrainte correspondante, revient alors à maximiser le critère suivant :

$$\begin{aligned} & \underset{\mathbf{w}}{\operatorname{argmax}} \|\mathbf{w}^t \mathbf{X}_{(L)}(\mathbf{V}_{i+1} \otimes \mathbf{U}_{i+1})\|_F^2, \\ & \text{t.q. } \mathbf{w}^t \mathbf{w} = 1. \end{aligned} \quad (5.20)$$

Trouver la solution optimale à ce problème revient simplement à appliquer la même étape que dans l'algorithme de la HOOI (Algorithme 3). Cette étape peut être vue comme le filtrage du tenseur \mathcal{X} en le projetant sur les sous-espaces générés par les matrices \mathbf{U}_{i+1} et \mathbf{V}_{i+1} , puis le calcul de l'espace dominant le mode des crises de ce tenseur filtré. Ces deux étapes reviennent à calculer le vecteur singulier dominant gauche de la matrice $\mathbf{X}_{(S)}(\mathbf{V}_{i+1} \otimes \mathbf{U}_{i+1})$. Une nouvelle fois, comme la solution optimale du critère de l'Équation (5.20) est obtenue, on a pour cette itération :

$$f(\mathbf{U}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_{i+1}) \leq f(\mathbf{U}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_i). \quad (5.21)$$

Algorithme mHOOI

Finalement, il est possible de proposer une implémentation différente de la HOOI, notée mHOOI (HOOI modifiée), minimisant le problème de la HOSVD de l'Équation (4.56) dans le cas où $K_L = K_T = K$ et $K_S = 1$.

Proposition 12

L'implémentation de la mHOOI est fournie dans l'Algorithme 6. On montre dans cet algorithme comment dériver le facteur de faible dimension \mathbf{F}_{mHOOI} . Cet algorithme est une procédure d'alternance des moindres carrés, dont on peut facilement prouver la convergence de la fonction de coût en combinant les Équations (5.18) et (5.21), on a :

$$f(\mathbf{U}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_{i+1}) \leq f(\mathbf{U}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_i) \leq f(\mathbf{U}_i, \mathbf{V}_i, \mathbf{w}_i). \quad (5.22)$$

L'algorithme HOOI, dans le cas $K_L = K_T = K$ et $K_S = 1$, a selon [16] et pour chaque itération, une complexité de $4(2K + 1)LTS$. Comme la mHOOI et la HOOI sont deux algorithmes qui réalisent plusieurs SVD, la complexité de l'Algorithme 6 est du même ordre mais est réduite à $4(K + 1)LTS$ opération pour chaque itération.

Il convient de noter que le critère d'arrêt que nous utilisons pour la procédure itérative est une métrique angulaire entre deux estimations successives de \mathbf{w}_i , ce qui épargne le calcul des coefficients $\mathbf{G}_{(L)i}$. En effet, si le vecteur \mathbf{w}_i n'est plus modifié à partir d'un certain nombre d'itérations, les solutions \mathbf{U}_i et \mathbf{V}_i resteront les mêmes. Comparé à l'algorithme HOOI, l'Algorithme 6 est composé de deux étapes au lieu de trois, et par conséquent, il devrait normalement converger plus rapidement. C'est ce qui est observé empiriquement dans l'Annexe 5.A

Algorithm 6 Estimation de \mathbf{F}_{mHOOI}

Require: \mathcal{X} , K , tolérance angulaire $\epsilon > 0$, et i_{max} .

$i = 1$

$\mathbf{w}_0 = \text{left_1_SVD}(\mathbf{X}_{(S)})$

while $i < i_{max}$ or $\text{acos}(\langle \mathbf{w}_i, \mathbf{w}_{i-1} \rangle) > \epsilon$ **do**

(A). $[\mathbf{U}_{i+1}^{(K)}, \mathbf{\Lambda}_{i+1}, \mathbf{V}_{i+1}^{(K)}] = K\text{-SVD}(\mathbf{X}_{(L)}(\mathbf{w}_i \otimes \mathbf{I}_T))$

(B). $\mathbf{w}_{i+1} = \text{left_1_SVD}(\mathbf{X}_{(S)}(\mathbf{V}_{i+1}^{(K)} \otimes \mathbf{U}_{i+1}^{(K)}))$

$i = i + 1$

end while

$\mathbf{F}_{mHOOI} = \mathbf{U}_i^{(K)} \mathbf{\Lambda}_i$

5.3 Implémentation de la HOSMR

On expose maintenant l'algorithme visant à trouver une solution au problème d'optimisation de l'Équation (5.7). La même logique que pour l'algorithme mHOOI (l'Algorithme 6) est suivie : le calcul de la SVD de la matrice de CFs générique est remplacé par une SVD parcimonieuse. Le principal avantage de cette décomposition est la prise en compte simultanée d'une régularisation parcimonieuse et de la contrainte elastic-Net. Nous présentons dans un premier temps l'algorithme de la SVD parcimonieuse utilisé en Section 5.3.1. L'algorithme de la HOSMR est alors présenté en Section 5.3.2

5.3.1 Implémentation d'une SVD parcimonieuse

La SVD parcimonieuse correspond à ajouter des contraintes de parcimonie sur chacun des facteurs d'une décomposition matricielle. Elle est déclinée en différentes versions, en imposant à la fois des contraintes d'orthogonalité et de parcimonie sur chaque facteur [204], ou en relaxant la contrainte d'orthogonalité [141]. On utilise la version relaxée, souvent préférée [6], [15], afin d'éviter le caractère compressif, et peu interprétable d'une contrainte d'orthogonalité. En considérant la matrice $\mathbf{X} \in \mathbb{R}^{L \times T}$, la version parcimonieuse de la SVD proposée cherche les matrices de rang faible $\mathbf{F} \in \mathbb{R}^{L \times K}$ et $\mathbf{V} \in \mathbb{R}^{T \times K}$, solutions du problème d'optimisation suivant :

$$\begin{aligned} \underset{\mathbf{F}, \mathbf{V}}{\operatorname{argmin}} \quad & \|\mathbf{X} - \mathbf{F}\mathbf{V}^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1, \\ \text{t.q.} \quad & \lambda_2 \|\mathbf{v}_{:k}\|_1 + \|\mathbf{v}_{:k}\|_F^2 \leq 1. \end{aligned} \quad (5.23)$$

Ici, les paramètres λ_1 et λ_2 permettent de régler le compromis entre la précision et la parcimonie de l'approximation. On résout ce problème en nous inspirant de l'article [141] qui présente un algorithme de SVD parcimonieuse. L'algorithme consiste à alterner deux étapes mettant à jour respectivement les matrices \mathbf{F} et \mathbf{V} .

Mise à jour des facteurs de CFs

À l'itération $i + 1$, on estime que la matrice \mathbf{V}_i est connue. Dans ce cas, le problème de l'Équation (5.23) consiste à trouver la matrice \mathbf{F} qui minimise :

$$\underset{\mathbf{F}}{\operatorname{argmin}} \quad \|\mathbf{X} - \mathbf{F}\mathbf{V}_i^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1. \quad (5.24)$$

Ce problème est bien connu, il s'agit d'une régularisation de lasso [184], [94]. Il existe de nombreux algorithmes pour le résoudre : méthode de Gauss-Siegel [94], méthode de direction alternée des multiplicateurs (ADMM) [37], [94], régression du moindre angle [67], [141]. Nous proposons d'utiliser, à la fois pour sa simplicité et ses bonnes performances [94], une méthode de gradient proximal alterné sur les colonnes de la matrice \mathbf{F} . Nous avons remarqué de meilleures performances en appliquant la méthode de gradient proximal sur chaque colonne, plutôt que sur chaque coordonnée ou sur l'ensemble de la matrice \mathbf{F} .

On rappelle que l'algorithme de descente de gradient permet de trouver le vecteur $\mathbf{y} \in \mathbb{R}^L$, qui minimise la somme de deux fonctions convexes et propres, telle que $h_1(\mathbf{y}) : \mathbb{R}^L \rightarrow \mathbb{R}$ soit différentiable et $h_2(\mathbf{y}) : \mathbb{R}^L \rightarrow \mathbb{R}$ potentiellement non différentiable en tout point. Une solution de :

$$\underset{\mathbf{y}}{\operatorname{argmin}} \quad h_1(\mathbf{y}) + h_2(\mathbf{y}), \quad (5.25)$$

peut être trouvée par applications successives de l'itération suivante :

$$\mathbf{y}_{j+1} = \operatorname{prox}_{\eta h_2(\bullet)}(\mathbf{y}_j - \eta \nabla h_1(\mathbf{y}_k)). \quad (5.26)$$

Avec $\operatorname{prox}_{\bullet}$ l'opérateur proximal décrit en Section 3.2.3, $\nabla \bullet$ l'opérateur gradient et η une constante. Si ∇h_1 est une fonction ν -Lipschitz, alors l'algorithme de gradient proximal converge si $\eta \in]0, \frac{1}{\nu}]$. Ainsi,

afin d'appliquer cette procédure sur une colonne $\mathbf{f}_{:k}$ de la matrice \mathbf{F} , on considère que (en éliminant pour simplifier l'index de l'itération i) :

$$h_1(\mathbf{f}_{:k}) = \left\| \left(\mathbf{X} - \sum_{k' \neq k} \mathbf{f}_{:k'} \mathbf{v}_{:k'}^t \right) - \mathbf{f}_{:k} \mathbf{v}_{:k}^t \right\|_F^2, \quad (5.27)$$

$$h_2(\mathbf{f}_{:k}) = \lambda_1 \|\mathbf{f}_{:k}\|_1. \quad (5.28)$$

Il suffit maintenant de déterminer le gradient, le coefficient de lipschitz du gradient de la fonction h_1 , et l'opérateur proximal de h_2 . Ce dernier correspond à la fonction de seuillage doux montrée dans l'Équation (3.70), notée $\text{SD}(\bullet)$:

$$\text{prox}_{\eta\lambda_1\|\bullet\|_1}(x) = \text{SD}_{\eta\lambda_1}(x) = \text{sign}(x) \max(|x| - \eta\lambda_1, 0). \quad (5.29)$$

Afin de déterminer le gradient de la fonction h_1 , on note dans un premier temps la matrice $\mathbf{R} = \mathbf{X} - \sum_{k' \neq k} \mathbf{f}_{:k'} \mathbf{v}_{:k'}^t$, on peut ainsi facilement déterminer que :

$$\nabla h_1(\mathbf{f}_{:k}) = 2\mathbf{f}_{:k} \mathbf{v}_{:k}^t \mathbf{v}_{:k} - 2\mathbf{R} \mathbf{v}_{:k}, \quad (5.30)$$

alors par simple développement de la matrice \mathbf{R} , on trouve :

$$\begin{aligned} \nabla h_1(\mathbf{f}_{:k}) &= 2\mathbf{f}_{:k} \mathbf{v}_{:k}^t \mathbf{v}_{:k} - 2\left(\mathbf{X} - \sum_{k' \neq k} \mathbf{f}_{:k'} \mathbf{v}_{:k'}^t\right) \mathbf{v}_{:k}, \\ &= 2\mathbf{F} \mathbf{V}^t \mathbf{v}_{:k} - 2\mathbf{X} \mathbf{v}_{:k}, \\ &= 2\mathbf{F} [\mathbf{V}^t \mathbf{V}]_{:k} - 2\mathbf{X} \mathbf{v}_{:k}. \end{aligned} \quad (5.31)$$

Enfin, le Lipschitzien de ce gradient se calcul directement à partir de sa définition :

$$\nu = 2\mathbf{v}_{:k}^t \mathbf{v}_{:k}. \quad (5.32)$$

Finalement, une itération de la méthode de gradient proximal, appliquée successivement aux colonnes de \mathbf{F} de l'Équation (5.24) correspond à :

$$(\mathbf{f}_{:k})_{j+1} = \text{SD}_{\frac{\lambda_1}{2\mathbf{v}_{:k}^t \mathbf{v}_{:k}}} \left[(\mathbf{f}_{:k})_j - \frac{1}{\mathbf{v}_{:k}^t \mathbf{v}_{:k}} (\mathbf{F}_j [\mathbf{V}^t \mathbf{V}]_{:k} - \mathbf{X} \mathbf{v}_{:k}) \right]. \quad (5.33)$$

La procédure afin de résoudre l'Équation (5.24) est fournie dans l'Algorithme 7.

Algorithm 7 Résout la régression parcimonieuse de l'Équation (5.24)

Require: \mathbf{X} , \mathbf{V}_i , K , le paramètre λ_1 , tolérance $\epsilon > 0$, et j_{max} . Une première supposition \mathbf{F}_0

$j = 1$

$\mathbf{D} = \mathbf{V}^t \mathbf{V}$

while $j < j_{max}$ or $\|\mathbf{X} - \mathbf{F}_j \mathbf{V}_i^t\|_F^2 + \lambda_1 \|\mathbf{F}_j\|_1 - \|\mathbf{X} - \mathbf{F}_{j+1} \mathbf{V}_i^t\|_F^2 - \lambda_1 \|\mathbf{F}_{j+1}\|_1 > \epsilon$ **do**

for $k \in \{1, \dots, K\}$ **do**

$\mathbf{y} = (\mathbf{f}_{:k})_j - \frac{1}{d_{kk}} (\mathbf{F}_j \mathbf{D}_{:k} - \mathbf{X} \mathbf{v}_{:k})$

$(\mathbf{f}_{:k})_{j+1} = \text{SD}_{\frac{\lambda_1}{2d_{kk}}}(\mathbf{y})$

end for

$j = j + 1$

end while

Mise à jour des facteurs de profils d'activation

Toujours à l'itération $i + 1$, on estime que la matrice \mathbf{F}_{i+1} est maintenant connue. Dans ce cas, le problème de l'Équation (5.23) consiste à trouver la matrice \mathbf{V} qui minimise :

$$\begin{aligned} \underset{\mathbf{V}}{\text{argmin}} \quad & \|\mathbf{X} - \mathbf{F} \mathbf{V}^t\|_F^2, \\ \text{t.q.} \quad & \lambda_2 \|\mathbf{v}_{:k}\|_1 + \|\mathbf{v}_{:k}\|_F^2 \leq 1. \end{aligned} \quad (5.34)$$

On obtient maintenant un problème de régression sous la contrainte élastic-net. Il s'agit d'un problème connu, nous appliquons ici l'implémentation proposée dans [141] qui consiste en une méthode de projection du gradient, appliquée sur les colonnes de \mathbf{V} .

On rappelle qu'une méthode de projection du gradient, appliquée à une fonction $h(\mathbf{y}) : \mathbb{R}^T \rightarrow \mathbb{R}$ soumise à une contrainte \mathcal{C} , consiste à appliquer l'opération suivante :

$$\mathbf{y}_{k+1} = \mathbb{P}_{\mathcal{C}} [\mathbf{y}_k + \eta \nabla h(\mathbf{y}_k)], \quad (5.35)$$

avec $\mathbb{P}_{\mathcal{C}} [\bullet]$ l'opération qui projette le résultat de la descente de gradient sur la contrainte considérée. Pour que l'algorithme converge, les conditions sur η sont similaires à celles d'une méthode de descente de gradient, soit $\eta \in]0, \frac{2}{\nu}[$, si ∇h est une fonction ν -Lipschitzienne. Dans notre cas on a :

$$h(\mathbf{v}_{:k}) = \left\| \mathbf{X} - \sum_{k' \neq k} \mathbf{f}_{:k'} \mathbf{v}_{:k'}^t - \mathbf{f}_{:k} \mathbf{v}_{:k}^t \right\|_F^2, \quad (5.36)$$

$$\mathcal{C} : \lambda_2 \|\mathbf{v}_{:k}\|_1 + \|\mathbf{v}_{:k}\|_F^2 \leq 1. \quad (5.37)$$

Le calcul du gradient ∇h et de son Lipschitzien est simple dans la mesure où il reprend exactement les étapes des Équations (5.30), (5.31) et (5.32), on a donc :

$$\nabla h(\mathbf{v}_{:k}) = 2\mathbf{V}[\mathbf{F}^t \mathbf{F}]_{:k} - 2\mathbf{X}^t \mathbf{f}_{:k} \quad (5.38)$$

$$\nu = 2\mathbf{f}_{:k}^t \mathbf{f}_{:k}. \quad (5.39)$$

La projection sur la contrainte elastic-Net de l'étape de descente de gradient (noté $\bar{\mathbf{y}}$), consiste à trouver le vecteur \mathbf{y} , minimisant sa distance avec $\bar{\mathbf{y}}$, et appartenant à l'espace formé par l'ensemble des solutions admissibles par la contrainte. Cette projection s'écrit :

$$\begin{aligned} & \underset{\mathbf{y}}{\operatorname{argmin}} \frac{1}{2} \|\bar{\mathbf{y}} - \mathbf{y}\|_F^2, \\ & \text{t.q. } \lambda_2 \|\mathbf{y}\|_1 + \|\mathbf{y}\|_F^2 \leq 1, \end{aligned} \quad (5.40)$$

selon :

$$\bar{\mathbf{y}} = \mathbf{v}_{:k} - \frac{2}{\nu} (\mathbf{V}[\mathbf{F}^t \mathbf{F}]_{:k} - \mathbf{X}^t \mathbf{f}_{:k}). \quad (5.41)$$

On utilise pour résoudre l'Équation (5.40) son Lagrangien noté :

$$\mathcal{L}(\mathbf{y}, \gamma) = \frac{1}{2} \|\bar{\mathbf{y}} - \mathbf{y}\|_F^2 + \gamma (\lambda_2 \|\mathbf{y}\|_1 + \|\mathbf{y}\|_F^2 - 1). \quad (5.42)$$

On sait alors que la solution optimale appartient à l'ensemble des solutions $\mathbf{y}^*(\gamma)$ minimisant le lagrangien selon la variable \mathbf{y} , cet ensemble correspond à :

$$\begin{aligned} \mathbf{y}^*(\gamma) &= \underset{\mathbf{y}}{\operatorname{argmin}} \mathcal{L}(\mathbf{y}, \gamma), \\ &= \operatorname{prox}_{\gamma(\lambda_2 \|\bullet\|_1 + \|\bullet\|_F^2 - 1)}(\bar{\mathbf{y}}), \\ &= \frac{\operatorname{SD}_{\gamma \lambda_2}(\bar{\mathbf{y}})}{1 + 2\gamma}. \end{aligned} \quad (5.43)$$

Les conditions complémentaires de relâchement de Karush-Kuhn-Tucker stipulent que si le résultat de la descente de gradient $\bar{\mathbf{y}}$ appartient à l'ensemble des solutions admissibles de la contrainte \mathcal{C} , alors la solution optimale du problème de l'Équation (5.40) est $\mathbf{y}^*(0) = \bar{\mathbf{y}}$. Si ce n'est pas le cas, on sait que la solution optimale se situe sur les bords de la contrainte. On cherche alors la valeur de γ tel que :

$$\lambda_2 \|\mathbf{y}^*(\gamma)\|_1 + \|\mathbf{y}^*(\gamma)\|_F^2 = 1. \quad (5.44)$$

En développant on obtient :

$$\frac{1}{(1 + 2\gamma)^2} \sum_{l \in \mathcal{S}(\lambda_2 \gamma)} \left[\lambda_2 |\bar{y}_l| + |\bar{y}_l|^2 - \lambda_2^2 \gamma (1 + \gamma) \right] = 1, \quad (5.45)$$

avec $\mathcal{S}(\lambda_2\gamma) = \{l \text{ t.q. } |\bar{y}_l| \geq \lambda_2\gamma\}$, l'ensemble des indices non nuls après application du seuillage doux. Afin de trouver une solution à l'Équation (5.45), on cherche dans un premier temps une sur-estimation de la valeur optimale de γ parmi les valeur du vecteurs $\bar{\mathbf{y}}$. On cherche alors l'élément de ce vecteur permettant de résoudre le problème suivant :

$$\begin{aligned} & \operatorname{argmax}_{l \in \{1, \dots, L\}} |\bar{y}_l|, \\ \text{t.q. } & \frac{1}{(1 + 2|\bar{y}_l|)^2} \sum_{l' \in \mathcal{S}(\lambda_2|\bar{y}_l|)} \left[\lambda_2 |\bar{y}_{l'}| + |\bar{y}_{l'}|^2 - \lambda_2^2 |\bar{y}_l| (1 + |\bar{y}_l|) \right] < 1. \end{aligned} \quad (5.46)$$

On utilisera l'algorithme proposé dans [141] afin de résoudre efficacement le problème de l'Équation (5.46). L'algorithme de [141] adapté à la définition de la contrainte elastic-net de l'Équation (5.37) est décrit dans l'Annexe (5.B). On note l^* l'indice résultant de l'optimisation de l'Équation (5.46) et L^* le nombre d'indice dans l'ensemble $\mathcal{S}(\lambda_2|\bar{y}_{l^*}|)$. On peut désormais obtenir une solution explicite γ^* permettant de résoudre l'Équation (5.45). En effet, grâce à la résolution de l'Équation (5.46), on sait que les indices contenus dans $\mathcal{S}(\lambda_2\gamma^*)$ et $\mathcal{S}(\lambda_2|\bar{y}_{l^*}|)$ seront les mêmes. Ainsi trouver γ^* revient à déterminer le zéro maximal du polynôme du second ordre suivant :

$$\frac{1}{(1 + 2\gamma)^2} \sum_{l \in \mathcal{S}(\lambda_2|\bar{y}_{l^*}|)} \left[\lambda_2 |\bar{y}_l| + |\bar{y}_l|^2 - \lambda_2^2 \gamma (1 + \gamma) \right] = 1, \quad (5.47)$$

$$\Leftrightarrow \gamma^2 + \gamma + \frac{1}{L^* \lambda_2^2 + 4} \left(1 - \sum_{l \in \mathcal{S}(\lambda_2|\bar{y}_{l^*}|)} \lambda_2 |\bar{y}_l| + |\bar{y}_l|^2 \right) = 0. \quad (5.48)$$

La démarche permettant de trouver la solution du problème de l'Équation (5.34) est résumée dans l'Algorithme 8.

Algorithm 8 Résout la régression sous contrainte Élastic-net de l'Équation (5.34)

Require: \mathbf{X} , \mathbf{F}_{i+1} , K , le paramètre λ_2 , tolérance $\epsilon > 0$, et j_{max} . Une première supposition \mathbf{V}_0

$j = 1$

$\mathbf{D} = \mathbf{F}^t \mathbf{F}$

while $j < j_{max}$ or $\|\mathbf{X} - \mathbf{F}_{i+1} \mathbf{V}_j^t\|_F^2 - \|\mathbf{X} - \mathbf{F}_{i+1} \mathbf{V}_{j+1}^t\|_F^2 > \epsilon$ **do**

for $k \in \{1, \dots, K\}$ **do**

$\mathbf{y} = (\mathbf{v}_{:k})_j - \frac{1}{d_{kk}} (\mathbf{V}_j \mathbf{D}_{:k} - \mathbf{X}^t \mathbf{v}_{:k})$

if $\lambda_2 \|\mathbf{y}\|_1 + \|\mathbf{y}\|_F^2 \leq 1$ **then**

$\mathbf{v}_{:k} = \mathbf{y}$

else

 Trouver l'indice l^* maximisant le problème de l'Équation (5.46)

 Déterminer γ^* comme zéro maximal du polynôme de l'Équation (5.48)

$\mathbf{v}_{:k} = \frac{\text{SD}_{\gamma^* \lambda_2}(\bar{\mathbf{y}})}{1 + 2\gamma^*}$

end if

end for

$j = j + 1$

end while

Algorithme de la SVD parcimonieuse

L'Algorithme 9 décrit l'implémentation de la SVD parcimonieuse (Équation (5.23)), qui consiste donc à alterner les deux algorithmes précédents (Algorithmes 7 et 8) jusqu'à convergence. Une preuve de convergence, sous conditions, est proposée dans [141]. Notons que cet algorithme correspond à un problème non convexe sensible à l'initialisation. Sauf indication contraire, on choisit comme initialisation des matrices \mathbf{F}_0 (resp. \mathbf{V}_0) qui sont des colonnes (resp. lignes) équitablement réparties de la matrice de CFs générique du tenseur \mathcal{X} .

Algorithm 9 Résout le problème de la SVD parcimonieuse de l'Équation (5.23)

Require: \mathbf{X} , K , les paramètres λ_1 et λ_2 , la tolérance $\epsilon > 0$, et i_{max} . Une première supposition $\mathbf{F}_0, \mathbf{V}_0$

$i = 1$

while $i < i_{max}$ or $\|\mathbf{X} - \mathbf{F}_i \mathbf{V}_i^t\|_F^2 + \lambda_1 \|\mathbf{F}_i\|_1 - \|\mathbf{X} - \mathbf{F}_{i+1} \mathbf{V}_{i+1}^t\|_F^2 - \lambda_1 \|\mathbf{F}_{i+1}\|_1 > \epsilon$ **do**

Estimer \mathbf{F}_{i+1} en utilisant l'Algorithme 7 avec pour entrée \mathbf{X} et \mathbf{V}_i et λ_1 .

Estimer \mathbf{V}_{i+1} en utilisant l'Algorithme 8 avec pour entrée \mathbf{X} et \mathbf{F}_{i+1} et λ_2 .

$i = i + 1$

end while

5.3.2 Algorithme de la HOSMR

L'algorithme de la HOSMR permettant de résoudre le problème de l'Équation (5.7), consiste, comme l'algorithme de la mHOOI à alterner :

Étape (A) Factorisation de la matrice de CFs générique

À l'itération $i + 1$, on suppose le vecteur permettant de calculer la matrice de CFs générique \mathbf{w}_i connue. La fonction de coût g de l'Équation (5.10) du problème de la HOSMR, peut être ré-écrite selon le mode des crises, et équivaut à :

$$\begin{aligned} g(\mathbf{F}, \mathbf{V}, \mathbf{w}_i) &= \|\mathbf{X}_{(L)} - \mathbf{F} \mathbf{I}_K (\mathbf{w}_i \otimes \mathbf{V})^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1, \\ &= \|\mathbf{X}_{(S)} - \mathbf{w}_i \text{vec}(\mathbf{I}_K)^t (\mathbf{V} \otimes \mathbf{F})^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1. \end{aligned} \quad (5.49)$$

Alors, comme on contraint $\mathbf{w}_i^t \mathbf{w}_i = 1$, on peut écrire :

$$g(\mathbf{F}, \mathbf{V}, \mathbf{w}_i) = \|\mathbf{w}_i^t \mathbf{X}_{(S)} - \text{vec}(\mathbf{I}_K)^t (\mathbf{V} \otimes \mathbf{F})^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1. \quad (5.50)$$

Finalement, en utilisant la formule de compatibilité entre le produit de Kronecker et la vectorisation de l'Équation (4.53), et en rappelant que :

$$\mathbf{w}_i^t \mathbf{X}_{(S)} = \text{vec}(\mathbf{X}_L (\mathbf{w}_i \otimes \mathbf{I}_T))^t, \quad (5.51)$$

on obtient directement :

$$\begin{aligned} g(\mathbf{F}, \mathbf{V}, \mathbf{w}_i) &= \|\text{vec}(\mathbf{X}_{(S)} (\mathbf{w}_i \otimes \mathbf{I}_T) - \mathbf{F} \mathbf{V}^t)^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1, \\ &= \|\mathbf{X}_{(S)} (\mathbf{w}_i \otimes \mathbf{I}_T) - \mathbf{F} \mathbf{V}^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1. \end{aligned} \quad (5.52)$$

On remarque alors que minimiser la fonction de coût $g(\mathbf{F}, \mathbf{V}, \mathbf{w}_i)$ sous contrainte d'elastic-Net sur les colonnes de \mathbf{V} correspond à résoudre la SVD parcimonieuse suivante :

$$\begin{aligned} \underset{\mathbf{F}, \mathbf{V}}{\text{argmin}} \quad & \|\mathbf{X}_{(S)} (\mathbf{w}_i \otimes \mathbf{I}_T) - \mathbf{F} \mathbf{V}^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1, \\ \text{t.q.} \quad & \lambda_2 \|\mathbf{v}_{:k}\|_1 + \|\mathbf{v}_{:k}\|_F^2 \leq 1. \end{aligned} \quad (5.53)$$

On utilise pour résoudre ce problème l'Algorithme 9. À partir de la seconde itération, on propose d'initialiser l'Algorithme 9 en utilisant les matrices de l'itération précédente \mathbf{F}_i et \mathbf{V}_i . Ainsi, en remarquant que cet algorithme converge comme montré dans [141], on a :

$$g(\mathbf{F}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_i) < g(\mathbf{F}_i, \mathbf{V}_i, \mathbf{w}_i). \quad (5.54)$$

Après avoir résolu le problème de l'Équation (5.53), on normalise les colonnes de \mathbf{F}_{i+1} et de \mathbf{V}_{i+1} telles que $\|\mathbf{v}_{:k}\|_F^2 = 1 \forall k \in 1, \dots, K$, afin que toute la variance soit contenue dans la matrice \mathbf{F}_{i+1} .

Étape (B) Mise à jour de la contribution de chaque crise

On suppose, pour la seconde étape de l'algorithme, que les matrices \mathbf{F}_{i+1} et \mathbf{V}_{i+1} sont connues. Alors la fonction de coût g du problème de la HOSMR de l'Équation (5.7) correspond à :

$$g(\mathbf{F}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}) = \|\mathbf{X}_{(S)} - \mathbf{w} \text{vec}(\mathbf{I}_K)^t (\mathbf{V}_{i+1} \otimes \mathbf{F}_{i+1})^t\|_F^2 + Ct.$$

En remarquant que la matrice \mathbf{I}_K est diagonale, par application de la deuxième formule de compatibilité entre le produit de Kronecker et le produit de Khatri-Rao, on obtient :

$$g(\mathbf{F}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}) = \|\mathbf{X}_{(S)} - \mathbf{w} \mathbf{1}^t (\mathbf{V} \odot \mathbf{F})^t\|_F^2 + Ct, \quad (5.55)$$

avec $\mathbf{1} \in \mathbb{R}^K$ vecteur dont chaque élément vaut un. Ainsi, minimiser l'Équation (5.55) selon le vecteur \mathbf{w} , avec la contrainte $\mathbf{w}^t \mathbf{w} = 1$ correspond à résoudre le problème suivant :

$$\begin{aligned} \underset{\mathbf{w}}{\text{argmin}} \quad & \|\mathbf{X}_{(S)} - \mathbf{w} \mathbf{1}^t (\mathbf{V} \odot \mathbf{F})^t\|_F^2, \\ \text{t.q.} \quad & \mathbf{w}^t \mathbf{w} = 1. \end{aligned} \quad (5.56)$$

Ce problème correspond à une régression des moindres carrés sous contrainte d'appartenance à la sphère unité, ce qui est un cas particulier de régression de Ridge [181]. En notant, pour simplifier, $\mathbf{p}^t = \mathbf{1}^t (\mathbf{V} \odot \mathbf{F})^t$, le lagrangien associé à l'Équation (5.56) correspond à :

$$\mathcal{L}(\mathbf{w}, \gamma) = \mathbf{w} \mathbf{p}^t \mathbf{p} \mathbf{w}^t - 2 \mathbf{w} \mathbf{p}^t \mathbf{X}_{(S)}^t + \gamma (\mathbf{w}^t \mathbf{w} - 1). \quad (5.57)$$

Cela donne, après dérivation de l'Équation (5.57) selon le vecteur \mathbf{w} et le paramètre γ , les conditions suivantes :

$$\mathbf{w}_{i+1} = \mathbf{X}_{(S)} (\mathbf{p}^t \mathbf{p} + \gamma) \mathbf{p} = \bar{\gamma} \mathbf{X}_{(S)} \mathbf{p}, \quad (5.58)$$

$$\mathbf{w}_{i+1}^t \mathbf{w}_{i+1} = 1. \quad (5.59)$$

Comme $\mathbf{X}_{(S)} \mathbf{p} \in \mathbb{R}^S$ est un vecteur, \mathbf{w}_{i+1} doit engendrer le même sous-espace. Ainsi, pour satisfaire la contrainte de l'Équation (5.59), il suffit de trouver une valeur de $\bar{\gamma}$ tel que \mathbf{w}_{i+1} soit de norme 1, fournissant la solution suivante :

$$\mathbf{w}_{i+1} = \frac{\mathbf{X}_{(S)} (\mathbf{V}_{i+1} \odot \mathbf{F}_{i+1}) \mathbf{1}}{\|\mathbf{X}_{(S)} (\mathbf{V}_{i+1} \odot \mathbf{F}_{i+1}) \mathbf{1}\|_F}. \quad (5.60)$$

Comme cette mise à jour du vecteur unitaire \mathbf{w}_{i+1} minimise l'Équation (5.56), on a :

$$g(\mathbf{F}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_{i+1}) \leq g(\mathbf{F}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_i). \quad (5.61)$$

Algorithme de la HOSMR

Proposition 13

L'implémentation de la HOSMR de l'Équation (5.7) est fournie dans l'Algorithme 10. On peut aussi établir la convergence de la fonction de coût, en considérant les Équations (5.54) et (5.61), on a :

$$g(\mathbf{F}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_{i+1}) \leq g(\mathbf{F}_i, \mathbf{V}_i, \mathbf{w}_i). \quad (5.62)$$

On rappelle que l'Algorithme 9 permet de réaliser une SVD parcimonieuse, ce dernier étant divisé en deux sous-algorithmes (Algorithmes 7 et 8) afin d'alterner la mise à jour des facteurs de CFs et les profils d'activation temporelle.

On précise que l'ajout de la contrainte de non-négativité, à la fois sur les matrices \mathbf{F} et \mathbf{V} , est ici très simple. Il suffit de remplacer dans l'Algorithme 9 de la SVD parcimonieuse, la fonction de seuillage doux $\text{SD}(\bullet)$, par sa version non négative, notée $\text{NSD}(\bullet)$ avec :

$$\text{NSD}_\lambda(x) = \max(x - \lambda, 0). \quad (5.63)$$

Cette fonction assure aux matrices \mathbf{F} et \mathbf{V} d'être non négatives. Finalement, on précise que le vecteur \mathbf{w} sera par défaut non négatif si la matrice $\mathbf{X}_{(S)}$ est non négative. En effet, la matrice $(\mathbf{V} \odot \mathbf{F})\mathbf{1}$ est aussi non négative, ainsi la multiplication matricielle des Équation (5.60) ne peut pas produire d'éléments négatifs. La contrainte de non-négativité n'est pas ajoutée car elle ne semble pas déterminante pour obtenir les résultats qui suivent.

Algorithm 10 Estimation of \mathbf{F}_{HOSMR}

Require: \mathcal{X} , K , λ_1 , λ_2 , tolérance $\epsilon > 0$, et i_{max} .

$i = 0$

$\mathbf{w}_0 = \text{left_1_SVD}(\mathbf{X}_{(S)})$

while $i < i_{max}$ or $\text{acos}(\langle \mathbf{w}_i, \mathbf{w}_{i-1} \rangle) > \epsilon$ **do**

(A). Trouver \mathbf{F}_{i+1} , \mathbf{V}_{i+1} en appliquant l'Algorithme 9 avec $\mathbf{X}_{(L)}(\mathbf{w}_i \otimes \mathbf{I}_T)$, \mathbf{F}_i , \mathbf{V}_i , λ_1 et λ_2 .

$\alpha_k = \|\mathbf{v}_{:k}\|_F^2$, $\mathbf{v}_{:k} = \frac{1}{\alpha_k} \mathbf{v}_{:k}$, $\mathbf{f}_{:k} = \alpha_k \mathbf{f}_{:k}$

(B). $\mathbf{w}_{i+1} = \mathbf{X}_{(S)}(\mathbf{V}_{i+1} \odot \mathbf{F}_{i+1})\mathbf{1} / \|\mathbf{X}_{(S)}(\mathbf{V}_{i+1} \odot \mathbf{F}_{i+1})\mathbf{1}\|_F$

$i = i + 1$

end while

$\mathbf{F}_{HOSMR} = \mathbf{F}_i$

5.4 Évaluations numériques sur un modèle

Dans cette section, nous évaluons les performances de la HOSMR pour décomposer un graphe dynamique et pour estimer des facteurs de CFs. On le compare à d'autres méthodes classiques de réduction dimensionnelle. À cette fin, et comme notre application concerne les signaux sEEG, on présente dans un premier temps, un modèle simplifié et original de graphe dynamique de CFs qui intègre quatre sources d'incertitudes capables de reproduire de nombreuses variabilités expérimentales constatées pour des signaux sEEG.

5.4.1 Un modèle de graphe dynamique pour les CFs

Une crise d'épilepsie démarre souvent par un groupe de CFs focales qui se propage dans les régions connectées du cerveau, et parfois se divise pour donner naissance à de nouveaux facteurs de CFs. La Figure 5.2.(a) représente une matrice caractéristique de CFs mesurée par Phase Locking Value (PLV, voir la Définition 9) [192]. Le modèle proposé est amené à reproduire la structure globale des CFs mesurées. Plus précisément, on considère une matrice $\mathbf{X} \in \mathbb{R}^{L \times T}$, où x_{lt} est une valeur élevée si l'indice de CF $l \in \{1, \dots, L\}$ est actif au temps $t \in \{1, \dots, T\}$, et correspond sinon à une valeur faible. Pour tenir compte des écarts moyens entre une CF active et inactive, on choisit respectivement des valeurs égales à 0.7 et 0.2. Une partition C_n , $n \in \{1, \dots, N\}$ est composée de l'ensemble des CFs qui sont actives sur la même période de temps T_n , il s'agit de la partie déterministe du modèle, définissant la structure et les temps d'activation des CFs. On superpose ensuite des aléas qui sont observés sur les signaux de CFs de crises réelles. Ces quatre sources d'incertitudes sont :

1. Durée aléatoire : chaque période d'activation T_n est uniformément distribuée entre une durée minimum (ici de 5 échantillons temporels) et $\frac{T}{N}$ échantillons. On associe à cette incertitude une variable binaire α , qui vaut 1 si l'on souhaite prendre en compte ce bruit et sinon 0.
2. Erreur d'activation : avec probabilité $\beta \in [0, 1]$, chaque CF d'un groupe C_n encourt le risque d'être remplacée par une autre CF choisie au hasard.
3. Bruit de connectivité : on ajoute à x_{lt} un bruit blanc Gaussien et centré de puissance σ^2 .
4. Gigue : Toutes les CFs d'un même groupe s'activent de manière indépendante avec une gigue, uniformément distribuée entre $[-\frac{\delta}{2}, \frac{\delta}{2}]$.

Le tracé (b) de la figure 5.2 représente une réalisation d'un réseau dynamique synthétique sous forme d'une matrice de CFs. En comparant avec des données réelles, le modèle réussit à reproduire globalement la dynamique des CFs. Plus important encore, ce modèle permet de contrôler la nature et l'intensité des incertitudes de chaque matrice. Cela permet d'évaluer la sensibilité et la robustesse des méthodes de réduction de dimension, par rapport à chaque source d'incertitude. Ainsi, on simule différentes

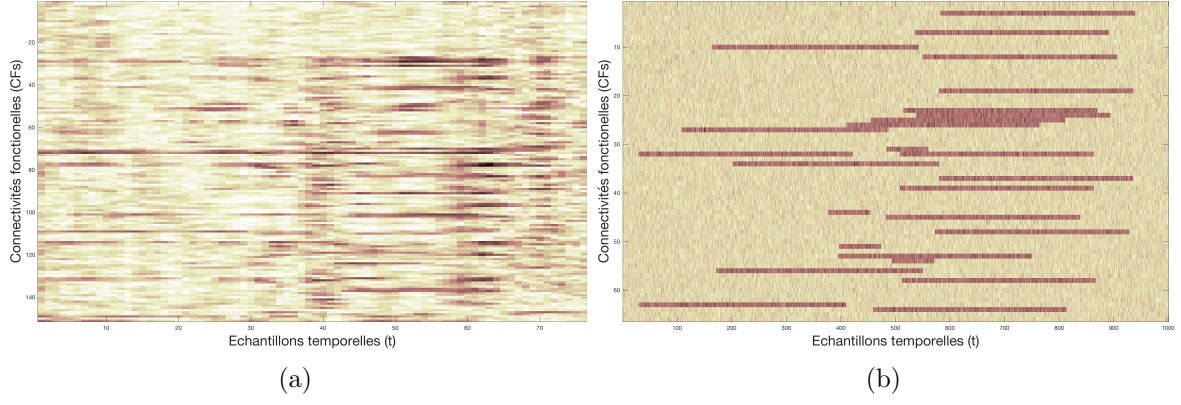


FIGURE 5.2 – (a) Matrice de CFs de données sEEG réelles, (b) exemple d'un modèle synthétique pour une crise avec comme paramètres d'incertitudes $\mathbf{b} = [1, 0.2, 0.1, 0.1]$.

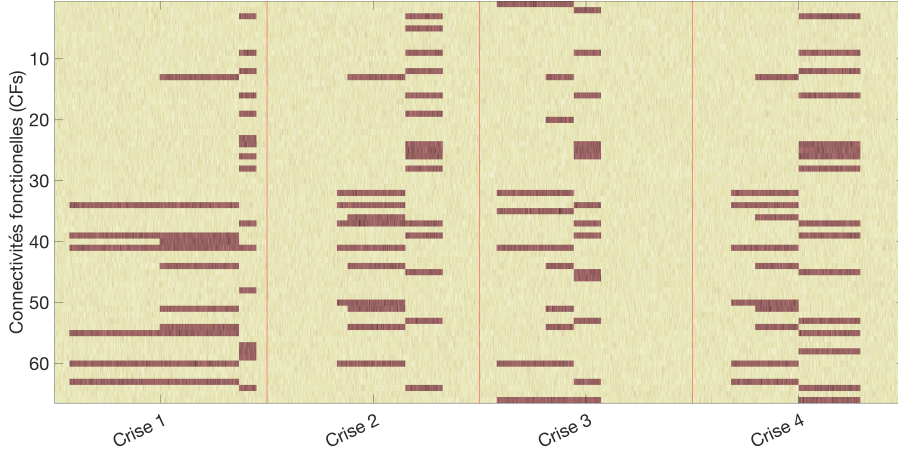


FIGURE 5.3 – Réalisation du tenseur matricé $\mathbf{X}_{(L)}$ correspondant à 4 crises (ou matrice de CFs) avec pour paramètres d'incertitudes $\mathbf{b} = [1, 0.2, 0.05, 0]$.

matrices de CFs provenant de plusieurs crises d'un même patient comme des réalisations indépendantes et identiquement distribuées de notre modèle avec les mêmes paramètres d'incertitude. Ceux-ci sont placés dans un vecteur $\mathbf{b} = [\alpha, \beta, \sigma, \delta]$. Les S crises simulées sont alors empilées dans un tenseur $\mathcal{X} \in \mathbb{R}^{L \times T \times S}$. La Figure 5.3 représente un exemple de tenseur matricé $\mathbf{X}_{(L)}$, correspondant au choix particulier $[\alpha = 1, \beta = 0.2, \sigma = 0.05, \delta = 0]$, $L = 66$, $T = 1000$ et $S = 4$.

5.4.2 Description de l'expérience

En utilisant le modèle de CFs, on peut maintenant comparer la méthode HOSMR proposée Section 5.3 à des méthodes classiques de réduction dimensionnelle. Chaque méthode produit des facteurs de CFs $\mathbf{F} \in \mathbb{R}^{L \times K}$ utilisés comme entrée pour l'algorithme k-means. Les performances se réfèrent ici à la capacité à récupérer les partitions de CFs du modèle global. Le Tableau 5.1 résume les 8 méthodes comparées, indiquant pour chacune l'implémentation utilisée ainsi que les valeurs des hyperparamètres déterminées empiriquement. Comme repère, on considère les deux approches directes consistant à appliquer k-means directement sur le tenseur matricé $\mathbf{F}_{Direct} = X_{(L)}$ ainsi que sur la moyenne de chaque crise $\mathbf{F}_{mean} = \frac{1}{S} \sum_{s=1}^S X_{::s}$. Aussi, d'autres approches classiques sont utilisées, comme la NMF (présentée en Section 4.1.3) appliquée sur $\mathbf{X}_{(L)}$, et la nnCP (présentée en Section 4.3.1) appliquée sur le tenseur \mathcal{X} . Pour ces méthodes, on normalise les matrices obtenues de sorte que l'ensemble de la variance soit contenue respectivement dans les matrices \mathbf{F}_{NMF} et \mathbf{F}_{nnCP} . Finalement, on considère les méthodes proposées comme la HOSMR, la SVD parcimonieuse appliquée sur $\mathbf{X}_{(L)}$ ainsi que la mHOOI. La dimension réduite K varie entre 1 et 4, et dans notre comparaison, seul le paramètre K donnant le meilleur score est considéré pour chaque méthode et chaque réalisation. Concernant l'algorithme k-means, on utilise la version *k-means++* de [10] que l'on arrête après 1000

Méthode	Section	Implémentation	Paramètres
1 - Direct	5.4.2	(1)	-
2 - Mean	5.4.2	(1)	-
3 - NMF	4.1.3	(1)	$\epsilon = 10^{-4}$
4 - nnCP	4.3.1	(2)	$\epsilon = 10^{-6}$
5 - SVD	4.1.3	(1)	-
6 - mHOOI	5.2.2	(1)-(3)	$\epsilon = 10^{-3}$
7 - sSVD	5.3.1	(1)-(4)	$\lambda_1 = 0.1, \lambda_2 = 1, \epsilon = 10^{-3}$
8 - HOSMR	5.3.2	(1)-(3)-(4)	$\lambda_1 = 0.1, \lambda_2 = 4, \epsilon = 10^{-3}$

Tableau 5.1 – Liste des méthodes comparées afin de réduire le tenseur \mathcal{X} en une matrice de facteur de CFs \mathbf{F} . Les implémentations utilisées pour les expériences peuvent être trouvées en utilisant les liens suivants : (1) Algorithmes utilisés dans l'article [77] FCTensDec; (2) N-way toolbox version 3.30[9]; (3) MATLAB Tensor Toolbox Version 2.6 [17]; (4) SPAMS toolbox version 2.6 [140]

$b_i = [\alpha, \beta, \sigma, \delta]$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
α	✓			✓		✓	✓			✓			✓	✓	✓
β			✓	✓	✓		✓		✓			✓	✓		✓
σ		✓			✓	✓	✓				✓	✓		✓	✓
δ								✓	✓	✓	✓	✓	✓	✓	✓

Tableau 5.2 – Une encoche est indiquée lorsque l'incertitude b_i est active dans la génération du modèle synthétique.

itérations, répétée 120 fois avec différentes initialisations. Le nombre de classes est fixé à $P = 4$ (3 facteurs d'activation plus le facteur de CFs désynchronisé). Finalement, afin d'évaluer les performances de partitionnement de chaque méthode, on utilise le score nommé Adjusted Rand Index (ARI) [99]. Ce score, calculé entre les partitions de CFs et la vérité terrain, vaut 1 pour une correspondance parfaite et 0 si la correspondance ne surpasse pas une solution de partition aléatoire. Une présentation du score ARI est fourni en Annexe 5.C.

5.4.3 Résultats et discussions

Sous ces conditions expérimentales, on évalue chaque méthode pour différentes configurations d'incertitudes résumées dans le Tableau 5.2. La Figure 5.4 représente, sous forme de boîtes à moustaches, la moyenne et l'écart type de 120 réalisations indépendantes du tenseur \mathcal{X} pour six configurations différentes impliquant les incertitudes suivantes : (i) durée aléatoire, (ii) erreur d'activation et (iii) bruit de connectivité, ici la gigue est constamment désactivée. La Figure 5.5, représente les résultats obtenus pour les configurations (8) à (15) du Tableau 5.2. À partir de la Figure 5.4, on observe que les méthodes appliquant des réductions dimensionnelles avant k-means ont de meilleures performances que les méthodes directes. Les méthodes tensorielles ont de meilleurs résultats que leurs homologues matricielles, et les méthodes parcimonieuses surpassent généralement les méthodes conventionnelles. La contrainte de non-négativité ne semble pas utile ici. Globalement, la HOSMR se distingue systématiquement, et surtout pour de faibles SNRs, notamment pour la configuration avec bruit de connectivité seule (expérience b_2).

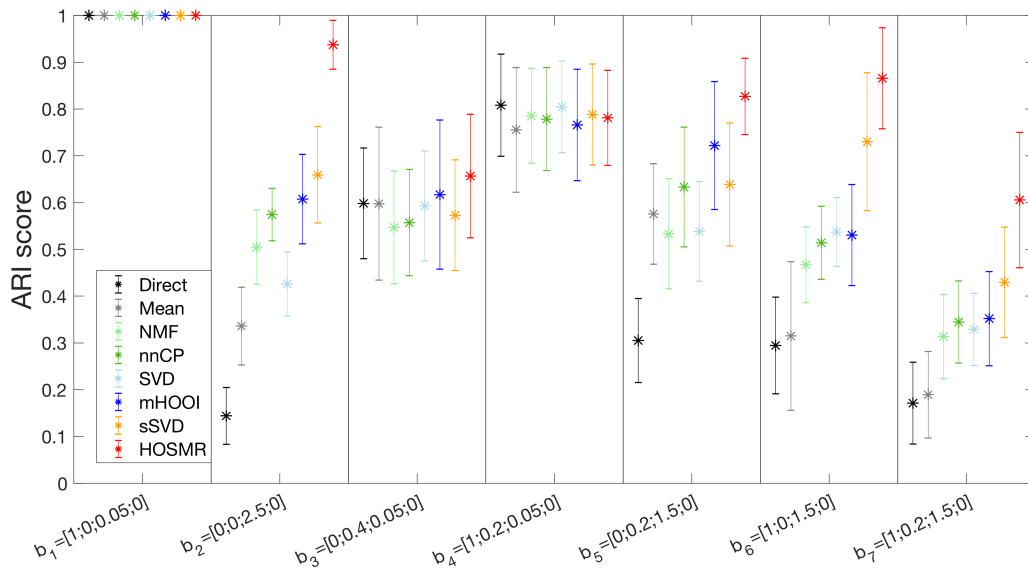


FIGURE 5.4 – ARI score pour chaque méthode, en fonction de 7 modèles associés à 7 vecteurs b_i pour les configurations (1) à (7) du Tableau 5.2.

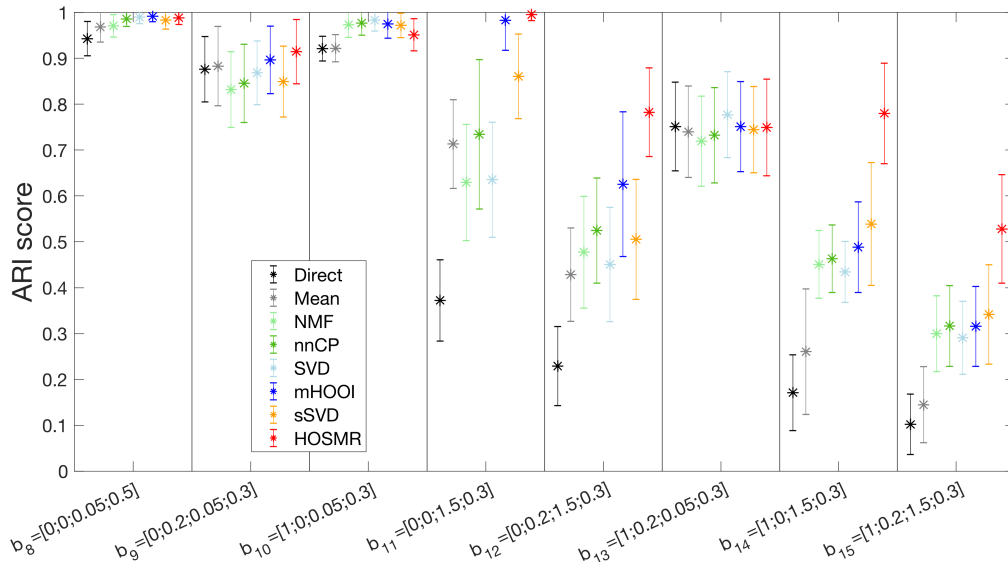


FIGURE 5.5 – ARI score pour chaque méthode, en fonction de 7 modèles associés à 7 vecteurs b_i pour les configurations (8) à (15) du Tableau 5.2.

On isole dans la Figure 5.5 l'impact de la gigue, il s'agit de la seule source d'incertitude pour laquelle la méthode HOSMR n'atteint pas systématiquement les meilleures performances. De ce point de vue, le pire des cas correspond à la combinaison de la gigue plus la durée aléatoire (cas b_{10}), où notre méthode fonctionne seulement mieux que les méthodes directes. Pour le cas b_{13} , les résultats sont similaires à ceux d'autres méthodes de réductions dimensionnelles. Expérimentalement, cela correspond à des situations où nous pourrions obtenir de meilleurs résultats si nous effectuions la décomposition du tenseur \mathcal{X} avec $K_S > 1$. En effet, la complexité générée par la gigue impacte l'activation commune des partitions, et des mélanges de différents groupes de FC peuvent apparaître. Enfin, HOSMR domine à nouveau en présence du bruit de connectivité.

Conclusion : La HOSMR a généralement de meilleures performances (ou des performances équivalentes) aux autres méthodes de réduction de dimensions étudiées ici, sauf dans deux cas. La méthode proposée surpasse les autres méthodes pour les modèles avec bruit de connectivité. Dans l'ensemble, les méthodes parcimonieuses obtiennent toujours de bonnes performances. La contrainte de non-négativité ne semble pas avantageuse ici.

5.5 Application sur des données réelles

5.5.1 Les données sEEG

On considère des données sEEG réelles d'un patient atteint d'une épilepsie focale. Afin de réduire le temps de calcul, 33 électrodes réparties équitablement sur les 108 électrodes de l'enregistrement sont sélectionnées. Les mesures dynamiques de CFs sont calculées via la PLV [192], aucun seuillage n'est appliqué. Après élimination des mesures impactées par les effets de bord, les données forment un tenseur $\mathcal{X} \in \mathbb{R}^{528 \times 96 \times 4}$.

5.5.2 Application

À partir du tenseur \mathcal{X} , on infère la matrice \mathbf{F}_{HOSMR} en utilisant l'Algorithme 10, avec pour paramètres, $K = 4$, $\lambda_1 = 1$, $\lambda_2 = 0.95$ et $\epsilon = 10^{-3}$ fixés empiriquement. Il est habituel de considérer entre 3 et 5 étapes différentes dans une crise d'épilepsie justifiant le choix $K = 4$. Pour les paramètres λ_1 et λ_2 , on recherche la valeur la plus élevée des paramètres pour lesquels les résultats semblent cohérents. Le tracé du haut de la Figure 5.6 montre les profils d'activation temporelle des facteurs de CFs \mathbf{F}_{HOSMR} (correspondant à la matrice \mathbf{V} provenant de l'Algorithme 10). Il y a 4 périodes d'activation, chacune d'elles forme un sous graphe caractéristique qui peut être associé à une des quatre étapes de la crise : "avant crise", "début", "propagation", "fin de crise". L'intervalle de temps aux alentours de 50 secondes est particulièrement intéressant car il ne présente pas d'activation de CFs. Cela correspond probablement à un découplage fonctionnel d'un début de crise [199], une courte période où les activités sEEG dans différentes régions du cerveau sont soudainement décorréliées.

On applique k -means sur les lignes de la matrice \mathbf{F}_{HOSMR} afin d'inférer $P = 5$ réseaux binaires. Comme \mathbf{F}_{HOSMR} est proche d'une matrice indicatrice normalisée (solution \mathbf{A}^* de l'Équation (5.5)), on associe les quatre graphes ayant le moins de liens aux 4 périodes d'activation, tandis que le 5^{ème} graphe correspond aux CFs non synchronisées. Le tracé du bas de la Figure 5.6 montre les positions des 33 électrodes projetées sur le plan transversal selon le repère de Talairach [102]. Les quatre sous-graphes de CFs sont représentés par des liens entre les paires d'électrodes. Chaque sous-graphe peut être associé aux quatre périodes de la crise :

- avant les crises, peu d'électrodes interagissent dans ce qui pourrait correspondre à la zone épileptogène,
- au début des crises, nous voyons une propagation de l'activation des CFs avec l'apparition d'une partition de CFs localisée autour de la zone épileptogène,
- durant la propagation des crises, d'autres CFs apparaissent spontanément dans l'autre hémisphère, en même temps, les CFs de l'hémisphère gauche sont diffusées,
- à la fin des crises, les deux hémisphères interagissent avec l'apparition de CFs communs.

Les graphes obtenus sont en accord avec les résultats cliniques, la zone focale du début de la crise étant proche du sous-graphe "avant crise" et "début". La propagation de l'épilepsie dans l'hémisphère droit du cerveau est bien représentée par les graphes "propagation" et "fin de crise".

5.6 Conclusion

Ce chapitre a présenté différentes méthodes de réduction dimensionnelle avant d'appliquer l'algorithme k -means, dans l'objectif de partitionner des mesures dynamiques de CFs provenant de plusieurs crises d'épilepsie d'un même patient. La procédure proposée permet notamment de décomposer un graphe dynamique (sans seuillage), en plusieurs ensembles de sous graphes caractéristiques. Comme on s'attend à ce que, pour chaque crise, les mêmes ordonnancements de facteurs de CFs s'activent, mais avec des durées différentes, on développe une décomposition tensorielle nommée HOSMR basée sur la la HOOL. Ce dernier est un algorithme déjà existant, qui peut être utilisé pour sélectionner des facteurs de CFs représentatifs de l'ensemble des crises. La HOSMR impose en plus des contraintes structurelles et temporelles pertinentes sur les facteurs de CFs obtenus et leurs profils d'activation. On compare notre méthode sur un modèle de graphe dynamique de CFs pouvant être altéré par

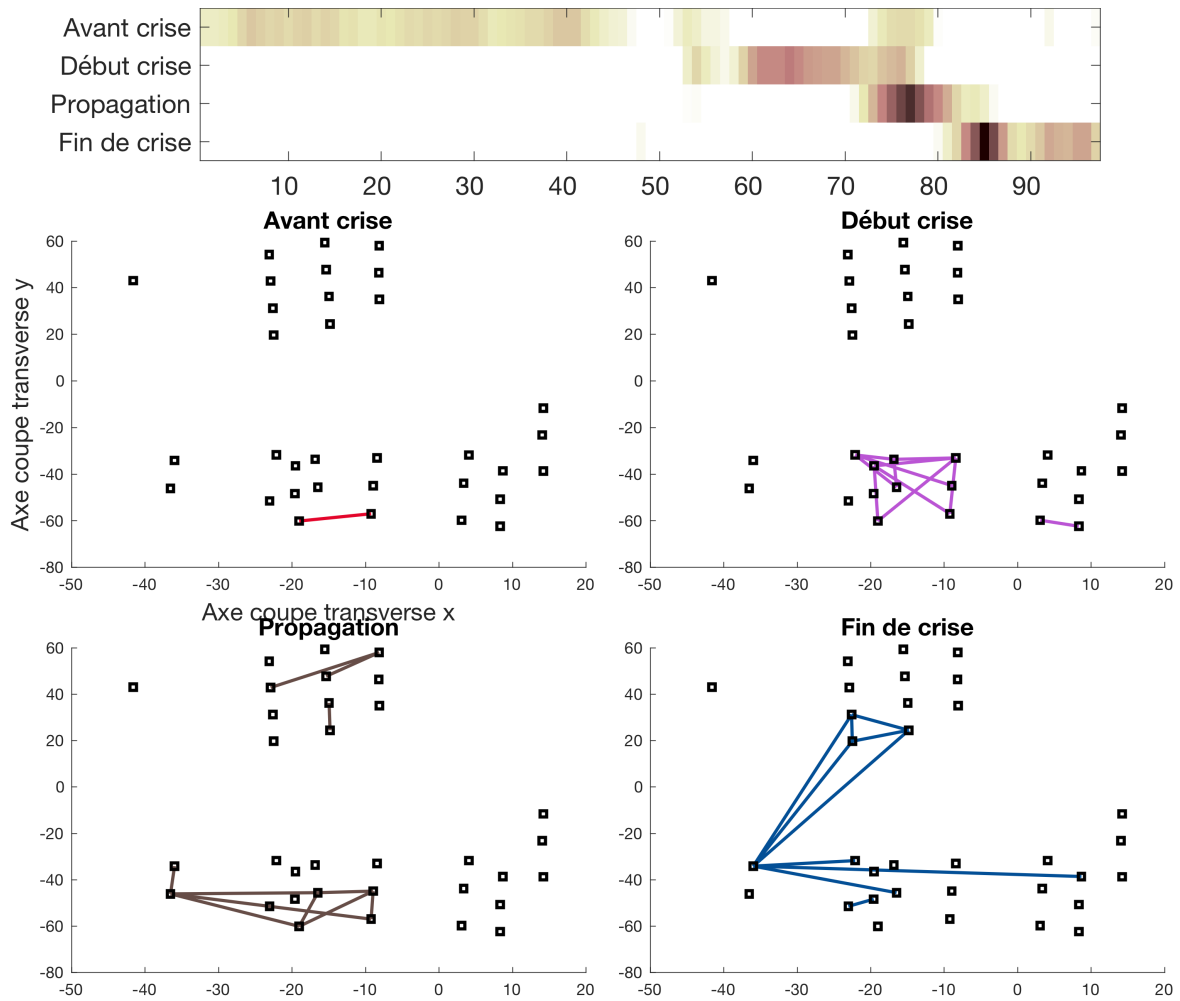


FIGURE 5.6 – (haut) Profils d’activation de la HOSMR (variable \mathbf{V} de l’Algorithme 9). (bas) Partition des CFs correspondant aux 4 étapes d’activation de la crise du patient considéré).

4 types d’incertitudes. Globalement, les méthodes utilisant des décompositions tensorielles et des contraintes de parcimonie obtiennent de meilleures performances, en particulier la HOSMR. Lorsqu’elle est appliquée comme prétraitement des données brutes, cette méthode de réduction dimensionnelle améliore significativement les performances de k -means. La méthode est enfin appliquée sur des données réelles de patient épileptique, permettant d’identifier 4 facteurs de CFs correspondant à 4 périodes différentes d’une crise. Le choix automatique des hyperparamètres est une perspective de ce travail, cela nous permettrait d’analyser davantage la dynamique de la connectivité fonctionnelle dans un ensemble de données cliniques plus riche.

Notons aussi que cette méthode demande de représenter les données sous la forme d’un tenseur, cela conduit à une phase de prétraitement où les données doivent être sélectionnées et centrées temporellement pour que chaque matrice de CFs contienne le même nombre d’échantillons. De plus, il arrive pour certaines crises que les différences de durée de chaque étape soient trop importantes pour pouvoir proprement former le tenseur \mathcal{X} . Enfin, pour certaines crises difficiles, il est possible que, d’une crise à l’autre, les facteurs principaux de CFs ne soient pas systématiquement les mêmes. Afin d’inclure ces variations et de faciliter le prétraitement des données, nous proposons au chapitre suivant des méthodes étendant la HOSMR, dans le cas où les données ne sont plus représentées sous forme d’un tenseur.

Annexe

5.A Comparaison de performances entre les algorithmes HOOI et mHOOI

Afin d'observer empiriquement que la convergence vers la solution optimale de l'algorithme mHOOI est plus rapide que l'algorithme HOOI, nous proposons l'expérience suivante : on considère des modèles de réseaux dynamiques de CFs sous forme de tenseur \mathcal{X} , ce modèle est décrit en Section 5.4.1. Il s'agit de trois facteurs de CFs qui s'activent successivement durant trois étapes, différents bruits sont ajoutés : des variations dans la durée de chaque étape, des erreurs d'activation, un bruit gaussien et des retards temporels sur le début des activations. Ces bruits sont associés à un vecteur \mathbf{b} indiquant l'intensité respective de chacun de ces bruits (plus de détails Section 5.4.1).

On fixe $K = 3$ pour la mHOOI, $K_L = K_T = 3$ et $K_S = 1$ pour la HOOI. On compare les performances de ces deux algorithmes en calculant la variance totale des composantes obtenues à chaque itération. On a donc $tvar(t) = \|\mathbf{G}_i\|_F^2$ pour la HOOI, et $tvar(t) = \|\mathbf{F}_i\|_F^2$ (elles correspondent toutes deux aux matrices contenant toute l'énergie). On calcule aussi le différentiel logarithmique de la variance pour deux itérations consécutives $ldvar(t) = \log(tvar(t) - tvar(t - 1))$. La Figure 5.A.1 montre la moyenne de $tvar(t)$ pour les 100 premières itérations en considérant 100 réalisations du tenseur modélisé \mathcal{X} associé au vecteur de bruit b_i . Un zoom de la dernière itération est affiché ainsi que les premières valeurs de $ldvar(t)$. On remarque souvent un manque de valeurs pour le score $ldvar$ quand la valeur moyenne de $tvar(t) - tvar(t - 1)$ atteint la précision de la machine (et peut devenir négative). Le nouvel algorithme proposé a toujours de meilleures performances que HOOI, avec convergence qui semble exponentielle, sur chaque scénario.

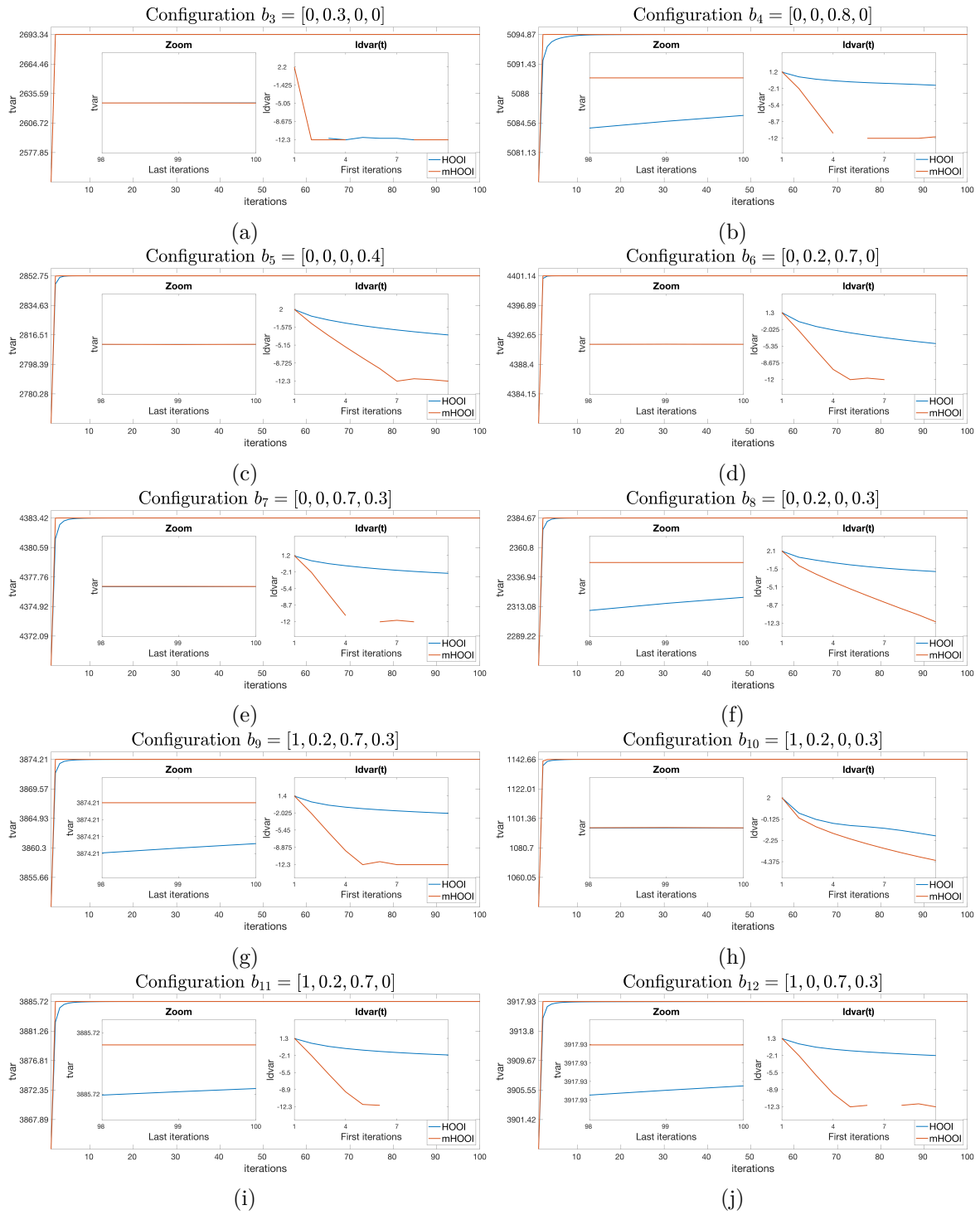


FIGURE 5.A.1 – score $tvar(t)$ pour les 100 premières itérations en utilisant, comme entrée des algorithmes HOOI et mHOOI, le tenseur \mathcal{X} avec le vecteur d'incertitudes \mathbf{b} correspondant. Un zoom est représenté sur les dernières itérations et un autre pour les premières itérations montrant les premières valeurs du score $ldvar(t)$.

5.B Projection efficace sur la contrainte elastic-net

Algorithm 11 Résout les étapes de projection sur la contrainte elastic-net de l'Algorithme 8, reprend l'implémentation de [141]

Require: \bar{y} , K , le paramètre λ_2 .

$U = 1, \dots, T$

$\zeta = 0$

$\rho = 0$

while $U \neq \emptyset$ **do**

 choisir $\tau \in U$ aléatoirement

 Partitionne U tel que :

$$G = \{t \in U \mid |\bar{y}_t| \geq |\bar{y}_\tau|\} \quad (5.64)$$

$$L = \{t \in U \mid |\bar{y}_t| < |\bar{y}_\tau|\} \quad (5.65)$$

$d\rho = |G|$

$d\zeta = \sum_{t \in G} \lambda_2 |\bar{y}_t| + (\bar{y}_t)^2$

if $\zeta + d\zeta - (\rho + d\rho)[\lambda_2 |\bar{y}_\tau| (1 + \frac{|\bar{y}_\tau|}{\lambda_2})] < (1 + 2\frac{|\bar{y}_\tau|}{\lambda_2})^2$ **then**

$\zeta = \zeta + d\zeta$

$\rho = \rho + d\rho$

$U = L$

else

$$U = \{t \in U \mid |\bar{y}_t| > |\bar{y}_\tau|\} \quad (5.66)$$

end if

end while

$$c = \frac{1-s}{r*\lambda_2^2+4}$$

$$\gamma^* = \frac{-1+\sqrt{1-4c}}{2}$$

5.C Le score ARI

On considère deux partitions composées de P classes : la partition qui constitue la vérité terrain notée C^{true} et la partition obtenue expérimentalement notée C^{exp} . Une table de contingence peut être utilisée pour comparer ces deux partitions. Il s'agit de compter le nombre d'éléments communs entre les classes de la partition C^{true} et C^{exp} . Le Tableau 5.C.1 correspond à la table de contingence quand on considère 4 classes par partition

C^{exp} / C^{true}	c_1^{true}	c_2^{true}	c_3^{true}	c_4^{true}	Sommes
c_1^{exp}	m_{11}	m_{12}	m_{13}	m_{14}	$m_{1.}$
c_2^{exp}	m_{21}	m_{22}	m_{23}	m_{24}	$m_{2.}$
c_3^{exp}	m_{31}	m_{32}	m_{33}	m_{34}	$m_{3.}$
c_4^{exp}	m_{41}	m_{42}	m_{43}	m_{44}	$m_{4.}$
Sommes	$m_{.1}$	$m_{.2}$	$m_{.3}$	$m_{.4}$	$m_{..}$

Tableau 5.C.1 – Table de contingence entre les partitions C^{exp} et C^{true} avec $P = 4$

m_{ij} est le nombre de CFs appartenant à la fois à la classe i de la partition C^{exp} et la classe j de la partition C^{true} . $m_{i.} = \sum_{j=1}^P m_{ij}$ est le nombre de CFs dans la classe i de la partition C^{exp} . De même, $m_{.j} = \sum_{i=1}^P m_{ij}$ est le nombre de CFs dans la classe j de la partition C^{true} . Finalement, $m_{..} = \sum_{i=1}^P \sum_{j=1}^P m_{ij} = \sum_{j=1}^P m_{.j} = L$ correspond au nombre total de CFs. Beaucoup de méthodes existent pour tester la similarité de deux partitions grâce à la table de contingence [202], [209]. On

sélectionne le score ARI [99], [169], car c'est un score couramment utilisé, normalisé et facilement interprétable. En considérant le score de performance $\phi = \sum_{i=1}^P \sum_{j=1}^P \binom{m_{ij}}{2}$ qui est maximal quand les deux partitions sont identiques, alors le score ARI est :

$$\phi_{ARI} = \frac{\phi - \mathbb{E}\{\phi\}}{\max(\phi) - \mathbb{E}\{\phi\}}, \quad (5.67)$$

avec $\mathbb{E}\{\phi\}$ la valeur moyenne de ϕ quand l'attribution des CFs dans les 4 classes de C^{exp} est aléatoire, $\max(\phi)$ correspond à la valeur de ϕ quand les deux partitions sont identiques. Ces deux quantités s'écrivent :

$$\max(\phi) = \sum_{i=1}^P \binom{m_{i.}}{2}, \quad \mathbb{E}\{\phi\} = \frac{\sum_{i=1}^P \binom{m_{i.}}{2} \sum_{j=1}^P \binom{m_{.j}}{2}}{\binom{m_{..}}{2}}. \quad (5.68)$$

Chapitre 6

Décomposition simultanée de réseaux dynamiques multi-crisés avec des durées hétérogènes

On considère de nouveau le problème de décomposer des réseaux dynamiques de CFs inférés à partir de plusieurs crises d'un même patient. Dans le chapitre précédent, les durées temporelles de toutes les séries temporelles de CFs pour un patient étant les mêmes, les données pouvaient commodément être représentées sous la forme d'un tenseur \mathcal{X} afin de pouvoir appliquer la méthode de décomposition HOSMR. Cependant, la construction du tenseur \mathcal{X} impose des étapes de prétraitement chronophage, qui sont : l'encadrement de la durée des différentes crises et le centrage des données afin que chaque crise démarre approximativement au même moment dans la fenêtre d'analyse. De plus, lorsque les durées de chaque crise varient beaucoup, la construction du tenseur \mathcal{X} semble arbitraire. Il est alors naturel de penser qu'une méthode plus souple, autorisant à travailler avec des matrices de CFs de durées hétérogènes, permettrait d'obtenir des résultats plus robustes.

Dans ce chapitre, on propose plusieurs nouvelles méthodes qui peuvent être vues comme des alternatives de la HOSMR dans le cas de données avec des durées hétérogènes. Ces méthodes sont présentées en Section 6.1. Elles se basent sur la procédure Parafac2 et la décomposition en matrices non négatives jointes et multivues. En plus de modifier les contraintes et régularisations appliquées à ces méthodes déjà existantes et afin de les rendre pertinentes dans le contexte de l'épilepsie, une nouvelle formulation de la NMF multivues est proposée. Cette réécriture, permet de simplifier l'implémentation de cette décomposition. De plus, elle offre un cadre plus général pour être plus robuste aux crises avec des comportements atypiques. En Section 6.2, les différentes méthodes proposées sont comparées sur un modèle similaire à celui du chapitre précédent mais qui permet en plus de reproduire l'hétérogénéité temporelle des données. Finalement, dans la Section 6.3, on applique les différentes méthodes sur plusieurs crises de deux patients différents. Le premier avec des crises de durées homogènes et le second avec des crises de durées hétérogènes.

6.1 Extensions de la HOsSVD appliquées aux matrices de CFs avec des durées hétérogènes

6.1.1 Modèle Parafac2 sous contraintes

On rappelle que la décomposition Parafac2 consiste à trouver une matrice de facteurs de CFs $\mathbf{F} \in \mathbb{R}^{L \times K}$ et S matrices d'activation temporelle (une pour chaque crise), notées $\mathbf{V}\{s\} \in \mathbb{R}^{T(s) \times K}$, $\forall s \in \{1, \dots, S\}$, ainsi qu'une matrice $\mathbf{W} \in \mathbb{R}^{S \times K}$ qui minimisent [117]

$$\begin{aligned} \operatorname{argmin}_{\mathbf{F}, \mathbf{V}\{1\}, \dots, \mathbf{V}\{S\}, \mathbf{W}} \sum_{s=1}^S \|\mathbf{X}\{s\} - \mathbf{F}\mathbf{D}(\mathbf{W}_{s:})\mathbf{V}\{s\}^t\|_F^2, \\ \text{t.q.} \quad \mathbf{V}\{s\}^t \mathbf{V}\{s\} = \mathbf{C}, \quad \forall s \in \{1, \dots, S\}. \end{aligned} \quad (6.1)$$

avec $D(\mathbf{w})$ la matrice diagonale dont les éléments sont le vecteur \mathbf{w} et \mathbf{C} la matrice de covariance commune des matrices $\mathbf{V}\{s\}$. À partir de cette décomposition, on peut considérer que les facteurs de CFs communs à toutes les crises correspondent à la matrice \mathbf{F} , ces facteurs sont associés aux profils d'activation temporelle contenus dans la matrice $D(\mathbf{W}_{s:})\mathbf{V}\{s\}^t$. Pour la crise s , on rappelle que pour résoudre le problème de l'Équation (6.1), on pose $\mathbf{V}\{s\} = \mathbf{P}\{s\}\mathbf{Q}$ avec $\mathbf{Q}^t\mathbf{Q} = \mathbf{C}$ et $\mathbf{P}^t\{s\}\mathbf{P}\{s\} = \mathbf{I}$ ($\forall s \in \{1, \dots, S\}$) afin d'assurer le respect de la contrainte de Parafac2.

Décomposition Parafac2 avec CP non-négative

La contrainte de Parafac2 de l'Équation (6.1) empêche de contrôler la structure des profils d'activation temporelle. En effet, toute contrainte temporelle appliquée sur ces profils s'exerce sur le produit de trois matrices $D(\mathbf{W}_{s:})\mathbf{P}\{s\}\mathbf{Q}$. Il est connu que la résolution de ce genre de problème est mal contrôlée [4]. On propose alors d'imposer seulement la non négativité au profil d'activation. Comme montré en Section 4.3.3, l'implémentation de Parafac2 consiste à alterner le calcul de S SVD, et une décomposition CP. L'ajout de régularisation sur des problèmes de décomposition CP est encouragé [54]. En effet, cela permet d'assurer qu'une solution existe en plus d'améliorer l'interprétabilité des résultats. On propose alors de remplacer dans l'algorithme Parafac2, le calcul de la CP, par celui d'une CP non-négative décomposition (nnCP). Cette nouvelle décomposition permet de résoudre le problème d'optimisation suivant :

$$\begin{aligned} \underset{\mathbf{F}, \mathbf{Q}, \mathbf{W}, \mathbf{P}(1), \dots, \mathbf{P}(S)}{\operatorname{argmin}} \quad & \sum_{s=1}^S \|\mathbf{X}\{s\} - \mathbf{F}D(\mathbf{W}_{s:})\mathbf{Q}^t\mathbf{P}\{s\}^t\|_F^2, \\ \text{t.q.} \quad & \mathbf{P}\{s\}^t\mathbf{P}\{s\} = \mathbf{I}_K, \quad \forall s \in \{1, \dots, S\}, \\ \text{t.q.} \quad & \mathbf{F} \geq 0, \quad \mathbf{Q} \geq 0, \quad \mathbf{W} \geq 0. \end{aligned} \quad (6.2)$$

Ainsi, la décomposition de l'Équation (6.2) permet d'obtenir des facteurs de CFs non négatifs. Dans le chapitre précédent, la contrainte de non-négativité dans la HOSMR était facultative, car les contraintes de parcimonie imposées aux facteurs de CFs et aux profils d'activation temporelle étaient suffisantes pour obtenir des résultats interprétables. Cependant, dans ce problème où la matrice $\mathbf{V}\{s\}$ ($\forall s \in \{1, \dots, S\}$) n'est pas contrainte, la non-négativité a alors un rôle important pour éviter une configuration trop gloutonne des facteurs de CFs, et donc un résultat difficile à interpréter.

Notons que, à l'itération i , la nnCP est appliquée sur un tenseur composé de S tranches où la tranche s est $\mathbf{X}\{s\}\mathbf{P}_i\{s\}$. Ce tenseur n'est pas forcément non négatif. Les valeurs négatives de ce tenseur ne pourront donc pas être reconstruites. Cependant, comme les matrices de CFs $\mathbf{X}\{s\}$ sont non négatives, on peut supposer que les CFs importantes seront contenues dans les valeurs positives de la matrice $\mathbf{X}\{s\}\mathbf{P}_i\{s\}$. La contrainte de non-négativité appliquée sur ce tenseur ne pose donc pas de problème et permet même de favoriser une certaine parcimonie des facteurs de CFs [65].

Parafac2 parcimonieuse

Une extension au problème de l'Équation (6.2) est maintenant de forcer les facteurs de CFs (\mathbf{F}) à être parcimonieux. On rappelle que la justification de la parcimonie pour notre problème est fournie dans la Définition 13. On propose alors la décomposition suivante :

Proposition 14 : Méthode snnP2

À partir d'une liste de matrices de CFs $\mathbf{X}\{s\} \forall s \in \{1, \dots, S\}$, la décomposition suivante permet d'extraire des facteurs de CFs \mathbf{F} parcimonieux et non négatifs à partir de la méthode Parafac2. On appelle cette méthode la Parafac2 parcimonieuse et (partiellement) non négative (snnP2) et elle consiste à résoudre le problème d'optimisation suivant :

$$\begin{aligned} \underset{\mathbf{F}, \mathbf{Q}, \mathbf{W}, \mathbf{P}(1), \dots, \mathbf{P}(S)}{\operatorname{argmin}} \quad & \sum_{s=1}^S \|\mathbf{X}\{s\} - \mathbf{F}D(\mathbf{W}_{s:})\mathbf{Q}^t\mathbf{P}\{s\}^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1, \\ \text{t.q.} \quad & \mathbf{P}\{s\}^t\mathbf{P}\{s\} = \mathbf{I}_K, \quad \forall s \in \{1, \dots, S\}, \\ \text{t.q.} \quad & \mathbf{F} \geq 0, \quad \mathbf{Q} \geq 0, \quad \mathbf{W} \geq 0. \end{aligned} \quad (6.3)$$

L'algorithme permettant de résoudre le problème de l'Équation (6.3) est fourni en Annexe 6.A.1. On appelle \mathbf{F}_{smnP2} les facteurs de CFs obtenus grâce à cette méthode.

Il est important, aussi, de remarquer qu'ajouter cette contrainte de parcimonie sur les facteurs de CFs, incite fortement la non-négativité des profils d'activation $D(\mathbf{W}_{s:})\mathbf{V}\{s\}$ ($\forall s \in \{1, \dots, S\}$) mais sans la garantir. En effet, on suppose que les valeurs non nulles (et forcément positives) de chaque facteur de CFs parcimonieux \mathbf{f}_k représentent majoritairement des CFs différentes. C'est-à-dire que la matrice de facteur de CFs \mathbf{F} obtenues est quasi orthogonale (c'est une supposition forte qui n'est pas toujours vérifiée dans la pratique). Dans ce cas, afin de satisfaire la fonction de coût de l'Équation (6.3), les profils d'activation $D(\mathbf{W}_{s:})\mathbf{V}\{s\}$ doivent aussi être des valeurs positives.

Contraintes appliquées sur les profils d'activation

L'application de contraintes sur les profils d'activation $D(\mathbf{W}_{s:})\mathbf{V}\{s\}$ ($\forall s \in \{1, \dots, S\}$) est complexe car, $\mathbf{V}\{s\}$ est le produit de deux matrices $\mathbf{P}\{s\}\mathbf{Q}$, avec $\mathbf{P}\{s\}$ qui doit être une matrice orthonormale. Dans [4], les auteurs arrivent à ajouter une contrainte de lissage temporel sur les profils d'activation, mais pas de parcimonie ou de non-négativité. On rappelle qu'à l'itération $i + 1$, en posant $\tilde{\mathbf{U}}\{s\}\tilde{\Sigma}\{s\}\tilde{\mathbf{V}}\{s\}$ la SVD de $\mathbf{X}\{s\}^t\mathbf{F}_iD(\mathbf{W}_{s:})_i\mathbf{Q}_i^t$, on a $\mathbf{P}_{i+1}\{s\} = \tilde{\mathbf{U}}\{s\}\tilde{\mathbf{V}}\{s\}^t$. Alors, afin de privilégier des profils d'activation parcimonieux, les calculs de SVD pourraient être remplacés par des SVDs parcimonieuses avec contrainte d'orthogonalité [204]. Cependant, même en ajoutant en plus une régularisation parcimonieuse sur la matrice \mathbf{Q} , il n'y a aucune garantie de parcimonie sur le produit des trois matrices parcimonieuses \mathbf{Q} , $\tilde{\mathbf{U}}\{s\}$ et $\tilde{\mathbf{V}}\{s\}$. Une autre alternative est d'imposer à la matrice $\mathbf{P}\{s\}$ d'être à la fois non-négative et orthonormale, dans ce cas $\mathbf{P}\{s\}$ serait proche d'une matrice indicatrice normalisée, solution de l'algorithme k -means. Cependant, pour revenir à un problème de partitionnement, il faudrait relâcher la contrainte d'orthonormalité des colonnes de $\mathbf{P}\{s\}$ en une contrainte d'orthogonalité. Ce simple changement n'est malheureusement pas facilement applicable au modèle Parafac2. On précise finalement que la contrainte de Parafac2, qui impose à la matrice $\mathbf{P}\{s\}$ d'être orthonormale, n'est pas une contrainte pertinente pour obtenir des profils d'activation interprétable. Ainsi, afin de mieux contrôler les contraintes appliquées aux profils d'activation, mais aussi afin de relâcher la contrainte d'orthogonalité sur $\mathbf{P}\{s\}$, on s'intéresse désormais à des simplifications du modèle Parafac2. Un des relâchements de Parafac2 est proposé dans [51] afin d'imposer la non négativité à toutes les composantes du problème de l'Équation 6.1. Il s'agit d'une approche intéressante qui pourrait être modifiée pour ajouter les contraintes désirées sur les profils d'activation. Cette approche ne sera pas poursuivie dans ce chapitre. À la place, nous nous intéressons à la décomposition jointe en matrices non-négatives qui peut être vue comme une simplification de la méthode Parafac2.

6.1.2 Décomposition jointe en matrices non-négatives

Une décomposition jointe en matrices non-négatives consiste à trouver les facteurs de CFs $\mathbf{F} \in \mathbb{R}^{L \times K}$ et les matrices d'activation $\mathbf{V}\{s\} \forall s \in \{1, \dots, S\}$, minimisant le problème suivant [196], [125], [118], [106] :

$$\begin{aligned} \underset{\mathbf{F}, \mathbf{V}\{1\}, \dots, \mathbf{V}\{S\}}{\operatorname{argmin}} \quad & \sum_{s=1}^S \zeta(s) \|\mathbf{X}\{s\} - \mathbf{F}\mathbf{V}\{s\}^t\|_F^2, \\ \text{t.q.} \quad & \mathbf{F} \geq 0, \quad \mathbf{V}\{s\} \geq 0 \quad \forall s \in \{1, \dots, S\}. \end{aligned} \quad (6.4)$$

Ce problème peut être vu comme une simplification de Parafac2 dans la mesure où sa contrainte principale est relâchée avec de plus $D(\mathbf{W}_{s:}) = \mathbf{I}$, $\forall s \in \{1, \dots, S\}$. En contrepartie, une contrainte de non-négativité est ajoutée sur chaque matrice $\mathbf{V}\{s\}$ et \mathbf{F} afin assurer l'unicité de la solution obtenue (au facteur d'échelle sur les colonnes près). La constante $\zeta(s)$ permet de pondérer l'importance de chaque crise dans le calcul de la solution. Finalement, on peut ajouter les régularisations et contraintes désirées pour notre problème, c'est-à-dire la parcimonie sur les facteurs de CFs et sur les profils d'activation. On propose à la place de la contrainte elastic-net une contrainte fused lasso (Définition 15). En effet, dans le chapitre précédent, la contrainte était appliquée sur la somme pondérée de différentes crises. Cette somme détériore la localisation temporelle des différentes étapes de la crise générique et il était

donc préférable d'appliquer la contrainte Elastic-net plutôt que fused lasso. Maintenant, les contraintes temporelles sont appliquées séparément sur chaque crise, encourageant le découpage en étapes de la crise et donc la contrainte fused-lasso. Cette contrainte permet de proposer une nouvelle méthode :

Proposition 15 : Méthode joinNMF

La décomposition suivante permet, à partir d'une liste de matrices de CFs $\mathbf{X}\{s\} \forall s \in \{1, \dots, S\}$, d'extraire des facteurs de CFs \mathbf{F} parcimonieux et non négatifs avec des profils d'activation $\mathbf{V}\{s\}$ parcimonieux compacts et non négatifs. Cette décomposition permet d'obtenir des facteurs de CFs communs à l'ensemble des crises, ainsi qu'un profil d'activation $\mathbf{V}\{s\}$ spécifique pour chaque crise. On appelle cette méthode décomposition jointe en matrices non négatives sous contraintes, notée joinNMF elle consiste à résoudre le problème d'optimisation suivant :

$$\begin{aligned} \operatorname{argmin}_{\mathbf{F}, \mathbf{V}\{1\}, \dots, \mathbf{V}\{S\}} & \sum_{s=1}^S \zeta(s) \|\mathbf{X}\{s\} - \mathbf{F}\mathbf{V}\{s\}^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1, \\ \text{t.q.} & \lambda_2(s) \|\mathbf{v}_{:k}\{S\}\|_1 + \lambda_3(s) \|\mathbf{\Gamma}\mathbf{v}_{:k}\{S\}\|_1 + \|\mathbf{v}_{:k}\{S\}\|_F^2 \leq 1, \\ \text{t.q.} & \mathbf{F} \geq 0, \quad \mathbf{V}\{s\} \geq 0, \quad \forall s \in \{1, \dots, S\}. \end{aligned} \quad (6.5)$$

Avec $\mathbf{\Gamma}$ l'opérateur de différence finie, définit dans l'Équation (4.9). L'algorithme permettant de résoudre le problème de l'Équation (6.5) est fourni en Annexe 6.A.2. On précise que la procédure d'accélération FISTA [27] est utilisée pour diminuer le temps de calcul. On appelle $\mathbf{F}_{\text{joinNMF}}$ les facteurs de CFs obtenus grâce à cette méthode.

La constante $\zeta(s)$ permet de pondérer l'importance de chaque crise dans le calcul de la solution. Afin que chaque crise soit prise en compte de la même manière, on fixe par la suite :

$$\zeta(s) = \frac{\|\mathbf{X}\{1\}\|_F^2}{\|\mathbf{X}\{s\}\|_F^2}, \quad (6.6)$$

$\lambda_2(s)$ et $\lambda_3(s)$ sont sélectionnés selon la durée de chaque crise. En effet, si la crise est courte, on s'attend à sélectionner des étapes de crises courtes. Au contraire, si la crise est longue, on autorise l'inférence de périodes d'activation plus longues. On a $\forall s \in \{1, \dots, S\}$:

$$\lambda_2(s) = \frac{T(1)}{T(s)} \lambda_2, \quad (6.7) \quad \lambda_3(s) = \frac{T(1)}{T(s)} \lambda_3. \quad (6.8)$$

L'utilisation de la méthode joinNMF se réduit donc au choix de 3 paramètres : λ_1 et λ_2 contrôlant respectivement la parcimonie des facteurs de CFs et des profils d'activation. Et λ_3 régulant la compacité temporelle des profils d'activation.

6.1.3 Décomposition multivues en matrices non-négatives

La régularisation multivues

Le décomposition jointe en matrices non-négatives permet d'extraire des facteurs de CFs communs à toutes les crises, en prenant en compte des durées hétérogènes pour chaque crise de CFs. Il existe une forme modifiée de la décomposition jointe en matrices non-négatives appelée décomposition en matrice non-négative multivues. Cela consiste à relaxer l'hypothèse que chaque facteur de CFs doit être exactement similaire pour l'ensemble des crises. Ainsi, on cherche pour chaque crise une matrice de facteurs de CFs différentes qui pour la crise s , sera notée $\mathbf{F}_{::s} \in \mathbb{R}^{L \times K}$. Pour ce nouveau problème, on peut choisir le taux de ressemblance entre chaque matrice $\mathbf{F}_{::s}$. Le problème de décomposition en matrices multivues consiste alors à trouver le tenseur $\mathcal{F} \in \mathbb{R}^{L \times K \times S}$ et les matrices d'activation $\mathbf{V}\{s\} \forall s \in \{1, \dots, S\}$, minimisant le problème suivant [136], [109] :

$$\begin{aligned} \operatorname{argmin}_{\mathcal{F}, \bar{\mathbf{F}}, \mathbf{V}\{1\}, \dots, \mathbf{V}\{S\}} & \sum_{s=1}^S \zeta(s) \|\mathbf{X}\{s\} - \mathbf{F}_{::s} \mathbf{V}\{s\}^t\|_F^2 + \delta_s \|\mathbf{F}_{::s} - \bar{\mathbf{F}}\|_F^2, \\ \text{t.q.} & \mathbf{F} \geq 0, \quad \mathbf{V}\{s\} \geq 0 \quad \forall s \in \{1, \dots, S\}. \end{aligned} \quad (6.9)$$

Ici, la matrice $\bar{\mathbf{F}}$ contient les facteurs moyennés à l'ensemble des crises et δ_s est un paramètre autorisant les matrices $\mathbf{F}_{::s}$ à s'écarter plus ou moins de la matrice de facteurs communs. La fonction de régularisation associée aux variables δ_s est appelée régularisation multivues. Il s'agit, lorsque les variables δ_s sont élevées, d'une alternative pour réaliser une décomposition jointe en matrices non-négatives. L'avantage de cette décomposition est de permettre à des crises avec un comportement singulier, d'avoir des profils de CFs légèrement spécifiques, permettant d'obtenir pour les autres crises des facteurs de CFs plus robustes. Le facteur de CFs communs à toutes les crises peut être obtenu en moyennant les facteurs de CFs obtenus pour chaque crise.

Interprétation sous forme de régularisation de Thikonov avec opérateur laplacien

Nous montrons dans cette section, que la régularisation multivues de l'Équation (6.9) peut être modifiée pour permettre d'explicitier les relations de similarité paire à paire entre chaque crise. Dans un premier temps, on peut noter que la matrice de facteurs communs $\bar{\mathbf{F}}$, lorsque l'on connaît le tenseur \mathcal{F} , se calcule ainsi :

$$\bar{\mathbf{F}} = \frac{\sum_{s=1}^S \delta_s \mathbf{F}_{::s}}{\bar{\delta}}, \quad \text{avec} \quad \bar{\delta} = \sum_{s=1}^S \delta_s. \quad (6.10)$$

À partir de cette observation, on peut montrer que :

$$\sum_{s=1}^S \delta_s \|\mathbf{F}_{::s} - \bar{\mathbf{F}}\|_F^2 = \sum_{s=1}^S \delta_s \text{Tr}(\mathbf{F}_{::s} \mathbf{F}_{::s}^t) - 2 \sum_{s=1}^S \sum_{s'=1}^S \frac{\delta_s \delta_{s'}}{\bar{\delta}} \text{Tr}(\mathbf{F}_{::s} \mathbf{F}_{::s'}^t) + \sum_{s=1}^S \delta_s \text{Tr}(\bar{\mathbf{F}} \bar{\mathbf{F}}^t), \quad (6.11)$$

avec :

$$\begin{aligned} \sum_{s=1}^S \delta_s \text{Tr}(\bar{\mathbf{F}} \bar{\mathbf{F}}^t) &= \bar{\delta} \text{Tr}(\bar{\mathbf{F}} \bar{\mathbf{F}}^t), \\ &= \sum_{s=1}^S \sum_{s'=1}^S \frac{\delta_s \delta_{s'}}{\bar{\delta}} \text{Tr}(\mathbf{F}_{::s} \mathbf{F}_{::s'}^t). \end{aligned} \quad (6.12)$$

Soit :

$$\begin{aligned} \sum_{s=1}^S \delta_s \|\mathbf{F}_{::s} - \bar{\mathbf{F}}\|_F^2 &= \sum_{s=1}^S \delta_s \text{Tr}(\mathbf{F}_{::s} \mathbf{F}_{::s}^t) - \sum_{s=1}^S \sum_{s'=1}^S \frac{\delta_s \delta_{s'}}{\bar{\delta}} \text{Tr}(\mathbf{F}_{::s} \mathbf{F}_{::s'}^t), \\ &= \sum_{s=1}^S \left(\frac{\delta_s (\bar{\delta} - \delta_s)}{\bar{\delta}} \right) \text{Tr}(\mathbf{F}_{::s} \mathbf{F}_{::s}^t) - \sum_{s, s' \neq s} \frac{\delta_s \delta_{s'}}{\bar{\delta}} \text{Tr}(\mathbf{F}_{::s} \mathbf{F}_{::s'}^t). \end{aligned} \quad (6.13)$$

En considérant la matrice $\mathbf{F}_{(S)} \in \mathbb{R}^{S \times LT}$ comme le matricage du tenseur \mathcal{F} selon le mode des crises :

$$\mathbf{F}_{(S)} = \begin{pmatrix} \text{vec}(\mathbf{F}_{::1}) \\ \vdots \\ \text{vec}(\mathbf{F}_{::S}) \end{pmatrix}, \quad (6.14)$$

on a :

$$\text{Tr}(\mathbf{F}_{::s} \mathbf{F}_{::s'}^t) = \text{vec}(\mathbf{F}_{::s})^t \text{vec}(\mathbf{F}_{::s'}) = \mathbf{F}_{(S)s} \mathbf{F}_{(S)s'}^t. \quad (6.15)$$

Ainsi, cela permet l'écriture simplifiée de l'Équation (6.13) telle que :

$$\sum_{s=1}^S \delta_s \|\mathbf{F}_{::s} - \bar{\mathbf{F}}\|_F^2 = \text{Tr}(\mathbf{F}_{(S)}^t \mathbf{\Delta} \mathbf{F}_{(S)}), \quad (6.16)$$

avec :

$$\mathbf{\Delta} = \frac{1}{\bar{\delta}} \begin{pmatrix} \delta_1(\bar{\delta} - \delta_1) & -\delta_1\delta_2 & \dots & -\delta_1\delta_S \\ -\delta_2\delta_1 & \delta_2(\bar{\delta} - \delta_2) & & \vdots \\ \vdots & & \ddots & \vdots \\ -\delta_S\delta_1 & \dots & \dots & \delta_S(\bar{\delta} - \delta_S) \end{pmatrix}. \quad (6.17)$$

On peut facilement remarquer que la matrice $\mathbf{\Delta}$ a une structure de matrice laplacienne [194], en effet :

$$\mathbf{\Delta}_{ss'} = -\frac{\delta_s\delta_{s'}}{\bar{\delta}} \leq 0, \quad \forall s' \neq s, \quad (6.18)$$

$$\mathbf{\Delta}_{ss} = -\sum_{s' \neq s}^S \mathbf{\Delta}_{ss'} = \frac{\delta_s(\bar{\delta} - \delta_s)}{\bar{\delta}}. \quad (6.19)$$

Ainsi l'Équation (6.18) montre que les éléments non diagonaux sont négatifs, l'Équation (6.19) montre que les éléments diagonaux correspondent à des degrés de graphes [194]. Cela implique que la somme des lignes et des colonnes de la matrice $\mathbf{\Delta}$ est nulle. À partir de cette observation, on peut remarquer que la régularisation multivues correspond à une régularisation de Thikonov avec un opérateur basé sur un Laplacien. On obtient alors :

$$\sum_{s=1}^S \delta_s \|\mathbf{F}_{::s} - \bar{\mathbf{F}}\|_F^2 = \|\mathbf{\Delta}^{\frac{1}{2}} \mathbf{F}_{(S)}\|_F^2. \quad (6.20)$$

Avec la structure particulière de l'Équation (6.17), on peut légèrement modifier le critère de régularisation multivues en considérant une matrice laplacienne quelconque qui correspondrait à un graphe de similarité entre les différentes crises d'un même patient. On propose alors la réécriture suivante de la NMF multivues :

Proposition 16

On considère une matrice laplacienne $\mathbf{\Delta}$ indiquant le niveau de similarité entre chaque paire de crise. La décomposition en matrice non-négative multivues modifiée correspond alors à :

$$\begin{aligned} \underset{\mathcal{F}, \mathbf{V}\{1\}, \dots, \mathbf{V}\{S\}}{\operatorname{argmin}} \quad & \sum_{s=1}^S \zeta(s) \|\mathbf{X}\{s\} - \mathbf{F}_{::s} \mathbf{V}\{s\}^t\|_F^2 + \eta \|\mathbf{\Delta}^{\frac{1}{2}} \mathbf{F}_{(S)}\|_F^2, \\ \text{t.q.} \quad & \mathbf{F} \geq 0, \quad \mathbf{V}\{s\} \geq 0 \quad \forall s \in \{1, \dots, S\}. \end{aligned} \quad (6.21)$$

Avec η paramètre associé à la régularisation multivues modifiée.

Comme montré en Annexe 6.B, la régularisation multivues modifiée correspond au critère de l'analyse canonique des corrélations [183]. On peut donc voir cette régularisation comme une méthode pour trouver une transformation linéaire des données $\mathbf{X}\{s\}$ (correspondant aux matrices $\mathbf{V}\{s\}$) qui maximise la corrélation paire à paire entre les crises d'un patient.

Cette décomposition ne doit pas être confondue avec des décompositions en matrices non négatives jointes avec régularisation laplacienne déjà existantes [138], [164], [106]. En effet, dans ces derniers cas, la régularisation laplacienne est utilisée pour lisser les facteurs de CFs à partir d'un a priori sur la structure de similarité entre eux. La régularisation n'est donc pas appliquée sur le mode des crises mais sur le mode des CFs.

Finalement, précisons que la contrainte multivues, ainsi que sa formulation sous forme de régularisation de Thikonov avec une matrice laplacienne quelconque, sont des contraintes convexes mais pas strictement convexes. En effet, une matrice laplacienne est semi-définie positive. Elle contient donc forcément au moins une valeur propre nulle. Pour éviter des problèmes de convergences lors de l'implémentation de ce problème, on peut alors rendre la matrice $\mathbf{\Delta}$ définie positive en l'additionnant à une matrice identité telle que $\mathbf{\Delta} \leftarrow \mathbf{\Delta} + \epsilon \mathbf{I}_S$, avec ϵ faible.

La décomposition en matrice non-négative multivues modifiée sous contraintes

On peut adapter la décomposition multivues en matrices non-négatives modifiées à notre problème en rajoutant les mêmes contraintes et régularisations que pour la joinNMF de l'Équation (6.5).

Proposition 17 : Méthode graphNMF et multNMF

La décomposition suivante permet, à partir d'une liste de matrices de CFs $\mathbf{X}\{s\} \forall s \in \{1, \dots, S\}$, d'extraire des facteurs de CFs $\mathbf{F}_{::s}$ parcimonieux et non négatifs avec des profils d'activation $\mathbf{V}\{s\}$ parcimonieux compacts et non négatifs. Cette décomposition permet d'obtenir des facteurs de CFs ainsi que des profils d'activation $\mathbf{V}\{s\}$ spécifiques pour chaque crise. Le niveau de similarité entre les facteurs de CFs $\mathbf{F}_{::s}$ est lié à la matrice Δ indiquant le niveau de similarité entre chaque paire de crise. On recherche alors le tenseur $\mathcal{F} \in \mathbb{R}^{L \times K \times S}$ et les matrices d'activation $\mathbf{V}\{s\} \forall s \in \{1, \dots, S\}$, minimisant le problème suivant :

$$\begin{aligned} \underset{\mathcal{F}, \mathbf{V}\{1\}, \dots, \mathbf{V}\{S\}}{\operatorname{argmin}} \quad & \sum_{s=1}^S \zeta(s) \|\mathbf{X}\{s\} - \mathbf{F}_{::s} \mathbf{V}\{s\}^t\|_F^2 + \eta \|\Delta^{\frac{1}{2}} \mathbf{F}_{(S)}\|_F^2 + \lambda_1 \|\mathbf{F}_{(S)}\|_1, \\ \text{t.q.} \quad & \lambda_2(s) \|\mathbf{v}_{:k}\{s\}\|_1 + \lambda_3(s) \|\mathbf{\Gamma} \mathbf{v}_{:k}\{s\}\|_1 + \|\mathbf{v}_{:k}\{s\}\|_F^2 \leq 1, \quad \forall s \in \{1, \dots, S\}, \\ \text{t.q.} \quad & \mathbf{F} \geq 0, \quad \mathbf{V}\{s\} \geq 0, \quad \forall s \in \{1, \dots, S\}. \end{aligned} \quad (6.22)$$

Avec $\mathbf{\Gamma}$ défini dans l'Équation (4.9). L'algorithme permettant de résoudre le problème de l'Équation (6.22) est fourni en Annexe 6.A.3. Alors :

- Si $\eta > 0$, et si Δ est le Laplacien d'un graphe binaire complet ($\delta_{ss'} = -1 \forall s \neq s'$), les tranches $\mathbf{F}_{::s}$ du tenseur \mathcal{F} seront d'autant plus similaires que η sera élevé. On appelle cette procédure **graphNMF**, produisant le facteur de CFs $\mathbf{F}_{\text{graphNMF}} = \frac{1}{S} \sum_{s=1}^S \mathbf{F}_{::s}$.
- Si $\eta = 0$, alors cela revient à appliquer S NMF parcimonieuses sous contraintes indépendantes. Dans ce cas il s'agit d'une décomposition en NMF multiples, nommée **multNMF**, produisant un tenseur de facteurs de CFs $\mathcal{F}_{\text{multNMF}}$.

On précise que les paramètres dépendants de la crise s qui sont $\zeta(s)$, $\lambda_2(s)$ et $\lambda_3(s)$ sont sélectionnés de la même manière que pour la méthode joinNMF, correspondant respectivement aux Équations (6.6), (6.7) et (6.8). Cette décomposition est donc associée à 4 paramètres : λ_1 , λ_2 et λ_3 communs à la méthode joinNMF ainsi que η , le paramètre de régularisation multivues modifiée.

Dans un cadre plus général où une matrice laplacienne quelconque est utilisée, une investigation plus poussée peut être réalisée afin de pouvoir former des groupes de crises qui ont des comportements similaires. Le taux de similarité entre deux facteurs de CFs $\mathbf{F}_{::s}$ et $\mathbf{F}_{::s'}$ sera alors proportionnel au poids $-\delta_{ss'}$ de la matrice laplacienne. Il faut donc avant tout trouver une manière pertinente de construire la matrice laplacienne Δ ; c'est-à-dire trouver des mesures permettant de calculer les similarités entre deux crises. Cela peut revenir à calculer les similarités entre deux matrices de CFs avec des durées hétérogènes (en utilisant par exemple la distance de Grassmann [206]). Ce cadre-là ne sera pas étudié dans ce chapitre.

6.2 Comparaison de différentes méthodes sur un modèle

Dans cette section, nous évaluons les performances des différentes procédures présentées précédemment pour estimer des facteurs de CFs. On utilise un modèle semblable à celui du chapitre précédent, c'est-à-dire un modèle de CFs de signaux sEEG avec 4 sources d'incertitudes. On précise en Section 6.2.1 les détails de ce nouveau modèle. L'expérience réalisée pour comparer chaque méthode est explicitée dans la Section 6.2.2, les résultats sont discutés en Section 6.2.3.

6.2.1 Un modèle de graphe multiplex avec durée hétérogène pour les FCs

On applique de nouveau le modèle de la Section 5.4, avec les mêmes paramètres d'incertitude $[\alpha, \beta, \sigma, \delta]$, correspondant respectivement à l'incertitude sur les durées aléatoires de chaque étape, les

erreurs d'activation, le bruit de connectivité et la gigue. On fixe comme précédemment $L = 66$, $S = 4$. Cependant, le nombre d'échantillons temporels est réduit à $T = 200$, à la fois pour que l'expérience soit moins coûteuse en temps de calcul, mais aussi en remarquant que la plupart des crises durent entre 80 et 300 secondes.

Afin de prendre en compte l'hétérogénéité temporelle des données, on commence par générer le tenseur $\mathcal{X} \in \mathbb{R}^{L \times T \times S}$ obtenu comme une réalisation du modèle de la Section 5.4 avec les paramètres d'incertitudes $[\alpha, \beta, \sigma, \delta]$. Ce tenseur est ensuite converti en une série de matrices de CF $\mathbf{X}\{s\} \in \mathbb{R}^{L \times T(s)}$, $\forall s \in \{1, \dots, S\}$. Ici $T(s)$ varie entre 200 et 300, on a $T(s) = T + \tau(s)$ avec $\tau(s)$ provenant du tirage aléatoire d'une loi uniforme discrète entre 0 et 100. Le début de la crise est aussi choisi aléatoirement en considérant deux matrices de bruit $\mathbf{B}^{(1)}$ et $\mathbf{B}^{(2)}$ qui correspondent respectivement à des enregistrements supplémentaires avant et après la crise. Chaque élément de ces matrices est une réalisation aléatoire d'un bruit blanc gaussien de puissance σ et de moyenne 0.2 (correspondant à la valeur moyenne des CFs non actives). On a alors :

$$\mathbf{X}\{s\} = [\mathbf{B}^{(1)}, \mathbf{X}_{::s}, \mathbf{B}^{(2)}]. \quad (6.23)$$

Avec $\mathbf{B}^{(1)} \in \mathbb{R}^{L \times \tau'(s)}$ et $\mathbf{B}^{(2)} \in \mathbb{R}^{L \times (\tau(s) - \tau'(s))}$ et $\tau'(s)$ provenant d'une loi uniforme discrète entre 0 et $\tau(s)$.

La Figure 6.1 présente quatre matrices de CFs $\mathbf{X}\{s\} \in \mathbb{R}^{L \times T(s)}$, $\forall s \in \{1, \dots, 4\}$ correspondant au choix particulier $[\alpha = 1, \beta = 0, \sigma = 0.05, \delta = 0]$.

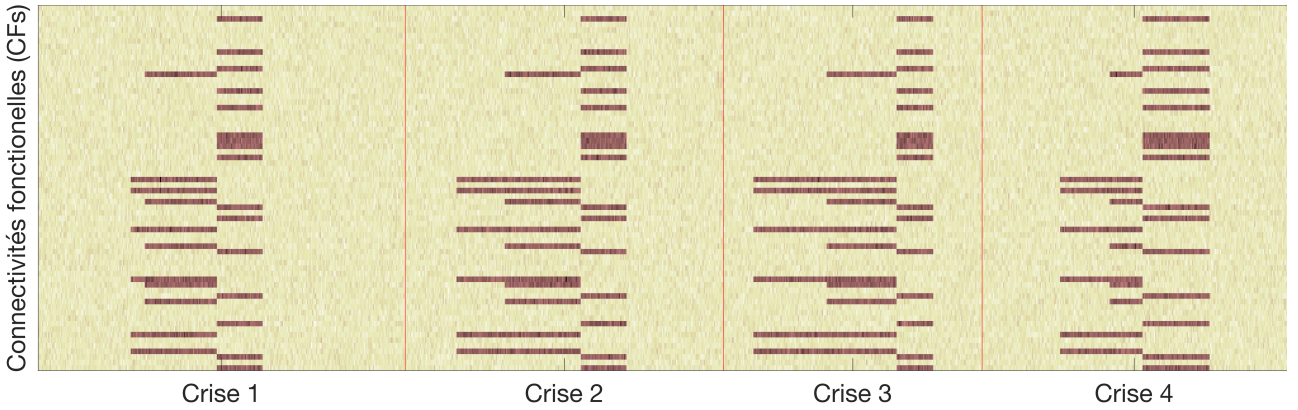


FIGURE 6.1 – Représentation de la concaténation de quatre matrices de CFs $\mathbf{X}\{s\} \in \mathbb{R}^{L \times T(s)}$, $\forall s \in \{1, \dots, 4\}$ correspondant au choix particulier $[\alpha = 1, \beta = 0, \sigma = 0.05, \delta = 0]$.

6.2.2 Description de l'expérience

Grâce au nouveau modèle, on peut maintenant comparer les différentes méthodes décrites dans la Section 6.1.1. Le Tableau 6.1 résume les 5 méthodes comparées. Comme repère, on applique la procédure HOSMR sur le tenseur \mathcal{X} utilisé comme base pour construire les matrices de CFs $\mathbf{X}\{s\}$, $\forall s \in \{1, \dots, S\}$. Les 4 autres méthodes sont la snnP2 correspondant à la Proposition 14, multNMF correspondant à la Proposition 15, joinNMF et la graphNMF correspondant à la Proposition 17.

Pour de nombreuses configurations des paramètres $(\lambda_1, \lambda_2, \lambda_3)$, les résultats obtenus sont similaires. On propose la démarche suivante pour simplifier la sélection de paramètres sans compromettre les performances du modèle. Les paramètres λ_1 et λ_2 , qui correspondent au niveau de parcimonie des composantes obtenues sont les plus critiques à sélectionner. On propose de les associer, on a ainsi $\lambda_1 = S\lambda_2$ (avec S le nombre de crise). On fait alors varier seulement le paramètre λ_2 de 0 à 1 par pas de 0.05. Cela permet de diminuer la complexité du modèle tout en permettant d'étudier un éventail assez complet des niveaux de parcimonie possible. Le paramètre λ_3 correspondant au niveau de cohérence temporelle souhaitée pour les profils d'activation, s'avère être un paramètre moins déterminant. Il est fixé à $\lambda_3 = 2$, permettant d'obtenir pour chaque expérience des profils d'activation contraints temporellement.

On cherche à déterminer quelle décomposition permet d'obtenir les facteurs de CFs ressemblant à la vérité terrain (Figure 6.2a). On considère alors seulement des matrices \mathbf{F} de dimensions $K = 3$ ou

Méthode	Section	Paramètres	Temps de calculs
1 - HOSMR	5.3.2	$\lambda_1 = \lambda_2, \lambda_2 \in \{0, 0.05, 0.1, \dots, 1\}$	0.05 (s)
2 - snnP2	6.1.1	$(\lambda_1/S) \in \{0, 0.05, 0.1, \dots, 1\}$	0.06 (s)
3 - MultNMF	6.1.3	$\eta = 0, \lambda_1 = \lambda_2, \lambda_2 \in \{0, 0.05, 0.1, \dots, 1\}$	1.1 (s)
4 - JoinNMF	6.1.2	$\lambda_1 = S\lambda_2, \lambda_2 \in \{0, 0.05, 0.1, \dots, 1\}, \lambda_3 = 2$	0.8 (s)
5 - GraphNMF	6.1.3	$\eta = 1, \lambda_1 = S\lambda_2, \lambda_2 \in \{0, 0.05, 0.1, \dots, 1\}, \lambda_3 = 2$	2.6 (s)

Tableau 6.1 – Liste des méthodes comparées afin de convertir les matrices de CFs $\mathbf{X}\{s\}, \forall s \in \{1, \dots, S\}$ en une matrice de facteur de CFs \mathbf{F} . Le temps de calcul correspond à la durée moyenne nécessaire à chaque méthode pour calculer les facteurs de CFs de toutes les expériences proposées.

$K = 4$. Les Figures 6.2b et 6.2c représentent respectivement un exemple de matrice \mathbf{F} obtenue avec les 3 facteurs de CFs visibles pour $K = 3$ et $K = 4$ (le quatrième facteur de la matrice Figure 6.2c contient principalement les CFs inactives).

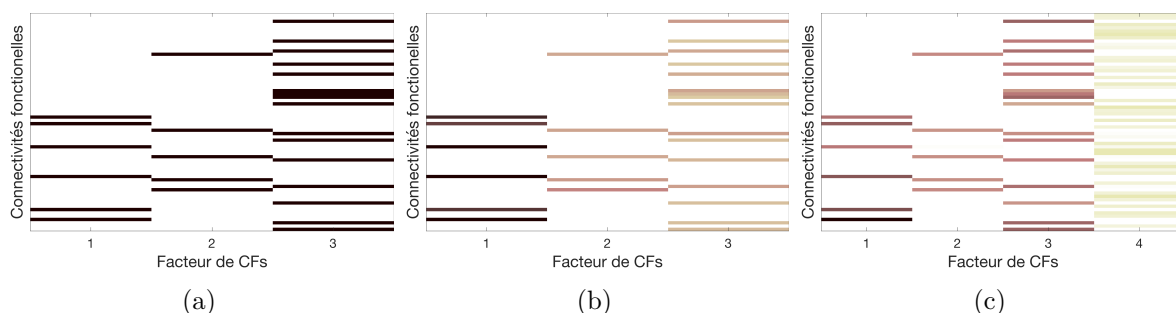


FIGURE 6.2 – (a) représentation des trois facteurs de CFs organisés selon 3 classes, (b)-(c) un exemple de matrice \mathbf{F} contenant les facteurs de CFs estimés respectivement pour $K = 3$ ou $K = 4$.

Afin de quantifier les performances de chaque méthode, on ne propose pas le même score que dans le chapitre précédent. En effet, dans le chapitre 5, la décomposition tensorielle était avant tout utilisée comme outil de réduction dimensionnelle avant d'appliquer k -means. Dans ce chapitre, on cherche à quantifier comment les décompositions proposées fournissent directement des résultats pertinents et interprétables.

Afin de quantifier les performances de chaque méthode, on applique la procédure suivante. Chaque colonne de la matrice de facteur de CFs est normalisée par son élément maximal, ce qui donne :

$$\hat{\mathbf{f}}_{:k} = \frac{\mathbf{f}_{:k}}{\max(\mathbf{f}_{:k})}, \quad \forall k \in 1, \dots, K. \quad (6.24)$$

Ainsi, chaque élément de la matrice de facteur sera compris entre 0 et 1. On considère la matrice contenant les facteurs de CFs idéaux de la Figure 6.2a, notée $\bar{\mathbf{F}} \in \mathbb{R}^{L \times 3}$. Les angles entre chaque colonne k de la matrice de facteurs estimés $\hat{\mathbf{F}}$ et chaque colonne k' de la matrice de facteurs idéaux sont calculées. Cette mesure angulaire se note $\rho_{kk'}$ et s'écrit :

$$\rho_{kk'} = \frac{\hat{\mathbf{f}}_{:k}^t \bar{\mathbf{f}}_{:k'}}{\sqrt{\hat{\mathbf{f}}_{:k}^t \hat{\mathbf{f}}_{:k} \bar{\mathbf{f}}_{:k'}^t \bar{\mathbf{f}}_{:k'}}}. \quad (6.25)$$

Chaque score $\rho_{kk'}$ est compris entre 0 (si les deux facteurs sont orthogonaux) à 1 (s'ils sont identiques). Finalement, pour chaque facteur idéal $\bar{\mathbf{f}}_{:k'}$, on détermine lequel des facteurs estimés $\hat{\mathbf{f}}_{:k}$ semble le mieux lui correspondre. Alors la moyenne des mesure angulaire entre chaque facteur idéal et sa meilleure estimation parmi les colonnes de la matrice $\hat{\mathbf{F}}$ est calculée, formant le score $\rho^{(+)}$ tel que :

$$\rho^{(+)} = \frac{1}{3} \sum_{k'=1}^3 \max(\rho_{:k'}). \quad (6.26)$$

De la même manière que pour le score ARI [99], on normalise le score proposé pour qu'il soit compris entre 0 et 1. Avec 1 indiquant une correspondance parfaite entre les facteurs estimés et idéaux, et 0 si la correspondance ne surpasse pas une solution $\hat{\mathbf{F}}$ où chaque élément est issu d'un tirage aléatoire selon une loi uniforme comprise entre 0 et 1. Le score $\rho^{(+)}$ normalisé est noté $\rho^{(*)}$ et correspond à :

$$\rho^{(*)} = \frac{\rho^{(+)} - \mathbb{E}\{\rho^{(+)}\}}{1 - \mathbb{E}\{\rho^{(+)}\}}. \quad (6.27)$$

$\mathbb{E}\{\rho^{(+)}\}$ est calculé empiriquement. Précisons que pour le tenseur $\mathcal{F}_{multNMF}$, $\rho^{(*)}$ est calculé pour chaque tranche $\mathbf{F}_{:,s}$, le score final étant la moyenne des scores $\rho^{(*)}$ obtenus.

6.2.3 Résultats et discussions

Sous ces conditions expérimentales, on évalue chaque méthode pour différentes configurations d'incertitudes. On présente les résultats pour différents niveaux de bruits de connectivités (qui peuvent être faibles $\sigma = 0.05$, moyens $\sigma = 0.5$ ou forts $\sigma = 1.5$). De nouveau, les résultats sont présentés sous forme de boîte à moustaches indiquant la moyenne et l'écart-type de 30 réalisations indépendantes des matrices de CFs $\mathbf{X}\{s\}$, $\forall s \in \{1, \dots, S\}$.

Bruit de connectivité faible

La Figure 6.3a présente les performances de chaque méthode lorsqu'il n'y a pas d'incertitude sur la durée de chaque étape ($\alpha = 0$). Lorsque les erreurs d'activation et la gigue sont actives, elles valent respectivement $\beta = 0.2$ et $\delta = 0.3$. On observe que la HOSMR domine chaque méthode pour ces configurations. A l'inverse, la méthode multNMF est celle qui fournit les plus mauvais résultats. La Figure 6.3b présente les performances de chaque méthode lorsque l'incertitude sur la durée de chaque étape est présente ($\alpha = 1$). Dans ce cas, la méthode HOSMR ne domine plus les autres procédures et a des performances similaires aux méthodes snnP2, graphNMF et joinNMF. Ce résultat montre que pour des crises avec des étapes de durées hétérogènes, ces trois méthodes peuvent être utilisées comme alternatives à la HOSMR.

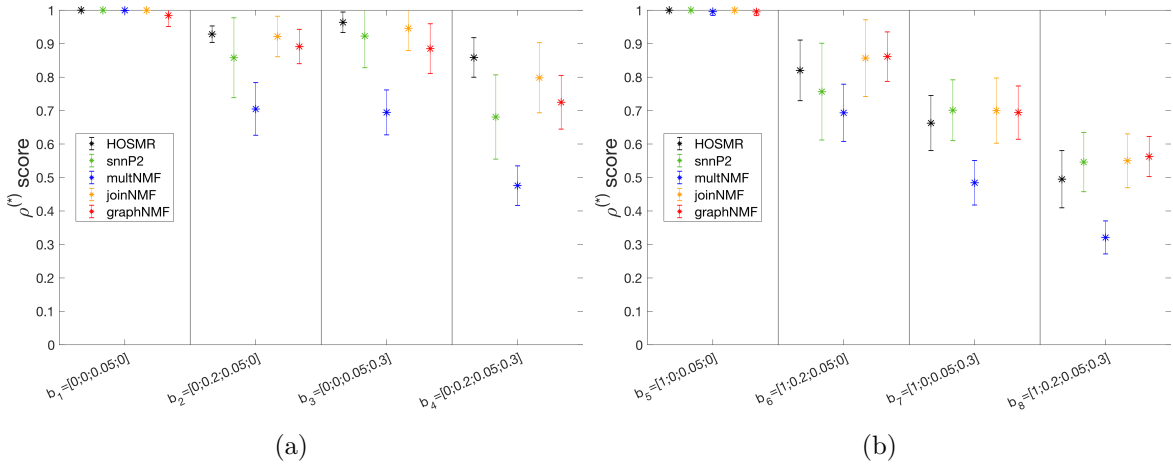


FIGURE 6.3 – Score $\rho^{(*)}$ atteint par les 5 méthodes proposées pour 4 conditions expérimentales avec $\sigma = 0.05$ et (a) $\alpha = 0$, (b) $\alpha = 1$

Bruit de connectivité moyen

La Figure 6.4a présente les performances de chaque méthode lorsqu'il n'y a pas d'incertitude sur la durée de chaque étape ($\alpha = 0$), ici le bruit de connectivité est moyen soit $\sigma = 0.5$. On observe des résultats similaires aux cas où le niveau de bruit est faible, c'est-à-dire une légère domination de la méthode HOSMR. La méthode multNMF fournit de nouveau les plus mauvais résultats montrant l'importance d'estimer des facteurs de CFs communs à l'ensemble des crises. La Figure 6.4b présente les performances de chaque méthode lorsque l'incertitude sur la durée de chaque étape est présente

($\alpha = 1$). Encore une fois, la HOSMR ne domine plus les autres méthodes, mais ce sont seulement les méthodes joinNMF et graphNMF qui fournissent les meilleurs résultats. En Comparaison du modèle avec bruit faible, les performances de la méthode snnP2 se dégradent, montrant l'importance de la contrainte fused-lasso dans ce cas.

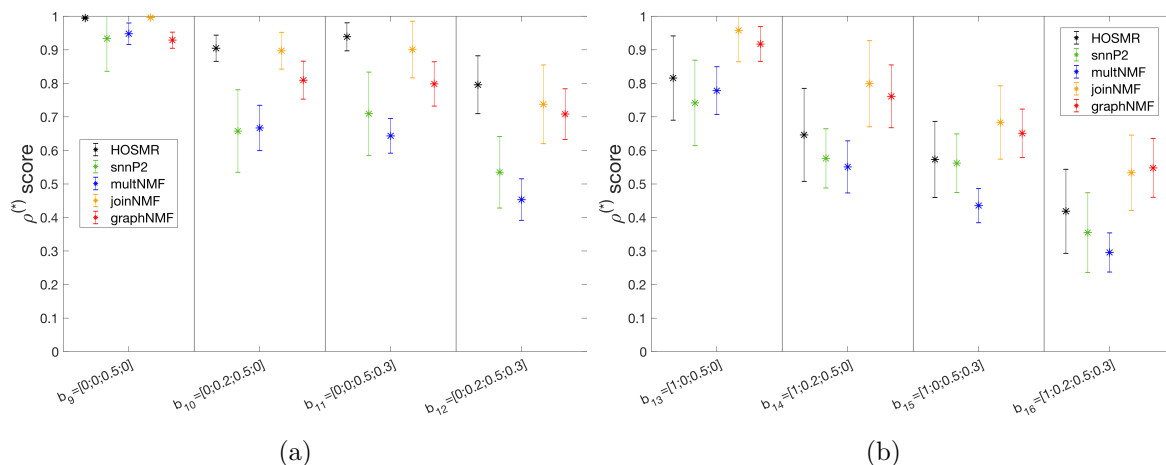


FIGURE 6.4 – Score $\rho^{(*)}$ atteint par les 5 méthodes proposées pour 4 conditions expérimentales avec $\sigma = 0.5$ et (a) $\alpha = 0$, (b) $\alpha = 1$

Bruit de connectivité fort

La Figure 6.5a présente les performances de chaque méthode lorsqu'il n'y a pas d'incertitude sur la durée de chaque étape ($\alpha = 0$) et avec un bruit de connectivité élevé $\sigma = 1.5$. Désormais, la méthode HOSMR fournit des résultats équivalents voire légèrement plus faibles que les méthodes joinNMF et graphNMF. L'importance de la contrainte fused-lasso devient évidente dans ce cas, comme le montre la méthode snnP2 qui ne l'applique pas. Finalement, la Figure 6.5b intègre l'incertitude sur la durée de chaque étape ($\alpha = 1$). L'avantage de la contrainte fused-lasso est confirmé dans ces situations. En effet, graphNMF et joinNMF dominent l'ensemble des autres méthodes, notamment la HOSMR. Il semblerait que la méthode graphNMF ait des performances légèrement meilleures que joinNMF. Le fait que la méthode graphNMF produise dans un premier temps des facteurs de CFs spécifiques à chaque crise, qui peuvent posséder quelques particularités malgré une forte similarité globale due à la contrainte multivues, semble être un avantage pour ces situations très bruitées. On rappelle que ces facteurs sont ensuite moyennés pour fournir la matrice $\mathbf{F}_{graphNMF}$, cela permet de diminuer l'impact des particularités contenues dans chaque matrice de CFs.

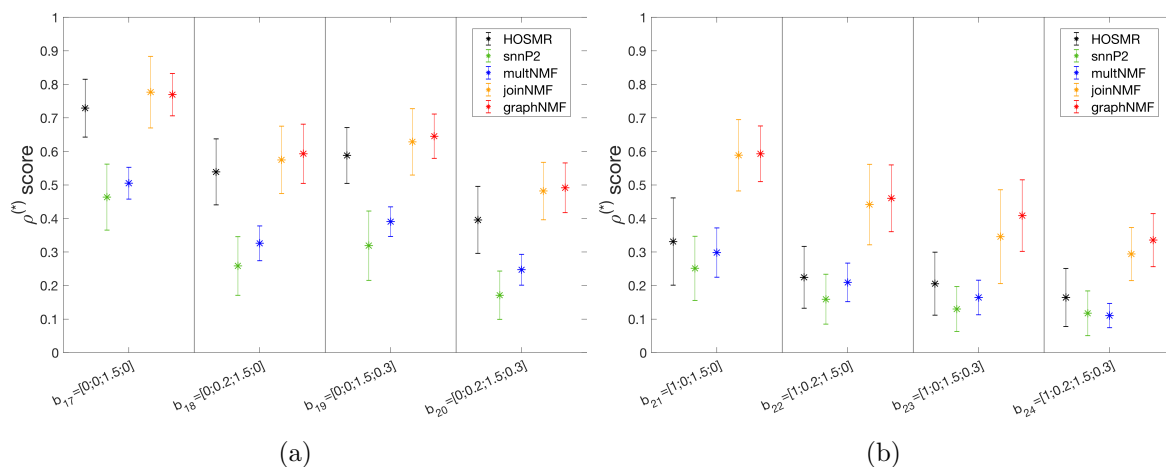


FIGURE 6.5 – Score $\rho^{(*)}$ atteint par les 5 méthodes proposées pour 4 conditions expérimentales avec $\sigma = 1.5$ et (a) $\alpha = 0$, (b) $\alpha = 1$

6.3 Application sur des données sEEG réelles.

On considère des données sEEG réelles de trois patients avec une épilepsie focale. Le premier patient est le même que celui de l’application de la Section 5.5. Les crises de ce patient ont des durées homogènes, ainsi on peut comparer les résultats obtenus entre les nouvelles méthodes proposées dans ce chapitre avec la HOSMR. Pour les deux autres patients, les crises ont des durées différentes.

6.3.1 Patient avec des crises de durées homogènes

Traitement : Le premier patient, qui est le même que dans la Section 5.5, présente 4 crises. Afin d’éviter des corrélations spatiales trop fortes entre les signaux, 33 électrodes réparties équitablement sur les 108 électrodes de l’enregistrement sont sélectionnées (soit 2 à 3 par tige). Les mesures dynamiques de CFs sont calculées via la PLV [192]. Contrairement à la section précédente, on applique maintenant un filtrage passe-bande afin de garder uniquement les fréquences comprises entre 20 et 100 Hz. Ce filtrage permet d’observer plus facilement les CFs qui s’activent en début de crise.

À partir du tenseur \mathcal{X} , on infère la matrice $\mathbf{F}_{HOSMR} \in \mathbb{R}^{L \times K}$, $\mathbf{F}_{snnP2} \in \mathbb{R}^{L \times K}$, $\mathbf{F}_{graphNMF} \in \mathbb{R}^{L \times K}$, ainsi que le tenseur $\mathcal{F}_{multNMF} \in \mathbb{R}^{L \times K \times S}$. La méthode graphNMF n’est pas testée car il s’agit d’un modèle plus complexe, qui demande un temps de calcul supérieur à joinNMF (voir le Tableau 6.1) et qui produit dans ce contexte des résultats similaires à cette méthode.

On fixe $K = 5$, les paramètres sont fixés empiriquement, afin d’avoir à la fois une forte parcimonie des facteurs de CFs et des résultats cohérents. Ainsi, pour la méthode HOSMR on fixe $\lambda_1 = 0.5$ et $\lambda_2 = 0.5$, pour la méthode snnP2 on a $\lambda_1 = 1$. Pour la méthode multNMF on a $\lambda_1 = 0.4$, $\lambda_2 = 0.4$, $\lambda_3 = 0.1$ et $\eta = 0$. Enfin, on utilise le même jeu de paramètres pour la méthode joinNMF que celui de la méthode multNMF sauf que $\lambda_1 = S\lambda_2$ et $\eta = 1$. Bien que le choix des paramètres pour chaque méthode soit important, on précise que pour de nombreuses configurations les facteurs de CFs obtenus sont similaires.

Finalement, comme étape de post-traitement on applique la transformation de l’Équation (6.24). Cette transformation consiste à diviser chaque facteur de CFs par une constante. Chacun des profils de CFs est alors multiplié par cette même constante pour conserver la même approximation $\mathbf{F}\mathbf{V}^t\{s\} \approx \mathbf{X}^t\{s\}$, $\forall s \in \{1, \dots, S\}$. Étant donné que la fonction de régularisation utilisée pour imposer la parcimonie est un relâchement de la fonction qui compte explicitement le nombre de CFs non nulles dans chaque facteur (norme ℓ_0), de nombreuses CFs associées à de faibles valeurs peuvent persister. On peut voir ces CFs comme des artefacts et vouloir les éliminer. On ne retient alors que les CFs telles que $\hat{\mathbf{f}}_{lk} \geq 0.2 \forall k \in \{1, \dots, K\}$, $\forall l \in \{1, \dots, L\}$.

Résultats : La Figure 6.6.(a) présente les 5 profils d’activation temporelle de la crise générique fournie par la méthode HOSMR (correspondant à la matrice \mathbf{V} provenant de l’Algorithme 10). On peut de nouveau observer différentes périodes d’activation à partir de ces profils. La Figure 6.6.(b)-(f) montre les positions des 33 électrodes projetées sur le plan transversal selon le repère de Talairach [102]. Les cinq sous-graphes de CFs, obtenus à partir de la matrice \mathbf{F}_{HOSMR} sont représentés par des liens entre les paires d’électrodes. L’épaisseur des liens indique le poids associé à chaque CFs dans le facteur $\hat{\mathbf{f}}_{:k}$ correspondant. Chaque sous-graphe s’active selon un des 5 profils d’activation temporelle.

Pour la méthode snnP2, les profils d’activation spécifiques à chaque crise sont fournis Figure 6.7. Bien que cette méthode n’impose pas spécifiquement la non négativité des profils d’activation, on remarque qu’ici ils sont bien non négatifs. Un trait vertical rouge est affiché pour indiquer le début de la crise, cette information est obtenue à partir de l’observation clinique du neurologue. Les graphes de CFs associés à chaque profil sont affichés dans la Figure 6.8.

Les profils d’activation spécifiques à chaque crise obtenus grâce à la méthode multNMF sont affichés dans la figure 6.9. Comme la méthode multNMF génère pour chaque crise des facteurs de CFs différents, on représente l’ensemble des sous-graphes de CFs obtenus sous forme d’un tableau de Figures 6.10. Les sous-graphes sont rangés par crises et par groupes qui semblent apporter une information clinique similaire (quand c’est possible).

Finalement, pour la méthode joinNMF, les profils d’activation sont présentés dans la Figure 6.11 tandis que leurs sous-graphes communs à toutes les crises sont exposés dans la Figure 6.12.

Discussion : La décomposition obtenue grâce à la HOSMR produit une crise générique censée résumer les quatre crises du patient 9. On peut alors observer, en analysant les profils d'activation de la méthode `snnP2` mais aussi ceux des méthodes `multNMF` et `joinNMF`, comment les profils d'activation de la HOSMR correspondent à une combinaison des profils d'activation spécifiques à chaque crise. Les profils d'activation obtenus grâce aux méthodes `multNMF` et `joinNMF` montrent l'influence de la contrainte `fused-lasso`. Le découpage des périodes temporelles de chaque crise est beaucoup plus prononcé pour ces deux méthodes que pour la méthode `snnP2`.

Pour toutes les méthodes, les facteurs de CFs obtenus sont similaires. Pour la méthode `multNMF`, qui propose pour chaque crise des facteurs de CFs spécifiques, on observe que le sous-graphe associé au groupe 1 de la crise 1 n'est pas reproductible. Globalement, même pour les sous-graphes très reproductibles, comme ceux associés aux groupes 2 et 5, des différences subsistent dans l'activation des liens d'une crise à l'autre. Les méthodes `snnP2`, `joinNMF` et HOSMR identifient les sous-graphes de CFs les plus fréquemment obtenus par la méthode `multNMF`. Cela montre bien que des facteurs de CFs plus robustes peuvent être obtenus en cherchant ceux qui sont communs à l'ensemble des crises.

Le ou les sous-graphes de CFs marquant le début de la crise sont particulièrement intéressants pour les épiléptologistes. On remarque ici que le début d'une crise est systématiquement suivi par la forte activation d'un à deux sous-graphes de CFs.

Conclusion : Ces crises de durées homogènes fournissent des résultats cohérents et similaires pour toutes les méthodes. On remarque cependant que la méthode `joinNMF` (et dans une moindre mesure la méthode `snnP2`) fournit des résultats aussi pertinents que la méthode HOSMR, en plus de fournir des profils d'activation spécifiques à chaque crise. La méthode `joinNMF` fournit des profils d'activation particulièrement faciles à interpréter grâce à la contrainte `fused-lasso`.

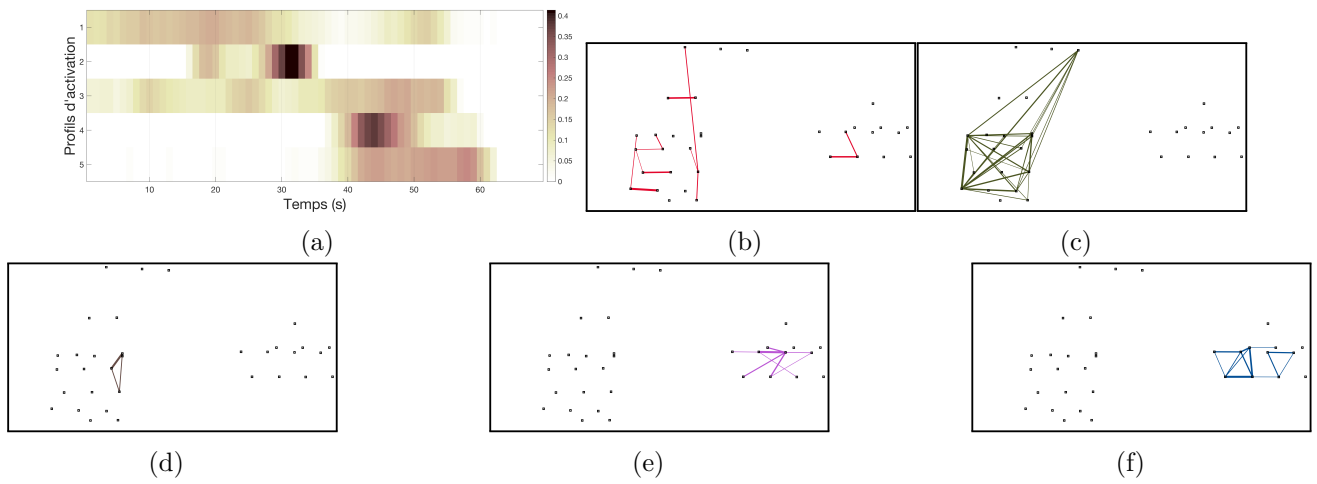


FIGURE 6.6 – (a) Les cinq profils d'activation de la crise générique, obtenus grâce à la méthode HOSMR pour le patient avec des crises de durées homogènes. (b) - (f) Sous-graphes obtenus à partir des facteurs de CFs 1 - 5 fournis par la méthode HOSMR.

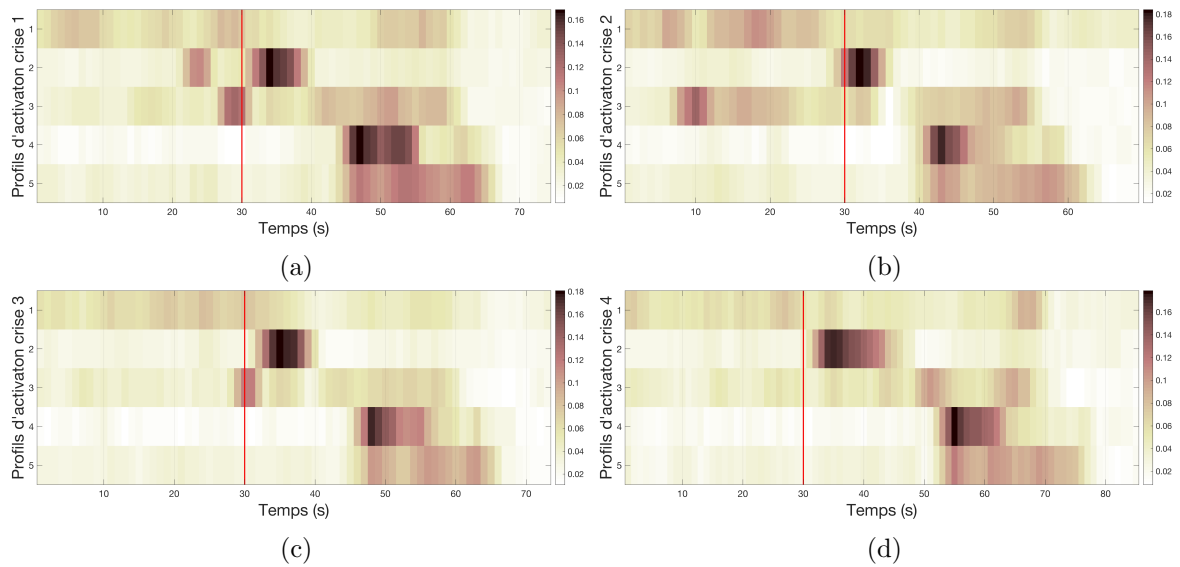


FIGURE 6.7 – Les cinq profils d'activation spécifiques à chacune des 4 crises obtenus grâce à la méthode smnP2 pour le patient avec des crises de durées homogènes. Le trait rouge indique le début de la crise.

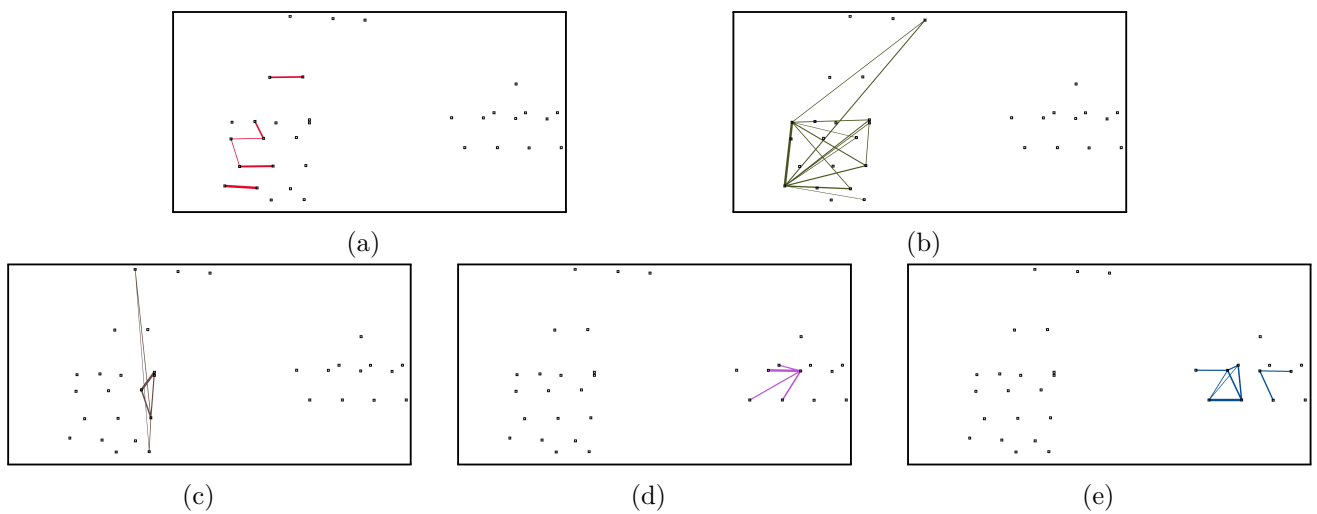


FIGURE 6.8 – (a) - (e) Sous-graphes obtenus à partir des facteurs de CFs 1 - 5 fournis par la méthode smnP2.

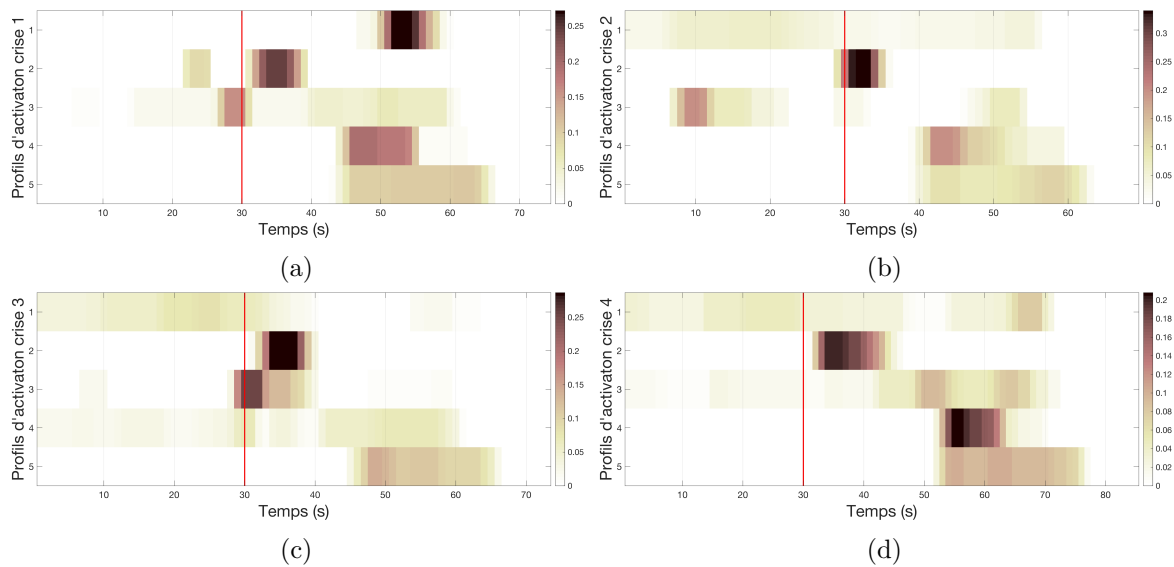


FIGURE 6.9 – Les cinq profils d'activation spécifiques à chacune des 4 crises obtenus grâce à la méthode multNMF pour le patient avec des crises de durées homogènes. Le trait rouge indique le début de la crise.

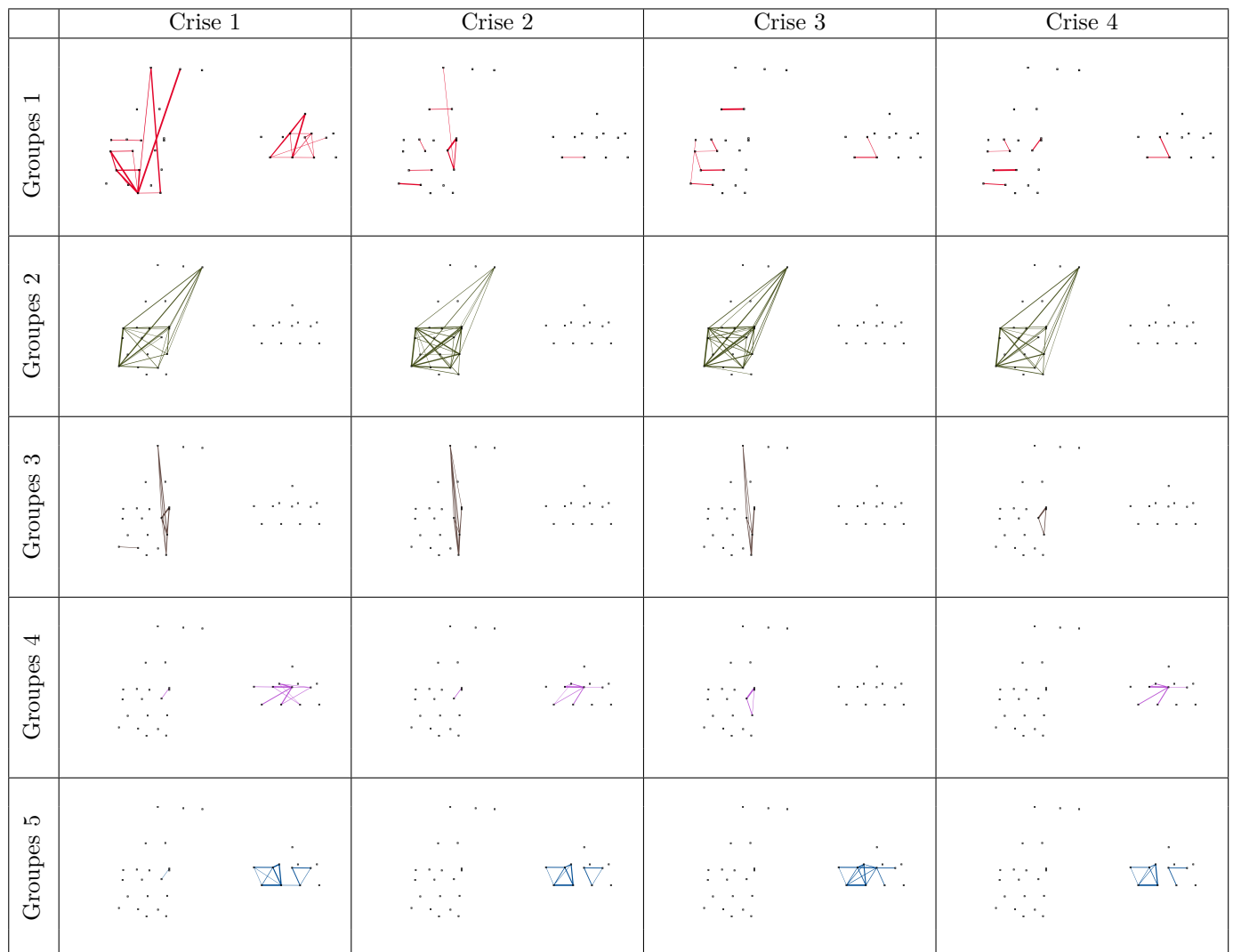


FIGURE 6.10 – Sous-graphes obtenus à partir des facteurs de CFs fournis par la méthode multNMF pour le patient avec des crises homogènes. Les graphes sont rangés par crises et en groupes qui semblent apporter une information clinique similaire (quand c'est possible).

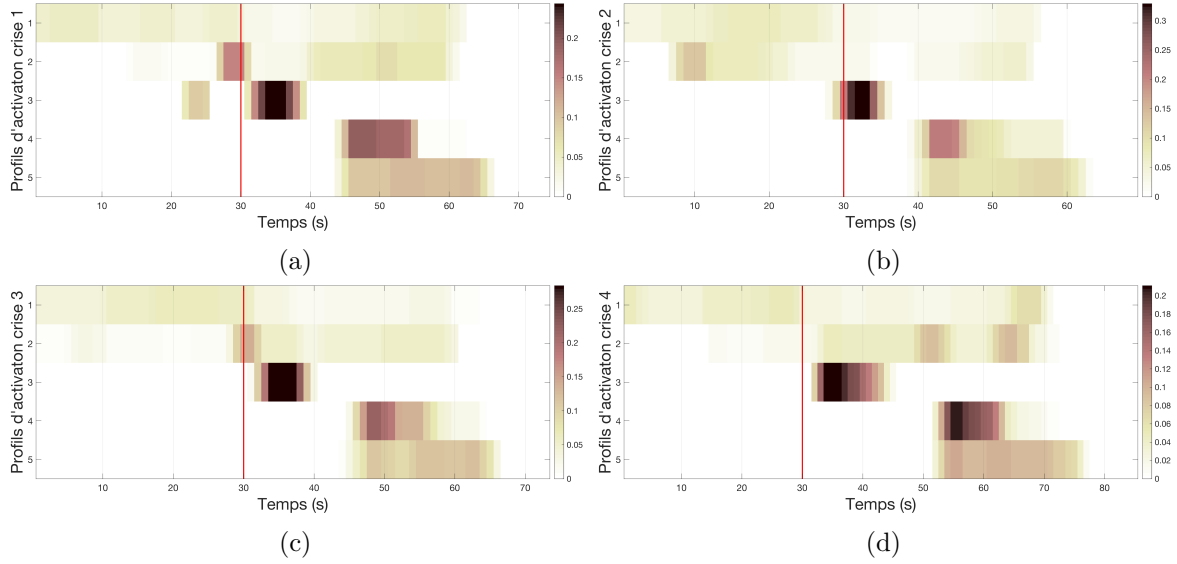


FIGURE 6.11 – (a) Les cinq profils d’activation de la crise générique, obtenus grâce à la méthode joinNMF pour le patient avec des crises de durées homogènes.

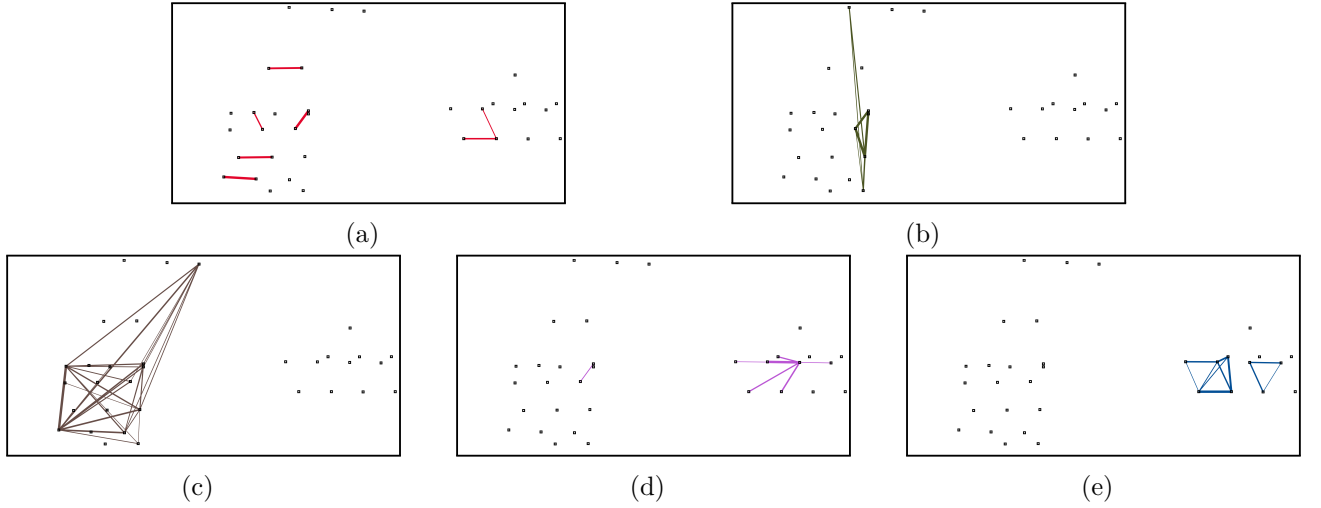


FIGURE 6.12 – (a) - (e) Sous-graphes obtenus à partir des facteurs de CFs 1 - 5 fournis par la méthode joinNMF.

6.3.2 Patient avec des crises hétérogènes

Traitement : Le second patient présente 3 crises avec des durées différentes. La première dure 134 secondes, la seconde et la troisième respectivement 56 et 25 secondes. Afin d’éviter des corrélations spatiales trop fortes entre les signaux, 32 électrodes réparties équitablement sur les 94 électrodes de l’enregistrement sont sélectionnées. Les mesures dynamiques de CFs sont calculées via la PLV [192], aucun seuillage n’est appliqué. De nouveau, on applique un filtrage passe-bande pour garder uniquement les fréquences comprises entre 20 et 100 Hz afin d’observer plus facilement les CFs qui s’activent en début de crise. Dans un premier temps, afin de pouvoir appliquer la méthode HOSMR, un tenseur contenant les mesures de CFs pour les trois crises est formé $\mathcal{X} \in \mathbb{R}^{L \times T \times S}$, avec $T = 134$ correspondant à la crise la plus longue. Les évènements post et pré-ictaux sont gardés pour arriver à une durée de 134 secondes pour les crises 2 et 3. À partir du tenseur \mathcal{X} , on infère la matrice \mathbf{F}_{HOSMR} . Les autres méthodes s’appliquent sur une liste de matrices de CFs $\mathbf{X}\{s\} \in \mathbb{R}^{L \times T(s) \times S}$, $s \in \{1, 2, 3\}$, avec respectivement $T(1) = 134$, $T(2) = 56$ et $T(3) = 25$. Grâce à cette liste de matrices, on infère les matrices de facteurs $\mathbf{F}_{snnP2} \in \mathbb{R}^{L \times K}$, $\mathbf{F}_{joinNMF} \in \mathbb{R}^{L \times K}$, ainsi que le tenseur $\mathcal{F}_{multNMF} \in \mathbb{R}^{L \times K \times S}$.

On fixe $K = 5$, les autres paramètres sont fixés empiriquement comme pour le cas précédent, afin d’avoir à la fois une forte parcimonie des facteurs de CFs et des résultats cohérents. Ainsi pour la

méthode HOSMR on a $\lambda_1 = \lambda_2 = 0.2$, pour la méthode snnP2 $\lambda_1 = 0.2$, pour la méthode multNMF on a $\lambda_1 = 0.3$, $\lambda_2 = 0.3$, $\lambda_3 = 0.1$ et $\eta = 0$. Enfin, on utilise le même jeu de paramètres pour la méthode joinNMF que ceux de la méthode multNMF sauf que $\lambda_1 = S\lambda_2$ et $\eta = 1$.

Finalement, comme étape de post-traitement on applique les mêmes étapes que pour le cas homogènes : (1) normalisation des facteurs de CFs et des profils d'activation à partir de la formule de l'Équation (6.24). (2) On ne retient alors que les CFs tel que $\hat{\mathbf{f}}_{tk} \geq 0.2 \forall k \in \{1, \dots, K\}$.

Résultats : La Figure 6.13.(a) présente les 5 profils d'activation temporelle de la crise générique fournie par la méthode HOSMR. La Figure 6.13.(b)-(f) montre les cinq sous-graphes de CFs, obtenus à partir de la matrice \mathbf{F}_{HOSMR} .

Pour la méthode snnP2, les profils d'activation spécifiques à chaque crise sont fournis dans la Figure 6.14. On précise qu'une valeur légèrement négative apparaît pour le facteur 3 de la crise 3 à la seconde 9. Cet effet est donc très négligeable. Cela illustre cependant comment la méthode snnP2 privilégie sans garantie, des profils d'activation positifs. Les sous-graphes de CFs associés à chaque profil sont affichés dans la Figure 6.15.

Les profils d'activation spécifiques à chaque crise obtenus grâce à la méthode multNMF sont dépeints Figure 6.16. On représente l'ensemble des sous-graphes de CFs obtenus dans le tableau de Figures 6.17.

Finalement, pour la méthode joinNMF, les profils d'activation sont présentés dans la Figure 6.18 tandis que leurs sous-graphes communs à chaque crise sont exposés dans la Figure 6.19.

Discussion : On remarque que les profils d'activation des méthodes HOSMR et snnP2 ne sont pas très faciles à interpréter. Pour HOSMR, c'est sûrement dû à la création du tenseur \mathcal{X} , qui contient des crises avec des comportements temporels trop différents. Pour la snnP2, on pense que c'est le manque de contraintes (notamment la parcimonie) appliquées sur les profils d'activation qui empêche d'inférer des composantes interprétables. Malgré tout, la HOSMR continue de fournir des facteurs de CFs cohérents, en particulier le 1, 2 et 3. On retrouve ces mêmes 3 facteurs pour la méthode snnP2. Le facteur 4 de la méthode snnP2 contient de nombreuses CFs d'arrière plan non informatives. C'est de nouveau la conséquence d'un manque de parcimonie sur les profils d'activation qui permet la génération de sous-graphes de CFs qui s'activent durant toute la durée de chaque crise.

L'intérêt de la contrainte fused-lasso est rendu évident lorsque l'on observe les profils d'activation des méthodes multNMF et joinNMF. Un découpage clair des différents événements apparaît. Pour ce patient, on observe beaucoup de variations dans les sous-graphes de CFs obtenus pour chaque crise par la méthode multNMF. Certains sous-graphes, comme ceux du groupe 4 pour les crises 2 et 3 semblent assez aléatoires. La méthode joinNMF offre des sous-graphes plus localisés spatialement qui semblent regrouper les motifs reproductibles de la méthode multNMF. En effet, le sous-graphe de la Figures 6.19.(a) peut-être associé aux sous-graphes de groupe 2 pour toutes les crises, et du groupe 3 pour la crise 2 dans la Figure (6.17). Le sous-graphe de la Figures 6.19.(b) correspond majoritairement aux sous-graphes du groupe 1 pour les 3 crises dans la Figure (6.17). De même pour le sous-graphe de la Figures 6.19.(c), celui-ci semble en plus contenir des éléments des sous-graphes du groupe 3 pour la crise 1 et 3, et du groupe 4 et 5 de la crise 3 et 4. Le sous-graphe de la Figures 6.19.(d) contient des éléments des sous-graphes du groupe 2 de la crise 2, du groupe 4 de la crise 1 et du groupe 5 de la crise 2 et 3.

Enfin, un facteur caractéristique et spécifique d'une crise, correspondant au facteur 5, crise 1 de la méthode multNMF, est aussi identifié par la méthode joinNMF dans la Figure 6.19.(e). Cela montre que que la méthode joinNMF, en plus de privilégier des facteurs de CFs communs à toutes les crises, est capable de révéler un événement particulier majeur d'une crise (on remarque que ce facteur ne s'active quasiment pas pour les autres crises). Dans cette situation, la méthode joinNMF a donc l'avantage de rester explicite dans l'identification des graphes de CFs importants pour chaque crise.

Conclusion : Pour des crises qui ont des durées très hétérogènes, une alternative à la méthode HOSMR est cruciale. La méthode joinNMF semble la plus adaptée pour à la fois obtenir des profils d'activation intelligibles tout en privilégiant des facteurs de CFs robustes communs à toutes les crises.

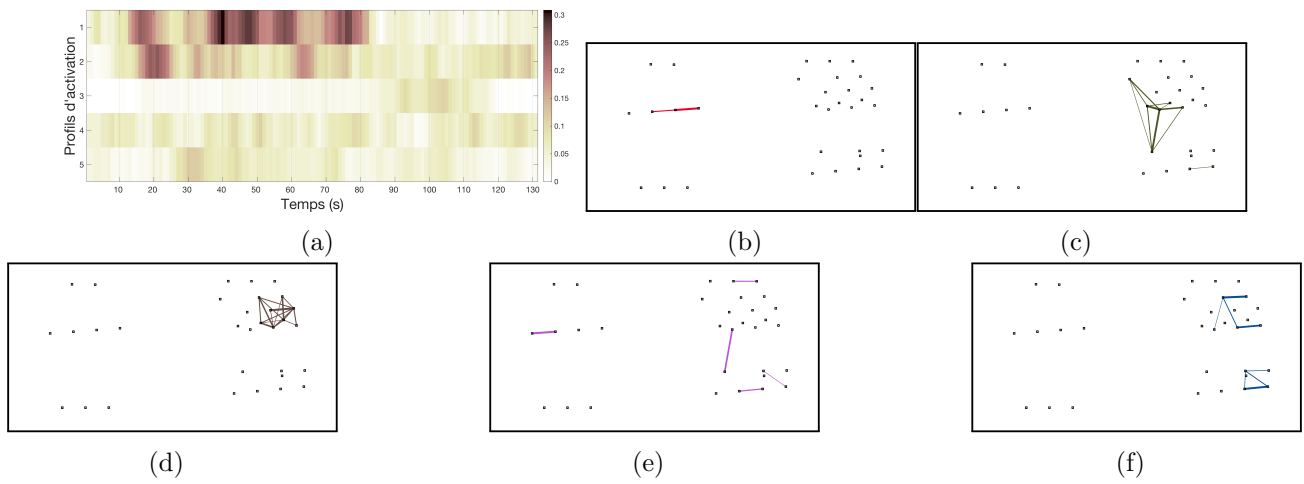


FIGURE 6.13 – (a) Les cinq profils d’activation de la crise générique obtenus grâce à la méthode HOSMR pour le patient avec des crises de durées hétérogènes. (b) - (f) Sous-graphes obtenus à partir des facteurs de CFs 1 - 5 fournis par la méthode HOSMR.

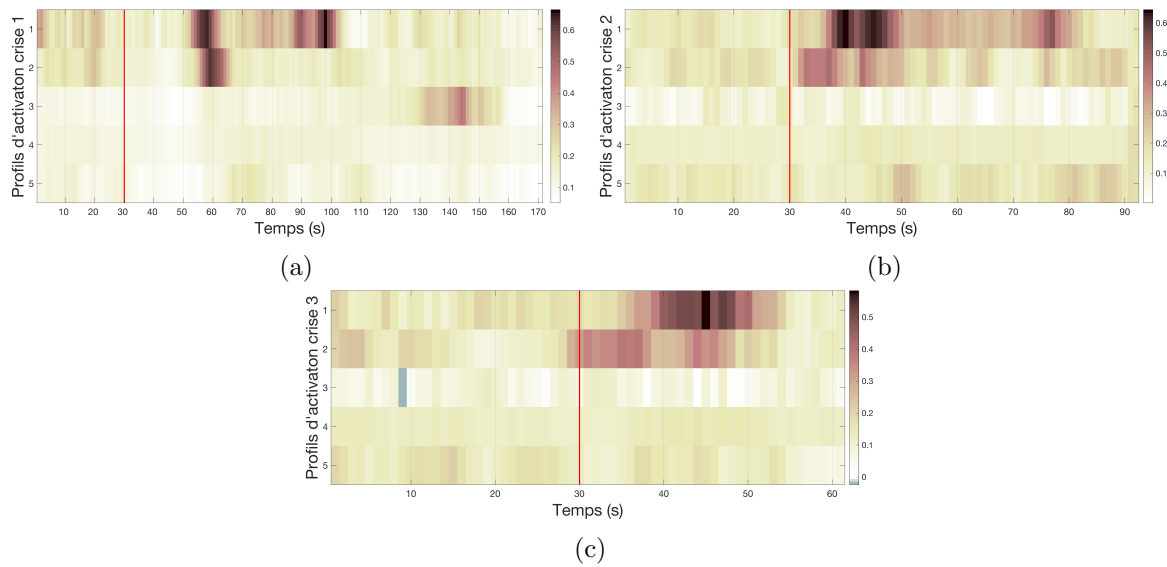


FIGURE 6.14 – Les cinq profils d’activation de la crise générique, obtenus grâce à la méthode snnP2 pour le patient avec des crises de durées hétérogènes.

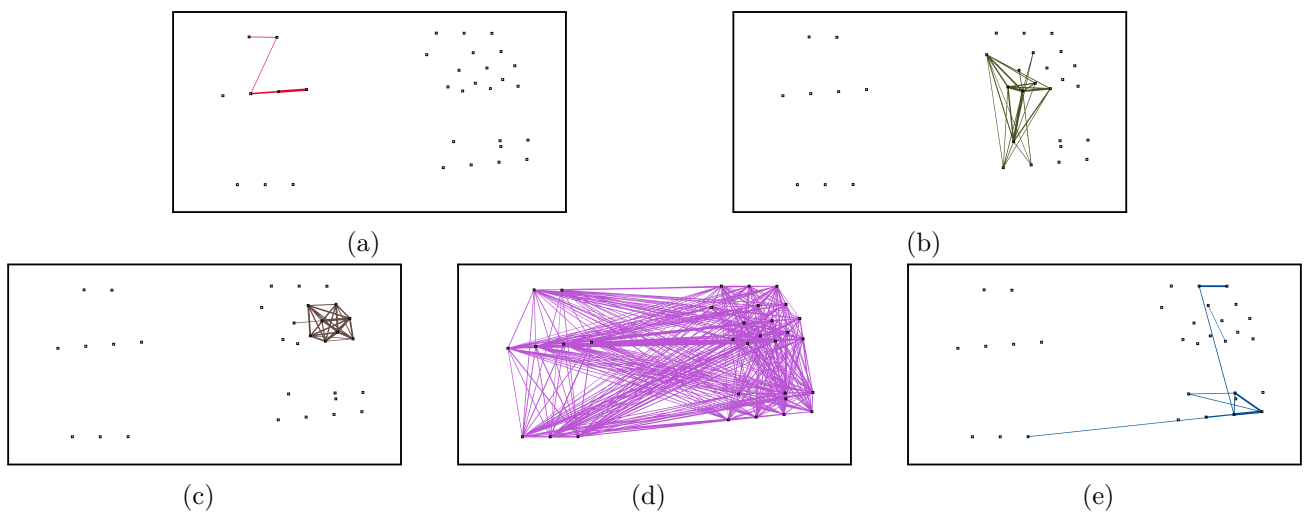


FIGURE 6.15 – Graphes binaires de CFs obtenus pour chaque profil d’activation selon la méthode snnP2. Les graphes de CFs (a) - (e) sont respectivement associés aux profils d’activation 1 - 5.

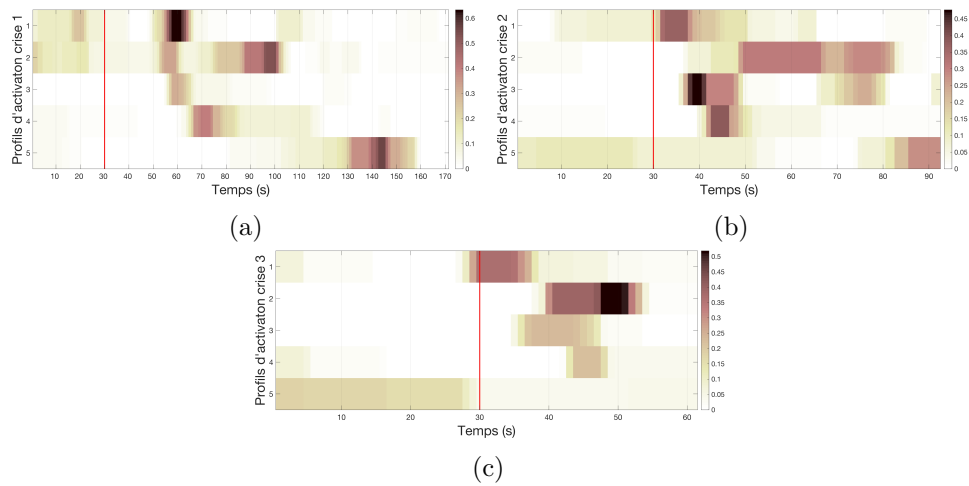


FIGURE 6.16 – Les cinq profils d’activation spécifiques à chacune des 4 crises obtenus grâce à la méthode multNMF pour le patient avec des crises de durées hétérogènes. Le trait rouge indique le début de la crise.

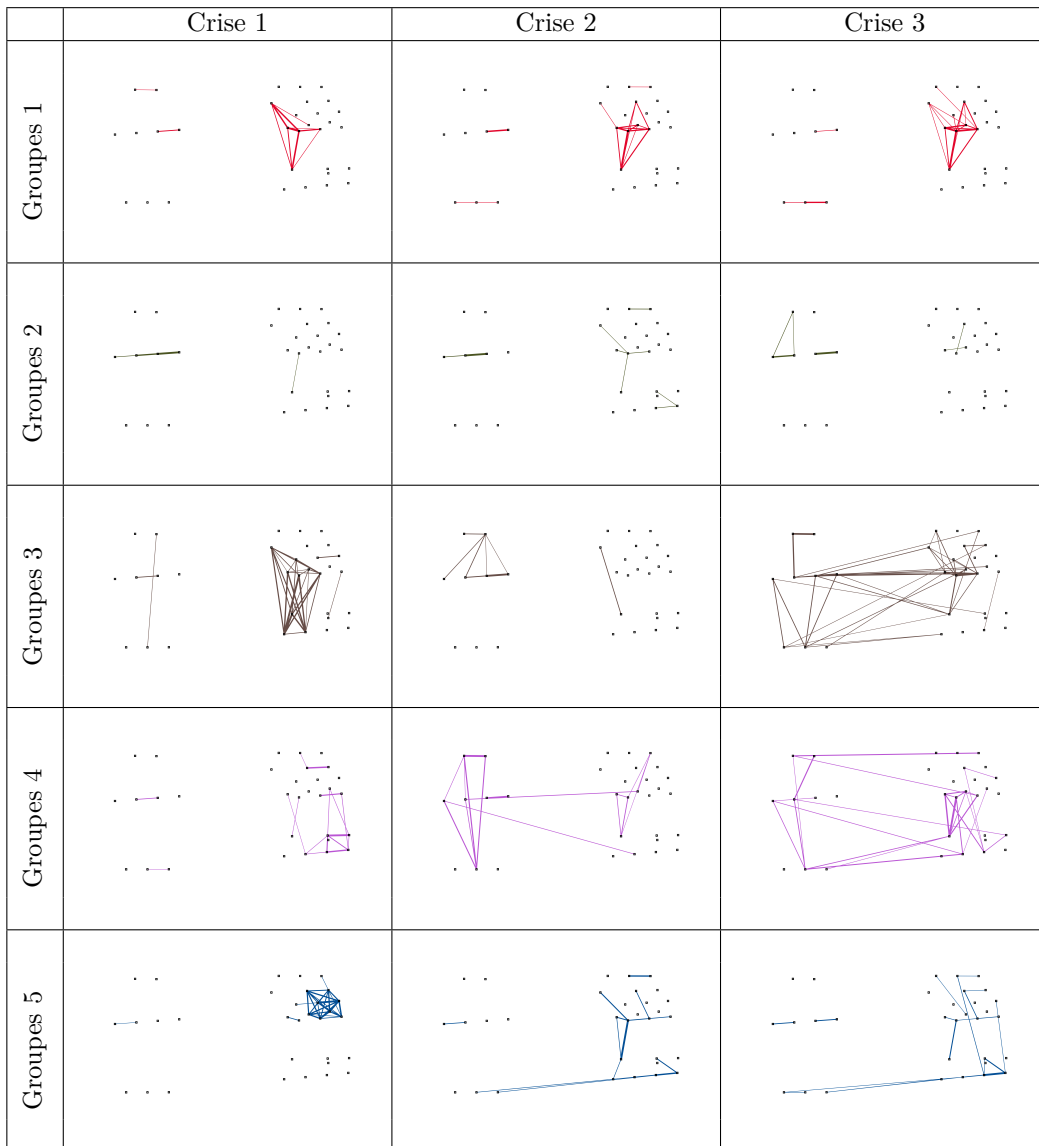


FIGURE 6.17 – Sous-graphes obtenus à partir des facteurs de CFs fournis par la méthode multNMF pour le patient avec des crises hétérogènes. Les graphes sont rangés par crise et en facteurs qui semblent apporter une information clinique similaire (quand c’est possible).

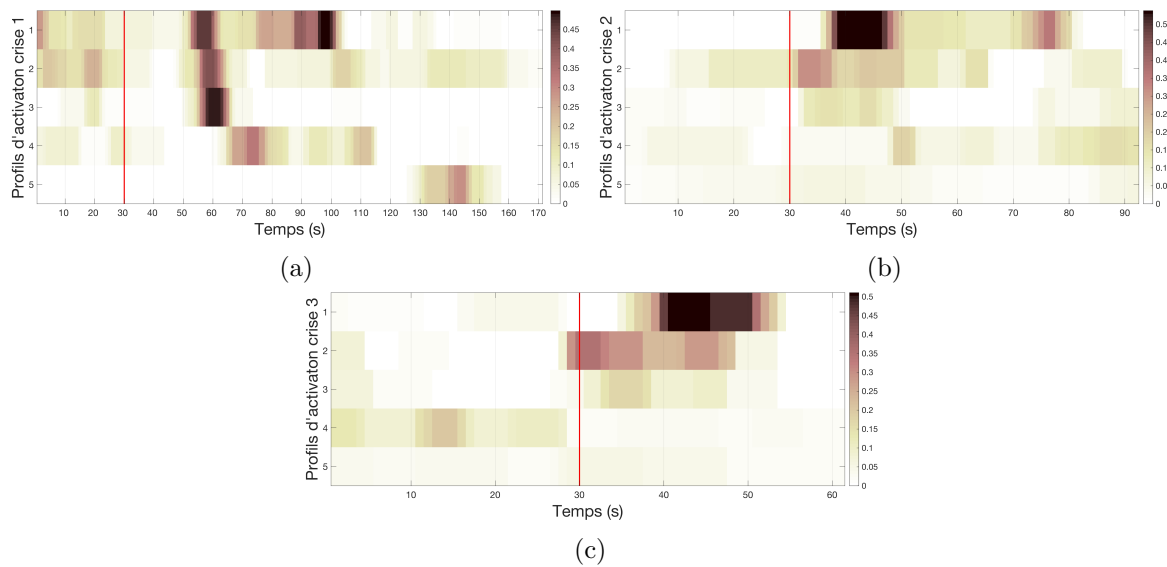


FIGURE 6.18 – (a) Les cinq profils d’activation de la crise générique, obtenus grâce à la méthode joinNMF pour le patient avec des crises de durées hétérogènes.

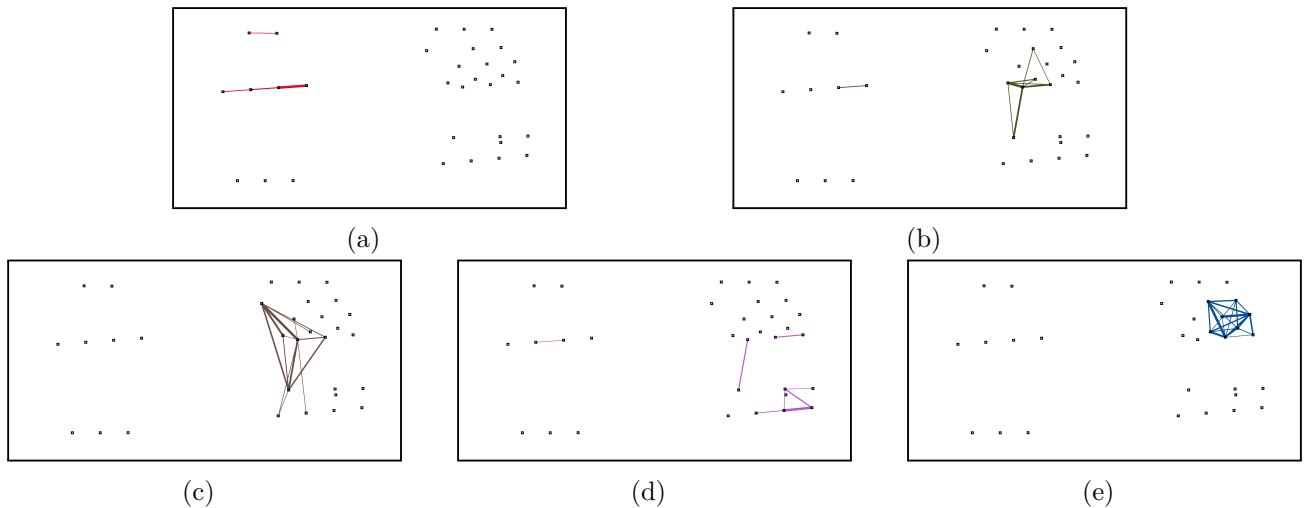


FIGURE 6.19 – (a) - (e) Sous-graphes obtenus à partir des facteurs de CFs 1 - 5 fournis par la méthode joinNMF.

6.4 Conclusion

Dans ce chapitre, des alternatives à la HOSMR ont été proposées dans le cas où les crises ont des durées hétérogènes. Dans un premier temps, des modifications de l’algorithme Parafac2 sont explicitées afin de favoriser l’inférence de composantes interprétables. Cependant, il n’est pas simple de contrôler les contraintes temporelles qui sont appliquées sur les profils d’activation via cette méthode. C’est pourquoi des décompositions en matrices non-négatives jointes ont été étudiées, elles peuvent être vues comme des simplifications de la méthode Parafac2. On propose une réécriture d’une méthode déjà existante, la décomposition en matrice non-négative multivues, en montrant que la régularisation multivues correspond à une régularisation de Tikhonov faisant intervenir une matrice laplacienne. Cette modification fournit une décomposition plus générale qui semble plus robuste à des crises bruitées. De plus, l’implémentation de cette décomposition est alors plus simple. Finalement, différentes méthodes sont comparées sur un modèle puis sur des données réelles. Les résultats montrent l’importance de l’application de la contrainte fused-lasso sur les profils d’activation dans le cas où les données sont bruitées. À partir de ces expériences, on recommande les méthodes graphNMF et JoinNMF comme alternatives à la HOSMR dans le cas où les crises ont des durées hétérogènes. Dans le chapitre suivant,

on applique la méthode multNMF dans un objectif clinique. La méthode multNMF est préférée à graphNMF pour l'application du chapitre suivant car : (i) la complexité de la méthode graphNMF est plus élevée (un paramètre supplémentaire) pour des résultats souvent similaires, (ii) Le temps de calcul de la méthode multNMF est beaucoup plus court que le temps de calcul de la méthode graphNMF.

Les travaux de ce chapitre pourraient faire l'objet d'analyses supplémentaires :

- Une investigation plus poussée de la méthode graphNMF pourrait être proposée. En effet, la méthode graphNMF peut être intéressante pour des cas cliniques complexes. Pour certains patients, les activations de CFs peuvent être similaires pour beaucoup de crises, mais varier considérablement pour certaines crises spécifiques. Il serait alors intéressant de prendre en compte la similarité des différentes crises dans la matrice laplacienne de la méthode graphNMF permettant d'obtenir des décompositions plus robustes pour ces patients.
- Comme conséquence du point précédent, il est désormais nécessaire de trouver une mesure de similarité entre les différentes crises d'un patient. On peut par exemple comparer les différentes matrices de CFs obtenues pour chaque crise. Cela implique l'utilisation de mesures applicables à des matrices dont le nombre de colonnes peut varier [206]. L'une des mesures de ce type couramment utilisée est la distance de Grassman. Cependant, cette dernière ne semble pas adaptée pour bien discriminer des crises avec des comportements différents.
- Les méthodes graphNMF et JoinNMF peuvent aussi être utilisées dans un objectif un peu différent. Effectivement, à l'instar de l'analyse canonique des corrélations, au lieu d'être directement utilisé pour découvrir des facteurs de CFs et des profils d'activation, elles peuvent être utilisées comme des méthodes pour plonger les matrices de CFs de durées différentes dans un espace où elles ont toutes la même durée. Ainsi, une méthode tensorielle comme la HOSMR pourrait ensuite être appliquée.
- Des extensions de Parafac2 afin d'obtenir des profils d'activation parcimonieux pourraient être étudiées. On a présenté à la fin de la Section (6.1.1) deux méthodes pouvant favoriser, mais sans la garantir, la parcimonie de ces profils. Un des relâchements de Parafac2 proposée dans [51] afin d'obtenir des composantes non négatives pourrait être une bonne approche pour ajouter les contraintes désirées sur les profils d'activation.
- Finalement, le choix des hyperparamètres est encore une fois un problème non résolu. Heureusement, pour de nombreuses combinaisons de paramètres, les décompositions obtenues sont pertinentes et fournissent des résultats semblables. De plus, pour des données de CFs qui sont la plupart du temps comprises entre 0 et 1, les valeurs des hyperparamètres obtenues oscillent généralement entre 0.01 et 10, permettant de borner le nombre de tests à réaliser pour trouver une bonne configuration.

Annexe

6.A Implémentation des algorithmes présentés dans ce chapitre

6.A.1 Implémentation de l'algorithme snnP2

L'implémentation de la méthode snnP2 dont le critère est fourni dans l'Équation (6.3) consiste à appliquer l'algorithme Parafac2 décrit dans l'Algorithme 5, en remplaçant le calcul de la CP décomposition par une CP non négative (nnCP) avec contrainte de parcimonie appliquée sur la matrice \mathbf{F} . L'implémentation de la snnP2 est décrite dans l'Algorithme 12.

Algorithm 12 Estimation de \mathbf{F} , \mathbf{W} , \mathbf{R} , $\mathbf{P}\{1\}$, ..., $\mathbf{P}\{S\}$ minimisant le critère de la snnP2 (si $\lambda_1 > 0$).

Require: $\mathcal{X} \in \mathbb{R}^{L \times T \times S}$, le paramètre de réduction K , l'incrément de tolérance $\epsilon > 0$, i_{max} , paramètre de parcimonie λ_1 , une première supposition pour \mathbf{F}_0 , \mathbf{Q}_0 , \mathbf{W}_0 et $\mathbf{P}\{s\}_0$

$i = 0$

while $i < i_{max}$ or $[f(\mathbf{F}_i, \mathbf{Q}_i, \mathbf{W}_i, \mathbf{P}_i\{1\}, \dots, \mathbf{P}_i\{S\}) - f(\mathbf{F}_{i+1}, \mathbf{Q}_{i+1}, \mathbf{W}_{i+1}, \mathbf{P}_{i+1}\{1\}, \dots, \mathbf{P}_{i+1}\{S\})] \leq \epsilon$

do

Calculer $\tilde{\mathbf{U}}\{s\}\tilde{\mathbf{\Sigma}}\{s\}\tilde{\mathbf{V}}\{s\}$ la SVD de $\mathbf{X}\{s\}_i^t \mathbf{F}_i \mathbf{D}(\mathbf{W}_{s \cdot})_i \mathbf{Q}_i^t$

$\mathbf{P}\{s\}_{i+1} = \tilde{\mathbf{U}}(s)\tilde{\mathbf{V}}\{s\}^t$

$\tilde{\mathbf{X}}_{::s} = \mathbf{X}\{s\}\mathbf{P}\{s\}_{i+1}^t$

Calculer \mathbf{F}_{i+1} , \mathbf{C}_{i+1} , \mathbf{W}_{i+1} via décomposition nnCP de $\tilde{\mathbf{X}}$ avec l'Algorithme 2 pour le paramètre λ_1

$i = i+1$

end while

Avec :

$$f(\mathbf{F}, \mathbf{Q}, \mathbf{W}, \mathbf{P}\{1\}, \dots, \mathbf{P}\{S\}) = \sum_{s=1}^S \|\mathbf{X}\{s\} - \mathbf{F}_i \mathbf{D}(\mathbf{W}_{s \cdot})_i \mathbf{Q} \mathbf{P}\{s\}_i\|_F^2 + \lambda_1 \|\mathbf{F}\|_1. \quad (6.28)$$

La décomposition nnCP de $\tilde{\mathbf{X}}$ est réalisée en appliquant la méthode des moindres carrés alternatifs hiérarchiques (HALS) [49]. Cette méthode permet d'appliquer la contrainte de non négativité sur les 3 modes du tenseur $\tilde{\mathbf{X}}$, plus la régularisation parcimonieuse appliquée au facteur de CFs \mathbf{F} .

Algorithm 13 Estimation \mathbf{F} , \mathbf{Q} et \mathbf{W} minimisant le critère d'une mnCP, l'implémentation utilisée est la méthode des moindres carrés alternatifs hiérarchiques (HALS) [49].

Require: $\tilde{\mathbf{X}} \in \mathbb{R}^{L \times K \times S}$, le paramètre de réduction K , l'incrément de tolérance $\epsilon > 0$, j_{max} , paramètre de parcimonie λ_1 , une première supposition pour $\mathbf{F}_0, \mathbf{Q}_0, \mathbf{W}_0$

$j = 0$

while $j < j_{max}$ or $[g(\mathbf{F}_j, \mathbf{Q}_j, \mathbf{W}_j) - g(\mathbf{F}_{j+1}, \mathbf{Q}_{j+1}, \mathbf{W}_{j+1})] \leq \epsilon$ **do**

$\mathbf{Z} = \mathbf{W} \odot \mathbf{Q}$

$\mathbf{D} = \mathbf{Z}^t \mathbf{Z}$

for $k \in \{1, \dots, K\}$ **do**

$\mathbf{y} = (f:k)_j - \frac{1}{d_{kk}}(\mathbf{F}_j \mathbf{D}_{:k} - \mathbf{X} \mathbf{z}_{:k})$

$(f:k)_{j+1} = \text{NSD}_{\frac{2\lambda_1}{d_{kk}}}(\mathbf{y})$ voir l'Équation (5.63)

end for

$\mathbf{Z} = \mathbf{W} \odot \mathbf{F}$

$\mathbf{D} = \mathbf{Z}^t \mathbf{Z}$

for $k \in \{1, \dots, K\}$ **do**

$\mathbf{y} = (r:k)_j - \frac{1}{d_{kk}}(\mathbf{Q}_j \mathbf{D}_{:k} - \mathbf{X} \mathbf{z}_{:k})$

$(r:k)_{j+1} = \text{NSD}_0(\mathbf{y})$

end for

$\mathbf{Z} = \mathbf{Q} \odot \mathbf{F}$

$\mathbf{D} = \mathbf{Z}^t \mathbf{Z}$

for $k \in \{1, \dots, K\}$ **do**

$\mathbf{y} = (w:k)_j - \frac{1}{d_{kk}}(\mathbf{W}_j \mathbf{D}_{:k} - \mathbf{X} \mathbf{z}_{:k})$

$(w:k)_{j+1} = \text{NSD}_0(\mathbf{y})$

end for

end while

Avec :

$$g(\mathbf{F}, \mathbf{Q}, \mathbf{W}) = \sum_{s=1}^S \|\tilde{\mathbf{X}} - \mathbf{F}(\mathbf{W} \odot \mathbf{Q})^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1. \quad (6.29)$$

6.A.2 Implémentation de l'algorithme joinNMF

L'implémentation permettant de trouver une solution au critère de l'Équation (6.5) correspond à alterner deux étapes : la mise à jour des facteurs de CFs communs \mathbf{F} et la mise à jours des profils d'activation de chaque crise $\mathbf{V}\{s\} \forall s \in \{1, \dots, S\}$.

Mise à jour des facteurs de CFs communs

À l'itération $i + 1$, on suppose les matrices \mathbf{F}_i et $\mathbf{V}_i\{s\}$, $\forall s \in \{1, \dots, S\}$ connues. La mise à jour de la matrice \mathbf{F}_i après avoir fixé les profils d'activation temporelle correspond à trouver une solution au problème :

$$\underset{\mathbf{F} \geq 0}{\operatorname{argmin}} \sum_{s=1}^S \zeta(s) \|\mathbf{X}\{s\} - \mathbf{F} \mathbf{V}_i\{s\}^t\|_F^2 + \lambda \|\mathbf{F}\|_1, \quad (6.30)$$

qui est équivalent à la régression de lasso suivante :

$$\operatorname{argmin}_{\mathbf{F}} \frac{1}{2} \left\| [\sqrt{\zeta(1)}\mathbf{X}\{1\}, \dots, \sqrt{\zeta(S)}\mathbf{X}\{S\}] - \mathbf{F}[\sqrt{\zeta(1)}\mathbf{V}_i\{1\}^t, \dots, \sqrt{\zeta(S)}\mathbf{V}_i\{S\}^t] \right\|_F^2 + \lambda \|\mathbf{F}\|_1, \quad (6.31)$$

On définit $\bar{\mathbf{X}} = [\sqrt{\zeta(1)}\mathbf{X}\{1\}, \dots, \sqrt{\zeta(S)}\mathbf{X}\{S\}]$ et $\bar{\mathbf{V}}_i = [\sqrt{\zeta(1)}\mathbf{V}_i\{1\}^t, \dots, \sqrt{\zeta(S)}\mathbf{V}_i\{S\}^t]$. Ce problème peut alors être résolu en appliquant l'Algorithme 7 fourni dans la Section 5.3.1 afin d'inférer les facteurs de CFs \mathbf{F}_{i+1} . Cependant, afin de diminuer le temps de calcul, on propose d'utiliser la technique d'accélération FISTA. Cette procédure d'accélération ne sera pas détaillée ici. On conseille, comme introduction à cette méthode l'article original [27]. L'idée de la technique d'accélération FISTA est d'utiliser l'information de la matrice \mathbf{F}_i et aussi \mathbf{F}_{i-1} pour inférer la mise à jour \mathbf{F}_{i+1} . Cela permet de réduire la complexité d'un algorithme de gradient proximal. Afin de pouvoir appliquer cette technique d'accélération, il faut dans un premier temps modifier l'Algorithme 7 afin que l'opérateur proximal soit utilisé pour mettre à jour l'intégralité de la matrice \mathbf{F}_{i+1} plutôt qu'une de ses colonnes.

En reprenant les notations de la Section 5.3.1, on a dans ce cas :

$$h_1(\mathbf{F}) = \|\bar{\mathbf{X}} - \mathbf{F}\bar{\mathbf{V}}_i\|_F^2, \quad (6.32)$$

$$h_2(\mathbf{F}) = \lambda_1 \|\mathbf{F}\|_1, \quad (6.33)$$

$$\nabla h_1(\mathbf{F}) = 2\mathbf{F}\bar{\mathbf{V}}_i^t\bar{\mathbf{V}}_i - 2\bar{\mathbf{X}}\bar{\mathbf{V}}_i \quad (6.34)$$

$$\nu = \|2\bar{\mathbf{V}}_i^t\bar{\mathbf{V}}_i\|_2^2. \quad (6.35)$$

Avec $\|\bullet\|_2^2$ correspondant à la norme nucléaire d'une matrice. L'opérateur proximal de $h_2(\bullet)$ en considérant en plus la contrainte de non négativité, correspond à la fonction de seuillage doux non négative NSD définie dans l'Équation (5.63) :

$$\operatorname{prox}_{\frac{1}{\nu}\lambda_1 h_2(\bullet)}(x) = \operatorname{NSD}_{\frac{1}{\nu}\lambda_1}(x). \quad (6.36)$$

Une itération de la méthode de gradient proximal, appliquée directement sur la matrice \mathbf{F} de l'Équation (6.31) correspond alors à :

$$\mathbf{F}_{i+1} = \operatorname{NSD}_{\frac{1}{\nu}\lambda_1} \left(\mathbf{F}_{i+1} - \frac{2}{\nu} (\mathbf{F}_i \bar{\mathbf{V}}_i^t \bar{\mathbf{V}}_i - \bar{\mathbf{X}} \bar{\mathbf{V}}_i) \right). \quad (6.37)$$

La procédure afin de résoudre l'Équation (6.31) avec la technique d'accélération FISTA est fournie dans l'Algorithme (14).

Algorithm 14 Résout la régression parcimonieuse de l'Équation (6.30)

Require: $\bar{\mathbf{X}}$, $\bar{\mathbf{V}}_i$, K , le paramètre λ_1 , tolérance $\epsilon > 0$, et j_{max} . Une première supposition \mathbf{F}_0

$j = 1$

$\mathbf{Z}^{(0)} = \mathbf{F}_0$,

$\theta^{(0)} = 1$

Score = $1 + \epsilon$

while $j < j_{max}$ or Score $\leq \epsilon$ **do**

$\mathbf{Y} = \mathbf{F}_j - \frac{2}{\nu} (\mathbf{F}_j \bar{\mathbf{V}}_i^t \bar{\mathbf{V}}_i - \bar{\mathbf{X}} \bar{\mathbf{V}}_i)$

$\mathbf{Z} = \operatorname{NSD}_{\frac{1}{\nu}\lambda_1}(\mathbf{Y})$

$\theta = \frac{1 + \sqrt{1 + 4\theta^2}}{2}$

$\mathbf{F}_{j+1} = \mathbf{Z} + \frac{\theta^{(0)}}{\theta} (\mathbf{Z} - \mathbf{Z}^{(0)})$

 Score = $\frac{1}{LK} \|\mathbf{Z} - \mathbf{Z}^{(0)}\|_F^2$

$\mathbf{Z}^{(0)} = \mathbf{Z}$

$\theta^{(0)} = \theta$

$j = j + 1$

end while

Mise à jour des profils d'activation

La mise à jour des S profils d'activation correspond à S problèmes indépendants, ainsi $\forall s \in \{1, \dots, S\}$ on a :

$$\begin{aligned} & \underset{\mathbf{v}\{s\} \geq 0}{\operatorname{argmin}} \quad \|\mathbf{X}\{s\} - \mathbf{F}_i \mathbf{V}\{s\}^t\|_F^2, \\ \text{t.q.} \quad & \lambda_2 \|\mathbf{v}_{:k}\{S\}\|_1 + \lambda_3 \|\mathbf{\Gamma} \mathbf{v}_{:k}\{S\}\|_1 + \|\mathbf{v}_{:k}\{S\}\|_F^2 \leq 1. \end{aligned} \quad (6.38)$$

Il s'agit d'un problème de régression sous contrainte de fused-lasso. On propose de résoudre ce problème de la même manière que pour le problème de régression sous contrainte Elastic-net dans la Section 5.3.1. Il s'agit d'une procédure de gradient projeté appliquée successivement sur chaque colonne de la matrice $\mathbf{v}_{:k}\{S\}$. La principale différence consiste à remplacer l'étape de projection sur l'ensemble des solutions admissibles de la contrainte Elastic-net par l'espace formé des solutions admissibles de la contrainte fused lasso. Cette projection s'écrit :

$$\begin{aligned} & \underset{\mathbf{y}}{\operatorname{argmin}} \quad \frac{1}{2} \|\bar{\mathbf{y}} - \mathbf{y}\|_F^2, \\ \text{t.q.} \quad & \lambda_2 \|\mathbf{y}\|_1 + \lambda_3 \|\mathbf{\Gamma} \mathbf{y}\|_1 + \|\mathbf{y}\|_F^2 \leq 1. \end{aligned} \quad (6.39)$$

avec :

$$\bar{\mathbf{y}} = \mathbf{v}_{:k}\{S\} - \frac{1}{\mathbf{v}_{:k}^t\{S\} \mathbf{v}_{:k}\{S\}} (\mathbf{V}\{S\} [\mathbf{F}_i^t \mathbf{F}_i]_{:k} - \mathbf{X}^t\{S\} \mathbf{f}_{:k}), \quad (6.40)$$

Le lagrangien du problème de l'Équation (6.39) se note :

$$\mathcal{L}(\mathbf{y}, \gamma) = \frac{1}{2} \|\bar{\mathbf{y}} - \mathbf{y}\|_F^2 + \gamma (\lambda_2 \|\mathbf{y}\|_1 + \lambda_3 \|\mathbf{\Gamma} \mathbf{y}\|_1 + \|\mathbf{y}\|_F^2 - 1). \quad (6.41)$$

Pour γ fixé, on peut montrer que si le vecteur \mathbf{z}^* est solution du problème [73], [141] :

$$\underset{\mathbf{z}}{\operatorname{argmin}} \quad \frac{1}{2} \|\bar{\mathbf{y}} - \mathbf{z}\|_F^2 + \gamma \lambda_3 \|\mathbf{\Gamma} \mathbf{z}\|_1,$$

alors le vecteur \mathbf{y}^* tel que :

$$\begin{aligned} \mathbf{y}^* &= \operatorname{prox}_{\gamma(\lambda_2 \|\cdot\|_1 + \|\cdot\|_F^2)}(\mathbf{z}^*), \\ &= \frac{ST_{\gamma \lambda_2}(\mathbf{z}^*)}{1 + 2\gamma} \end{aligned} \quad (6.42)$$

est solution du problème :

$$\underset{\mathbf{y}}{\operatorname{argmin}} \quad \mathcal{L}(\mathbf{y}, \gamma). \quad (6.43)$$

De nombreux algorithmes permettent de résoudre le problème de l'Équation (5.40). Dans [141], un changement de variables est effectué, remplaçant le problème de l'Équation (5.40) par une régression parcimonieuse pondérée. Il est possible aussi de trouver l'argument qui minimise le dual de ce problème, alors cela revient à calculer une régression des moindres carrés sous une contrainte simple à gérer (chaque variable doit être, en valeur absolue, inférieure ou égale au paramètre de régularisation). On pourra alors utiliser un algorithme de gradient proximal dual [45], [27]. Il est conseillé dans ce cas d'appliquer une procédure d'accélération comme dans l'algorithme FISTA [27]. Une autre solution afin de considérer simultanément le problème primal et dual est d'utiliser le principe de division de Douglas-Rachford [66], qui permet de construire un algorithme similaire à l'ADMM [82]. Malheureusement, comme l'ADMM, cet algorithme introduit un nouveau paramètre dont la valeur risque de fortement impacter sa vitesse de convergence. Finalement des méthodes moins générales, spécialement conçues

pour ce problème ont été développées dans [56]. On retrouve dans le site web de cet auteur¹, un algorithme particulièrement efficace consistant à appliquer en parallèle deux régressions isotoniques bornant la solution optimale de l'Équation (6.42). C'est cet algorithme qui sera utilisé par la suite. Finalement, l'algorithme permettant de résoudre la régression des moindres carrés sous contrainte fused-lasso est fourni dans l'Algorithme 15.

Algorithme 15 Résout la régression sous contrainte fused lasso de l'Équation (6.38)

Require: \mathbf{X} , \mathbf{F}_i , K , le paramètre λ_2 , tolérance $\epsilon > 0$, et j_{max} . Un nombre de décimales de précision a_{max} pour le paramètre γ^* Une première supposition \mathbf{V}_0

```

j = 1
D = F_i^t F_i
while j < j_max or || X - F_i V_j^t ||_F^2 - || X - F_i V_{j+1}^t ||_F^2 > epsilon do
    for k in {1, ..., K} do
        y = (v:k)_j - 1/d_kk (V_j D:k - X^t v:k)
        if h(y, 1) <= 1 then
            v:k = y
        else
            c = 1
            a = 1
            while a <= a_max do
                Calculer y_tilde = TV(y, c lambda_3) en utilisant l'algorithme 2
                y_tilde = NSD_{c lambda_2}(y_tilde) / (1 + 2c)
                if h(y_tilde, gamma) >= 1 then
                    c = c + 10^{-a+1}
                else
                    c = c - 10^{-a+1}
                    a = a + 1
                end if
            end while
            (v:k)_{j+1} = y_tilde
        end if
    end for
    j = j + 1
end while
    
```

Avec :

$$h(y, \gamma) = \gamma \lambda_2 \| \mathbf{y} \|_1 + \gamma \lambda_3 \| \mathbf{\Gamma} \mathbf{y} \|_1 + \gamma \| \mathbf{y} \|_F^2. \quad (6.44)$$

Algorithme joinNMF

L'algorithme général pour résoudre la joinNMF consiste alors à alterner les deux étapes précédentes, son implémentation est fournie dans l'Algorithme 16.

1. <https://lcondat.github.io/>

Algorithm 16 Résout le problème de la joinNMF de l'Équation (6.5)

Require: \mathbf{X} , K , les paramètres λ_1 , λ_2 et λ_3 , la tolérance $\epsilon > 0$, et i_{max} . Une première supposition \mathbf{F}_0 , $\mathbf{V}_0\{1\}, \dots, \mathbf{V}_0\{S\}$.

$i = 1$

while $i < i_{max}$ or $\sum_{s=1}^S \|\mathbf{X}\{s\} - \mathbf{F}_i \mathbf{V}_i^t\{s\}\|_F^2 + \lambda_1 \|\mathbf{F}_i\|_1 - \sum_{s=1}^S \|\mathbf{X}\{s\} - \mathbf{F}_{i+1} \mathbf{V}_{i+1}^t\{s\}\|_F^2 - \lambda_1 \|\mathbf{F}_{i+1}\|_1 > \epsilon$ **do**

Estimer \mathbf{F}_{i+1} en utilisant l'Algorithme 14 avec pour entrée $\bar{\mathbf{X}} = [\sqrt{\zeta(1)}\mathbf{X}\{1\}, \dots, \sqrt{\zeta(S)}\mathbf{X}\{S\}]$, $\bar{\mathbf{V}}_i = [\sqrt{\zeta(1)}\mathbf{V}_i\{1\}^t, \dots, \sqrt{\zeta(S)}\mathbf{V}_i\{S\}^t]$ et λ_1 .

for $s \in \{1, \dots, S\}$ **do**

Estimer $\mathbf{V}_{i+1}\{s\}$ en utilisant l'Algorithme 15 avec $\mathbf{X}\{s\}$ et \mathbf{F}_{i+1} , λ_2 et λ_3

end for

$i = i + 1$

end while

6.A.3 Implémentation de l'algorithme graphNMF (et multNMF)

L'implémentation permettant de trouver une solution au critère de l'Équation (6.22) correspond, comme pour la décomposition joinNMF à alterner deux étapes : la mise à jour des facteurs de CFs \mathcal{F} et la mise à jour des profils d'activation de chaque crise $\mathbf{V}_i\{s\} \forall s \in \{1, \dots, S\}$.

Mise à jour des profil d'activation

À l'itération $i+1$, on estime que la matrice \mathbf{V}_i est connue. Dans ce cas le problème de l'Équation (6.22) consiste à trouver le tenseur \mathcal{F} qui minimise :

$$\operatorname{argmin}_{\mathcal{F} \geq 0} \sum_{s=1}^S \zeta(s) \|\mathbf{X}\{s\} - \mathbf{F}_{::s} \mathbf{V}\{s\}^t\|_F^2 + \eta \|\Delta^{\frac{1}{2}} \mathbf{F}_{(S)}\|_F^2 + \lambda_1 \|\mathbf{F}_{(S)}\|_1, \quad (6.45)$$

Il s'agit de S régularisations de lasso [184], [94], avec un terme de régularisation liant chaque facteur de CFs. On utilise pour résoudre ce problème une méthode de gradient proximal, appliquée de nouveau sur chaque colonne de chaque facteur de CFs $\mathbf{f}_{:ks}$ (de même, il semble qu'appliquer la procédure sur chaque colonne semble converger plus rapidement que d'alterner la procédure sur chaque facteur de CFs $\mathbf{F}_{::s}$). La technique d'accélération FISTA n'est pas applicable ici, car l'opérateur proximal ne peut être utilisé pour mettre intégralement à jour le tenseur \mathcal{F} . En reprenant les notations de la Section 5.3.1 il s'agit d'appliquer la mise à jour suivante :

$$(\mathbf{f}_{:ks})_{j+1} = \operatorname{prox}_{\frac{1}{\nu} h_2(\bullet)}((\mathbf{f}_{:ks})_j - \frac{1}{\nu} \nabla h_1((\mathbf{f}_{:ks})_j)), \quad (6.46)$$

avec :

$$h_1(\mathbf{f}_{:ks}) = \zeta(s) \|\mathbf{X}\{s\} - \mathbf{f}_{:ks} \mathbf{V}\{s\}^t\|_F^2 + \eta \|\Delta^{\frac{1}{2}} \mathbf{F}_{(S)}\|_F^2, \quad (6.47)$$

$$h_2(\mathbf{f}_{:ks}) = \lambda_1 \|\mathbf{f}_{:ks}\|_1. \quad (6.48)$$

Alors, on a de nouveau en considérant la définition du seuillage doux non négatif Équation 5.63.

$$\operatorname{prox}_{\eta \lambda_1 \|\bullet\|_1}(\mathbf{f}_{:ks}) = \operatorname{NSD}_{\eta \lambda_1}(\mathbf{f}_{:ks}). \quad (6.49)$$

Le gradient de la fonction h_1 , s'écrit :

$$\nabla h_1(\mathbf{f}_{:ks}) = 2\zeta_s(\mathbf{F}_{::s})[\mathbf{V}\{s\}^t \mathbf{V}\{s\}]_{:k} - 2\zeta_s \mathbf{X}\{s\} \mathbf{v}_{:k}\{s\} + 2\eta [[\Delta_s: \mathbf{F}_{(S)}]]_{:k}^{L \times K}, \quad (6.50)$$

avec $[[\mathbf{y}]]^{L \times K}$ l'opération qui reconstitue le vecteur $\mathbf{y} \in \mathbb{R}^{LK}$ en matrice de dimension $L \times K$ avec

$$[[\mathbf{y}]]^{L \times K} = \begin{pmatrix} y_1 & \dots & y_K \\ y_{K+1} & \dots & y_{2K} \\ \vdots & & \vdots \\ y_{(L-1)K+1} & \dots & y_{LK} \end{pmatrix} \quad (6.51)$$

Enfin, le calcul du Lipschitzien de ce gradient correspond à :

$$\nu = 2\zeta_s \mathbf{v}_{:k}\{s\}^t \mathbf{v}_{:k}\{s\} + 2\eta \Delta_{ss} \quad (6.52)$$

Ce qui permet de déduire le pas $\frac{1}{\nu}$. On a alors les éléments nécessaires pour mettre à jour le tenseur \mathcal{F} . La procédure permettant de réaliser cette mise à jour est décrite dans l'Algorithme 17

Algorithm 17 Résout la régression de lasso multivues de l'Équation (6.45)

Require: \mathbf{X} , \mathbf{V}_i , K , le paramètre λ_1 , tolérance $\epsilon > 0$, et j_{max} . Une première supposition \mathbf{F}_0

```

j = 1
for s ∈ {1, ..., S} do
    D{s} = V{s}^t V{s}
    for k ∈ {1, ..., K} do
        psk =  $\frac{1}{2\zeta_s \mathbf{d}_{kk}\{s\} + 2\eta \Delta_{ss}}$ 
    end for
end for
while j < jmax or f(Fj, Vit{1}, ..., Vit{S}) - f(Fj+1, Vi+1t{1}, ..., Vi+1t{S}) > ε do
    for s ∈ {1, ..., S} do
        for k ∈ {1, ..., K} do
            y = (f:ks)j - psk(F::s(j) D:k{s} - X v:k{s} - η[[Δs:F(s)]]:kL×K)    voir l'Équation (6.51)
            (f:ks)j+1 = NSD2λ1psk(y)
        end for
    end for
    j = j + 1
end while
    
```

Avec :

$$f(\mathcal{F}, \mathbf{V}^t\{1\}, \dots, \mathbf{V}^t\{S\}) = \sum_{s=1}^S \|\mathbf{X}\{s\} - \mathbf{F}\mathbf{V}^t\{s\}\|_F^2 + \lambda_1 \|\mathbf{F}_{(S)}\|_1 + \eta \|\Delta^{\frac{1}{2}} \mathbf{F}_{(S)}\|_F^2. \quad (6.53)$$

Mise à jour des profils d'activation

Pour mettre à jour les profils d'activation, on applique les mêmes étapes que pour l'algorithme joinNMF. Cette étape est décrite en Section 6.A.2.

Algorithme graphNMF

L'Algorithme 18 résume l'implémentation pour trouver une solution au critère de l'Équation (6.22).

Algorithm 18 Résout le problème graphNMF ($\eta > 0$) et multNMF ($\eta = 0$) de l'Équation (6.5)

Require: \mathbf{X} , K , les paramètres λ_1 , λ_2 et λ_3 , la tolérance $\epsilon > 0$, la matrice laplacienne Δ et i_{max} . Une première supposition \mathcal{F}_0 , $\mathbf{V}_0\{1\}, \dots, \mathbf{V}_0\{S\}$.

```

i = 1
while i < imax or f(Fi, Vit{1}, ..., Vit{S}) - f(Fi+1, Vi+1t{1}, ..., Vi+1t{S}) > ε do
    Estimer Fi+1 en utilisant l'Algorithme 18 avec pour entrée les matrices de CFs
    √ζ1X{1}, ..., √ζsX{S} et les profils d'activation √ζ1V{1}t, ..., √ζsV{S}t ainsi que les paramètres
    η et λ1.
    for s ∈ {1, ..., S} do
        Estimer Vi+1{s} en utilisant l'Algorithme 15 avec X{s} et F::s(i+1), λ2 et λ3
    end for
    i = i + 1
end while
    
```

Avec la fonction f écrite dans l'Équation (6.53).

6.B Lien entre la régularisation multivues et l'analyse canonique des corrélations

Il est possible de démontrer un lien intéressant entre la régularisation multivues modifiée de l'Équation (6.21) et l'analyse canonique des corrélations, fournissant une interprétation supplémentaire à cette régularisation. L'analyse canonique des corrélations consiste, pour deux matrices $\mathbf{X}\{s\}$ et $\mathbf{X}\{s'\}$, à trouver deux transformations linéaires $\mathbf{Z}\{s\} \in \mathbb{R}^{L \times K}$ et $\mathbf{Z}\{s'\} \in \mathbb{R}^{L \times K}$ qui permettent de maximiser la corrélation entre les sous-espaces formés par les matrices $\mathbf{F}_{::s} = \mathbf{X}\{s\}\mathbf{Z}\{s\}$ et $\mathbf{F}_{::s'} = \mathbf{X}\{s'\}\mathbf{Z}\{s'\}$. En d'autres termes, il s'agit de trouver deux transformations qui maximisent la correspondance entre les données. Ces méthodes sont souvent utilisées afin de savoir si les deux jeux de données $\mathbf{X}\{s\}$ et $\mathbf{X}\{s'\}$ décrivent le même phénomène. L'analyse canonique de la corrélation entre les variables $\mathbf{X}\{s\}$ et $\mathbf{X}\{s'\}$ revient donc à trouver les matrices $\mathbf{Z}\{s\}$ et $\mathbf{Z}\{s'\}$ qui minimisent [188], [182] :

$$\begin{aligned} & \underset{\mathbf{Z}\{s\}, \mathbf{Z}\{s'\}}{\operatorname{argmin}} - \operatorname{Tr}(\mathbf{Z}\{s\}^t \mathbf{X}\{s\}^t \mathbf{X}\{s'\} \mathbf{Z}\{s'\}), \\ \text{t.q. } & \mathbf{Z}\{s\}^t \mathbf{X}\{s\}^t \mathbf{X}\{s\} \mathbf{Z}\{s\} = \mathbf{I}_K, \quad \mathbf{Z}\{s'\}^t \mathbf{X}\{s'\}^t \mathbf{X}\{s'\} \mathbf{Z}\{s'\} = \mathbf{I}_K. \end{aligned} \quad (6.54)$$

La contrainte permet de maximiser la corrélation entre les sous-espaces les plus corrélés obtenus à partir d'une transformation linéaire des matrices $\mathbf{X}\{s\}$ et $\mathbf{X}\{s'\}$. Sans perdre en généralité, ce critère s'écrit aussi :

$$\begin{aligned} & \underset{\mathcal{F}}{\operatorname{argmin}} - \operatorname{Tr}(\mathbf{F}_{::s}^t \mathbf{F}_{::s'}), \\ \text{t.q. } & \mathbf{F}_{::s}^t \mathbf{F}_{::s} = \mathbf{I}_K, \quad \mathbf{F}_{::s'}^t \mathbf{F}_{::s'} = \mathbf{I}_K, \\ \text{t.q. } & \mathbf{F}_{::s} = \mathbf{X}\{s\}\mathbf{Z}\{s\}, \quad \mathbf{F}_{::s'} = \mathbf{X}\{s'\}\mathbf{Z}\{s'\}. \end{aligned} \quad (6.55)$$

Une extension de l'analyse canonique des corrélations existe lorsque l'on désire chercher les transformations $\mathbf{Z}\{1\}, \dots, \mathbf{Z}\{S\}$ qui maximisent la correspondance entre chaque paire de matrices $\mathbf{X}\{1\}, \dots, \mathbf{X}\{S\}$ [183], on a dans ce cas :

$$\begin{aligned} & \underset{\mathcal{F}}{\operatorname{argmin}} - \sum_{s, s' \neq s'} a_{ss'} \operatorname{Tr}(\mathbf{F}_{::s}^t \mathbf{F}_{::s'}), \\ \text{t.q. } & \mathbf{F}_{::s}^t \mathbf{F}_{::s} = \mathbf{I}_K, \quad \forall (s) \in \{1, \dots, S\}, \\ \text{t.q. } & \mathbf{F}_{::s} = \mathbf{X}\{s\}\mathbf{Z}\{s\}, \quad \forall (s) \in \{1, \dots, S\}. \end{aligned} \quad (6.56)$$

Avec $a_{ss'}$ un terme permettant de pondérer l'importance de chaque paire de matrices. On précise que ce problème est souvent NP-hard, d'autres formulations de ce problème ont été proposées pour simplifier sa résolution [113], [47]. En contrepartie, les solutions de ces autres formulations sont souvent différentes de celles du problème de l'Équation (6.56).

La fonction de coût du critère de l'Équation (6.56) s'écrit aussi :

$$\sum_{s, s' \neq s'} a_{ss'} \operatorname{Tr}(\mathbf{F}_{::s}^t \mathbf{F}_{::s'}) = \operatorname{Tr}(\mathbf{F}_{(S)}^t \mathbf{A} \mathbf{F}_{(S)}), \quad (6.57)$$

avec \mathbf{A} une matrice d'adjacence telle que $\mathbf{A}_{ss} = 0, \forall (s) \in \{1, \dots, S\}$ et $\mathbf{A}_{ss'} = a_{ss'}$ quand $s' \neq s$. On observe ici un terme proche de la régularisation multivues surtout si l'on observe l'Équation (6.16), à la différence que l'on a ici une matrice d'adjacence au lieu d'une matrice laplacienne. Cette différence est importante. L'utilisation de la fonction de coût de l'analyse canonique des données de l'Équation (6.56) ne peut pas être utilisée comme terme de régularisation, car il s'agit d'une fonction non convexe qui n'a pas de minimum global. Cependant on peut remarquer, en notant \mathbf{D} la matrice de degrés telle que $d_{ss} = \sum_{s'}^S a_{ss'}$, et aux vues des contraintes du critère de l'Équation (6.56), que :

$$\operatorname{Tr}(\mathbf{F}_{(S)}^t \mathbf{D} \mathbf{F}_{(S)}) = \sum_{s=1}^S d_{ss}, \quad \text{si} \quad \mathbf{F}_{::s}^t \mathbf{F}_{::s} = \mathbf{I}, \quad \forall (s) \in \{1, \dots, S\}. \quad (6.58)$$

Alors le terme de l'Équation (6.58) est un terme constant qui n'influence pas la fonction de coût de l'Équation (6.56) si les contraintes sont respectées. On remarque que :

$$-\text{Tr}(\mathbf{F}_{(S)}^t \mathbf{A} \mathbf{F}_{(S)}) + \text{Tr}(\mathbf{F}_{(S)}^t \mathbf{D} \mathbf{F}_{(S)}) = \text{Tr}(\mathbf{F}_{(S)}^t \mathbf{\Delta} \mathbf{F}_{(S)}), \quad (6.59)$$

avec $\mathbf{\Delta} = \mathbf{D} - \mathbf{A}$ la matrice laplacienne formée à partir de la matrice de degrés \mathbf{D} et la matrice d'adjacence \mathbf{A} , le critère de l'analyse multiple des corrélations peut aussi bien s'écrire :

$$\begin{aligned} \underset{\mathcal{F}}{\text{argmin}} \quad & \|\mathbf{\Delta}^{\frac{1}{2}} \mathbf{F}_{(S)}\|_F^2, \\ \text{t.q.} \quad & \mathbf{F}_{::s}^t \mathbf{F}_{::s} = \mathbf{I}_K, \quad \forall (s) \in \{1, \dots, S\}, \end{aligned} \quad (6.60)$$

$$\text{t.q.} \quad \mathbf{F}_{::s} = \mathbf{X}\{s\} \mathbf{Z}\{s\}, \quad \forall (s) \in \{1, \dots, S\}. \quad (6.61)$$

On remarque alors que la fonction de coût de l'analyse canonique des corrélations correspond à la régularisation multivues modifiée. On peut donc voir cette régularisation comme une méthode pour trouver une transformation linéaire des données $\mathbf{X}\{s\}$ (correspondant aux matrices $\mathbf{V}\{s\}$) qui maximise la corrélation paire à paire entre les crises d'un patient.

Chapitre 7

Extraction semi-automatique de réseaux dynamiques décrivant les crises d'épilepsie d'un patient

Des études sur des enregistrements sEEG ont montré que, pendant une crise, l'activité épileptique se propage à travers plusieurs régions anatomiques. Cette propagation se fait de la zone de début de crise vers des zones cérébrales éloignées. Une caractérisation complète et objective de ce réseau variant en fonction du temps est cruciale pour un traitement chirurgical optimal. Ceci est conforme aux études de neurosciences computationnelles qui modélisent les maladies neurologiques en tant que défaut du réseau cérébral ([91], [191], [189], [168], [178], [7]). Cependant, l'extraction d'informations significatives à partir de ces grands ensembles de données, prend du temps et nécessite une grande expertise.

C'est pourquoi on propose, à partir de signaux sEEG, d'appliquer la méthode joinNMF présentée dans le chapitre 6 pour caractériser le réseau épileptogène quantitativement et dans le temps. Dans un premier temps, le calcul des CFs dynamiques est réalisé (on utilise dans cette étude de nouveau la PLV comme mesure de CFs [124], [13]). À partir du réseau dynamique de CFs produit, la méthode joinNMF est appliquée pour extraire des sous-graphes pathologiques avec leurs profils d'activation temporelle spécifiques à chaque crise du patient. On valide la méthode proposée grâce à une étude clinique.

Dans la Section 7.1 on détaille la chaîne de traitement utilisée, notamment le cadre d'utilisation de la méthode joinNMF dans le contexte de cette étude clinique. Dans la Section 7.2, on valide notre approche en l'appliquant sur des données sEEG enregistrées chez 9 patients. Une à 4 crises ont été enregistrées pour chaque patient, avec au total 27 crises étudiées.

7.1 Méthode d'extraction de réseaux caractérisant une crise d'épilepsie

7.1.1 Description de la méthode

Pour chacune des crises d'un patient, les données sont composées d'un enregistrement sEEG. Les crises peuvent avoir des durées différentes. La stratégie proposée peut être résumée en quatre étapes : (a) On découpe chaque enregistrement en segments courts (avec potentiellement du recouvrement), (b) Pour chaque segment, on calcule la CF pour chaque paire d'électrodes. (c) les mesures de CF sont placées dans une liste de matrices représentant, pour chaque crise d'un patient, leurs évolutions temporelles. (d) La liste des matrices représentant les réseaux dynamiques multi-crisis du cerveau est décomposée en sous-graphes de CF caractéristiques d'un patient mais communs à toutes ses crises, ainsi que leurs profils d'activation spécifique à chaque crise. Les principales étapes de la méthode sont illustrées sur la Figure 7.1. Cette procédure de décomposition du réseau dynamique multi-crisis du cerveau est nommée BTND (Brain-wide Time-varying Network Decomposition).

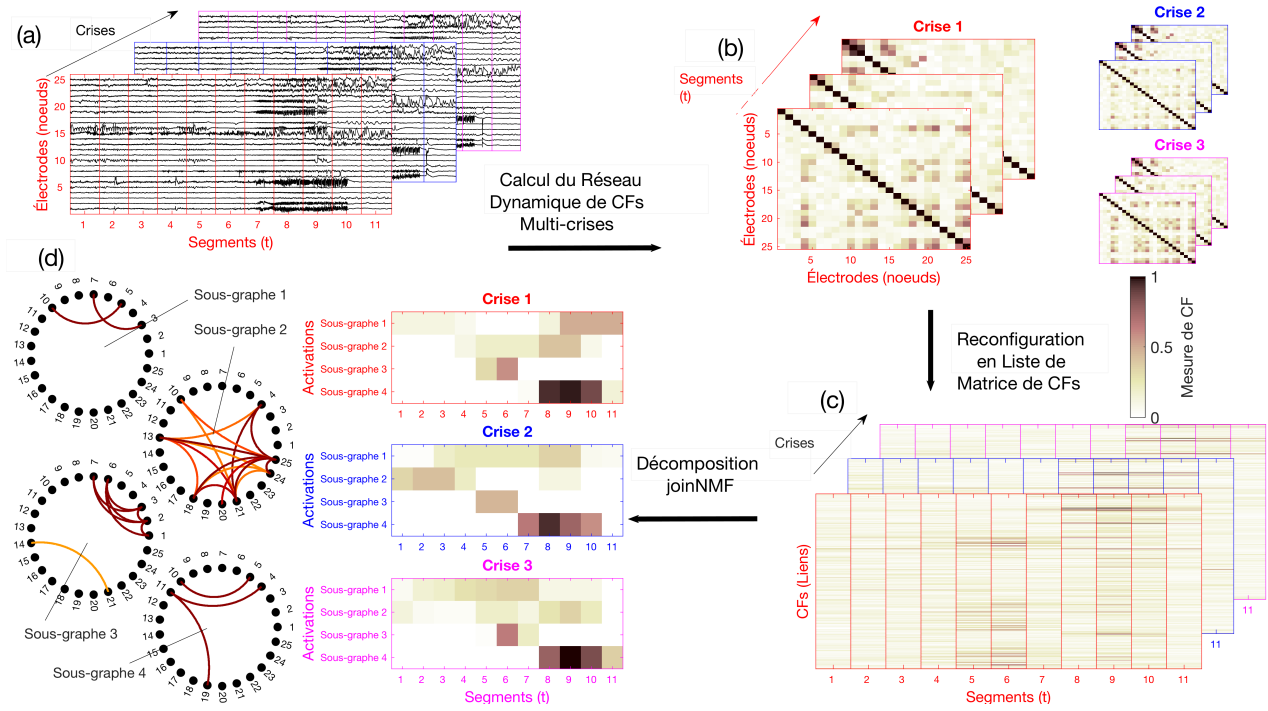


FIGURE 7.1 – Aperçu de la méthode BTND : (a) On découpe chaque enregistrement en segments temporels, avec potentiellement du recouvrement. (b) Pour chaque segment, on calcule pour chaque paire d'électrodes le niveau de connectivité grâce à une mesure de CFs. (c) On réorganise les mesures FC dans un vecteur. Ces dernières, empilées dans le temps, forment pour chaque crise une matrice CF. (d) la liste des matrice de CFs est décomposée en sous-graphes avec leurs profils d'activation temporelle.

7.1.2 La méthode joinNMF pour extraire les sous graphes fonctionnels

Comme dans le chapitre 6, les mesures de CFs sont placées dans une liste de matrices notée $\mathbf{X}_{lt}\{s\}$, correspondant à la CF l au segment temporel t et pour la crise s . On rappelle que la liste de matrices $\mathbf{X}\{s\} \in \mathbb{R}^{L \times T(s)} \forall s(1, \dots, S)$ est une représentation mathématique du réseau dynamique multi-crisis du cerveau. La Figure 7.1.(c) illustre un exemple de liste de matrices $\mathbf{X}\{s\}$.

On décompose cette liste de matrices en appliquant la méthode joinNMF de l'Équation (6.5) présentée dans le chapitre 6. On rappelle que cette décomposition permet d'approximer la liste de matrices telle que $\mathbf{X}\{s\} \approx \mathbf{F}\mathbf{V}^t\{s\}$, avec $\mathbf{F} \in \mathbb{R}^{L \times K}$ les K facteurs de CFs et $\mathbf{V}\{s\} \in \mathbb{R}^{T(s) \times K}$ leurs profils d'activation temporelle spécifiques à la crise s . Les facteurs obtenus contiennent le poids des liens des sous-graphes recherchés, comme on peut le voir sur la Figure 7.1.(d).

En appliquant les recommandations de la Section 6.1.2 pour choisir les paramètres $\zeta(s)$, $\lambda_2(s)$ et $\lambda_3(s)$, il ne reste que trois paramètres à sélectionner : λ_1 et λ_2 contrôlant respectivement la parcimonie des facteurs de CFs et des profils d'activation, λ_3 régulant la compacité temporelle des profils d'activation.

7.1.3 Détails sur l'utilisation de la méthode joinNMF

Initialisation

Comme le problème d'optimisation de l'Équation (6.5) n'est pas convexe, différentes initialisations peuvent conduire à des solutions différentes. Pour contourner cette limitation, on utilise la stratégie habituelle pratiquée par d'autres méthodes d'apprentissage automatique non convexes (comme pour l'algorithme k -means par exemple [103]). Ainsi, on calcule pour différentes initialisations un minimum local de l'Équation (6.5), on en retient uniquement la décomposition qui minimise la fonction de coût considérée. Dans le cadre de l'étude clinique, 20 différentes initialisations semblent suffisantes. On considère à chaque fois, pour l'une des initialisations de \mathbf{F} , comme la matrice singulière gauche issue de

la SVD de la matrice $[\mathbf{X}\{1\}, \mathbf{X}\{2\}, \dots, \mathbf{X}\{S\}]$. Pour les autres initialisations, on utilise des matrices aléatoires dont chaque élément est tiré selon une loi uniforme entre 0 et 1.

Sélection de paramètres

On fixe empiriquement, pour tous les patients, $\lambda_3 = 0.2$. Cette valeur semble suffisante pour obtenir à chaque fois un profil d'activation cohérent. On propose une procédure simple pour sélectionner automatiquement les paramètres λ_1 et λ_2 pour chaque patient. D'abord, comme ces paramètres visent à obtenir des composantes parcimonieuses, on propose de les fusionner en fixant $\lambda_1 = S\lambda_2$ (avec S le nombre de crise). On est maintenant réduit à la sélection d'un paramètre qu'on note $\lambda_2 = \lambda$ pour simplifier.

Comme on a remarqué que pour différentes valeurs de K , l'impact de la régularisation parcimonieuse reste similaire donc on fixe $K = 5$ pour sélectionner le paramètre λ . On cherche alors une valeur de λ tel que le taux de reconstruction $a_\lambda = \sum_{s=1}^S \zeta_s \|\mathbf{X}\{s\} - \mathbf{F}\mathbf{V}^t\{s}\|_F^2$ est suffisamment élevé pour garantir que la plupart des connexions peu actives sont supprimées. Ici, \mathbf{F} et $\mathbf{V}^t\{s\}$ correspondent à la décomposition obtenue à partir de l'Équation (6.5). On précise que a_λ augmente avec λ puisque la régularisation limite la capacité des matrices initiales $\mathbf{X}\{s\}$ à être reconstruites. Alors, la reconstruction la plus fidèle fournit le taux de reconstruction a_0 , qui est associé à des sous-graphes entièrement connectés et complexes. Quand la régularisation est trop forte, les matrices \mathbf{F} et $\mathbf{V}\{s\}$ sont nulles, le taux de reconstruction dans ce cas est noté $a_\infty = \sum_{s=1}^S \zeta_s \|\mathbf{X}\{s\}\|_F^2$. Alors on sélectionne le paramètre λ , tel que 80% de l'information contenue dans la reconstruction la plus fidèle (avec $\lambda = 0$) soit éliminée. C'est-à-dire avec un taux de reconstruction de $a_\lambda \approx 0.8(a_\infty - a_0) + a_0$. Afin de trouver cette approximation, on calcule la solution du problème de l'Équation (6.5) avec λ valant de 0 à 1 par pas de 0.05 (avec 20 initialisations différentes à chaque fois). On choisit un seuil de 80% car cela conduit à ne sélectionner que les connectivités fonctionnelles les plus actives. L'utilisation d'un pourcentage inférieur permettrait à beaucoup de connexions d'arrière-plan d'apparaître. Un pourcentage plus élevé produit des sous-graphes plus parcimonieux ; cependant, ils deviennent moins significatifs en termes de connexions pathologiques.

Pour sélectionner le nombre de modules nécessaires, on calcule pour $K = 3$ à 10 la décomposition avec les paramètres $(\lambda, \gamma; \eta)$ sélectionnés. Le critère "Elbow" [121] et une inspection visuelle de l'activation temporelle des sous-graphes, sont utilisés pour sélectionner le meilleur nombre de composantes K .

Post-traitement

On précise que les facteurs de CF \mathbf{F} , solutions du problème de l'Équation (6.5), ne contiennent pas nécessairement des valeurs comprises entre 0 et 1 ; cela gêne pour les interpréter comme des mesures de connectivité fonctionnelle. De plus, le manque d'échelle commune entre ces facteurs de CF ne permet pas de bien les comparer. Afin de favoriser des résultats interprétables, on effectue, comme étape de post-traitement, la normalisation suivante imposant \mathbf{F} à ne contenir que des valeurs comprises entre 0 et 1. Soit $m_k = \max(\mathbf{f}_{:,k})$ la valeur maximale du facteur k . Les facteurs normalisés qui correspondent aux liens des sous-graphes de CFs, sont $\hat{\mathbf{f}}_{:,k} = \frac{1}{m_k} \mathbf{f}_{:,k}$. Afin de garder des profils d'activation pertinents, on lui transfère le terme m_k tel que $\hat{\mathbf{v}}_{:,k}\{s\} = m_k \mathbf{v}_{:,k}\{s\}$, $\forall s \in \{1, \dots, S\}$. Étant donné que la fonction de régularisation parcimonieuse utilisée est une relaxation de la fonction qui compte explicitement le nombre de CFs non nulles dans chaque facteur (norme ℓ_0), des artefacts peuvent apparaître. On propose alors de diminuer leur présence, en considérant uniquement les liens l tel que $\hat{\mathbf{F}}_{lk} \geq 0.2 \forall k \in \{1, \dots, K\}$.

7.2 Application de la procédure BTND sur 9 patients

7.2.1 Détails sur les données sEEG utilisées

Patients

Pour illustrer la pertinence clinique de la méthode BNTD, on l'applique aux crises enregistrées chez 9 patients épileptiques.

Il s'agit de 9 adultes souffrant d'épilepsie focale résistante aux médicaments, suivis dans le département de neurologie et d'épileptologie aux Hospices Civils de Lyon (hôpital universitaire de Lyon). Les enregistrements sEEG respectent les critères suivants : (i) au moins une crise a été enregistrée au cours du séjour à l'hôpital, (ii) l'analyse visuelle conventionnelle des signaux sEEG a permis de localiser la région déclenchant la crise.

Les détails cliniques de chaque patient sont fournis dans le Tableau 7.1

Parmi eux, 8 patients ont présenté des crises suggérant une atteinte du lobe temporal mais avec des caractéristiques cliniques ou des altérations morphologiques sur l'IRM cérébrale atypiques pour des épilepsies temporales médiales. C'est ce qui a nécessité d'appliquer des enregistrements sEEG. Pour un patient, la sémiologie clinique a suggéré une implication du cortex operculo-insulaire. 3 patients ont subi une résection chirurgicale du cortex épileptogène. Le résultat chirurgical est bon dans chaque cas, après 4 à 48 mois de suivi, les patients n'ont jamais eu de crises (class Engel I-A). Pour deux patients, une thermolésion focale a été réalisée grâce aux électrodes sEEG, entraînant une diminution notable de la maladie. Après 4 à 5 mois de suivi, les patients n'ont jamais eu de crise (class Engel I-A). Pour deux patients, la chirurgie est contre-indiquée en raison d'une implication des deux lobes temporaux lors des crises. Pour deux patients, une résection chirurgicale est prévue sur la base de l'analyse des signaux sEEG, mais n'a pas encore été réalisée au moment de la présente étude.

Enregistrement sEEG

Des électrodes intracérébrales multi-contacts (5 à 15 contacts, de diamètre 0.8 mm, longueur 2 mm et positionnées à 1.5 mm d'intervalle) sont implantées selon la méthode stéréotactique de Talairach [102]. L'emplacement des électrodes a été vérifié par IRM post-implantation. Des enregistrements extra-opératoires prolongés ont été réalisés pour recueillir les crises de chaque patient.

Les données sEEG ont été acquises avec un système de surveillance vidéo et d'EEG de 256 canaux. Il s'agit d'un système d'acquisition Micromed video EEG (SD LTM express, Micromed, Treviso, Italy) utilisant les paramètres suivants : taux d'échantillonnage 256 Hz, filtre passe-haut à 0.15 Hz, filtre passe-bas à 200 Hz, filtre coupe-bande à 50 Hz.

Le nombre médian de contacts bipolaires enregistrés par patient est de 101 (variant de 64 à 130 électrodes). Les principales structures cérébrales ciblées par les électrodes intracrâniennes pour chaque patient sont répertoriées dans le Tableau (7.2). Dans l'ensemble, pour tous les patients, le lobe temporal médial (hippocampe antérieur, hippocampe postérieur, amygdale, cortex entorhinal), le néocortex latéral temporal et le cortex insulaire, ont toujours été ciblés. Selon les résultats électrocliniques de chaque patient, le lobe frontal, le lobe pariétal et le cortex occipital ont été ciblés.

27 crises au total ont été extraites sur les 9 patients. Pour 6 patients, 3 crises ont été analysées. Pour 2 patients, 4 crises ont été analysées et pour un patient, seulement une crise est disponible pour l'analyse. Pour chaque crise, 1 minute d'événements pré-ictaux ont été gardées ainsi que l'intégralité de la crise.

Les durées des crises sont largement hétérogènes, tant au niveau inter qu'intra-individuelles. La durée médiane des 27 crises est de 96 s (allant de 18 s à 337 s).

Analyse des signaux sEEG

On considère le montage bipolaire des signaux sEEG (voir la Section 1.A.1). Un filtre passe-haut, avec une fréquence de coupure égale à 20 Hz a été appliquée sur les signaux sEEG afin de mettre en évidence l'activité haute fréquence typique de l'activité épileptique, en particulier au début de crise [20].

Pour chaque crise, on a calculé les graphes de CF évoluant dans le temps en utilisant la PLV [124], [13]. Pour calculer les graphes de CF, les signaux sEEG ont été segmentés avec des fenêtres glissantes de quatre secondes se déplaçant par pas de 1 seconde.

La méthode BTND est appliquée pour chaque patient, afin de décomposer toutes les crises.

Patient	BRAIN MRI	EEG INTERICTAL	EEG ICTAL	FDG PET	SEMIOLOGY ICTAL	CHIRURGIE	RÉSULTAT CHIRURGICAL
1	atrophie des pôles LT	activité onde lente LT et pointe temporelle	activité LT ictale	hypoM temporal médiale gauche + hypoM néocorticale temporal antérieure gauche	automatismes oro-alimentaires + dystonie droite + perte de conscience	ALT gauche	la (24 m)
2	séquelles post-chirurgica LT	pointes LT&RT avec prédominance gauche	activité ictale LT	hypoM temporal médiale gauche + hypoM médial droit temporal	perte de conscience + arrêt de discours	NA	NA
3	HS droit	pointe temporelle + ralentissement du fond temporel	activité basale ictale RT	temporal médial droit + hypo M pôles RT + hypoM lateral neocortical RT	paresthésie gauche + sensation viscérale ascendante + tachycardie + perte de conscience confusion post-ictale	ATL droite	la (16 m)
4	HS gauche, atrophie du pôle LT	pointes LT&RT avec prédominance gauche	activité LT ictale	pôle LT + hypo M néocortical latérale LT	automatismes oro-alimentaire + ralentissement mental avec une conscience préservée	ATL gauche	la (48 m)
5	lésion temporal postopératoire droite	pointes RT	activité ictale RT	hypoM pôle RT	perte de conscience + posture bilatérale dystonique du bras + automatismes buccaux	NA	NA
6	normal	normal	activité centrale gauche	hypo M périsylvien gauche incluant le gyrus temporal antérieur + pôle temporal + insula	posture tonique bilatérale des bras + déviation droite tête + paresthésies bras droit	thermolésion insulaire opercule gauche	la (4 m)
7	hyperintensité amygdale gauche	pointes LT	activité ictal LT	hypoM pôle LT + hypoM lobe temporal	sensation céphalique + sensation onirique +troubles du langage	lésion thermique temporal médiale gauche	la (5 m)
8	HS droite	activité onde lente RT	activité ictale RT	hypoM temporal médial droit + hypoM néocorticale temporal antérieur droit	automat. oro alimentairet. + conscience préservée + clonus facial gauche	NA	NA
9	HS droite	pointes RT	activité ictale RT	hypoM temporal médiale droite	automatismes verbaux + dysgueusie + perte de conscience	NA	NA

Tableau 7.1 – Détails cliniques de chaque patient. **légende**; HS : Sclérose hippocampique / LT : temporal gauche / RT : temporal droit / hypoM : hypometabolisme / ATL : Lobectomie temporal antérieure / NA : pas enregistré / Le résultat chirurgical est exprimé en classe Engel.

Patient	1	2	3	4	5	6	7	8	9
Nombre de contacts bipolaires	79	116	130	103	101	135	64	85	95
ANT HIPPOC	L & R	L & R	R	L & R	R	L	L	R	L & R
POST HIPPOC	L	L & R	R	L & R	R	NA	L	R	R
AMYGDALA	L & R	R	R	L & R	R	NA	L	R	R
ENTORHINAL CORTEX	NA	R	R	NA	R	NA	L	R	R
TEMPORAL POLE	L	NA	R	L & R	R	L	L	R	L & R
ANT TEMPORAL NEOCORTEX	L & R	L & R	R	L & R	R	L	L	R	L & R
POST TEMPORAL NEOCORTEX	L	L & R	R	L	R	NA	L	R	L & R
INSULA	L	L & R	R	L	R	L	L	NA	R
PRECENTRAL OPERCULUM	NA	NA	R	NA	NA	L	NA	NA	NA
POST CENTRAL OPERCULUM	NA	NA	R	L	NA	L	NA	NA	NA
PARIETAL CORTEX	NA	NA	R	NA	NA	L	NA	NA	NA
OCCIPITAL CORTEX	NA	NA	R	NA	NA	NA	NA	R	NA
OFC	L	L & R	NA	NA	R	NA	NA	NA	NA
ANT CING	NA	NA	NA	L	R	L	NA	NA	NA
LAT FRONTAL CORTEX	NA	NA	NA	NA	R	L	NA	NA	NA

Tableau 7.2 – Liste des principales structures cérébrales ciblées par les contacts sEEG. **Légende** ; L : gauche, R : droite, NA : pas enregistré, ANT HIPPOC : hippocampe antérieur, POST HIPPOC : hippocampe postérieur, OFC : Orbito Frontal Cortex, ANT CING : Anterior Cingulate, LAT FRONTAL : Lateral Frontal Cortex.

7.2.2 Comparaison du réseau dynamique estimé par analyse visuelle conventionnelle et par la méthode BTND

Chaque crise est segmentée en trois périodes majeures via l'analyse visuelle du clinicien : début de crise, propagation et fin de crise. Le début de la crise correspond à une courte période associée à des changements considérables de certains signaux sEEG. Ces changements peuvent être une activité haute fréquence (typiquement supérieure à 20 Hz) et de basse amplitude ou des pointes d'amplitudes élevées qui apparaissent à intervalle régulier. La propagation des crises correspond à une période de temps prolongée. Durant cette période, les comportements atypiques des signaux sEEG ictaux se propagent à plusieurs structures cérébrales, localement ou à distance de la zone de début de crise. Le recrutement de ces régions dans la zone de propagation peut se produire soit par activations indépendantes de zones individuelles, soit par des activations simultanées de plusieurs zones. Enfin, la crise prend fin quand les activations dans les différentes zones touchées se synchronisent et se stabilisent.

Pour chaque période majeure, les électrodes qui ciblent des zones impliquées dans la crise sont déterminées grâce à l'inspection visuelle conventionnelle du clinicien. Les électrodes ont ensuite été regroupées dans plusieurs régions anatomiques prédéfinies.

Pour chaque patient, la méthode BTND fournit le profil d'activation temporel de chaque sous-graphe pour chaque crise. Les sous-graphes qui s'activent dans chaque période majeure sont retenus. Chaque sous-graphe contient plusieurs électrodes fortement connectées. Alors, il est possible de déterminer les régions anatomiques impliquées dans la crise à l'aide des sous-graphes qui sont actifs durant chaque période majeure.

Enfin, une comparaison qualitative entre l'ensemble des structures activées et déterminées par l'analyse visuelle ou par la méthode BTND est réalisée pour chaque crise.

7.2.3 Résultats sur 9 patients

L'organisation des CFs est extraite pour les 27 crises en utilisant la méthode BTND. On rappelle que toutes les crises d'un même patient sont traitées ensembles avec cette méthode. Le Tableau 7.3 fournit la comparaison qualitative entre l'ensemble de la structure retenue par l'analyse visuelle et la méthode BTND pour les patients 1 et 2. De même, les Tableaux 7.4 et 7.5 montrent respectivement la même comparaison qualitative pour les patients 2-4 et les patients 5-9.

En utilisant cette méthode, on a constaté, dans le cas de 6 patients, que $K = 6$ sous-graphes fonctionnels distincts permettent de caractériser l'organisation de leurs crises. $K = 7$ sous-graphes sont utilisés pour 1 patient. $K = 5$ sous-graphes pour un autre et seulement $K = 4$ sous-graphes caractérise la crise du dernier patient. Cependant, pour chaque patient, certains sous-graphes sont fortement actifs tout le long de l'enregistrement (ou avant la crise). Ces sous-graphes sont considérés comme non spécifiques. Pour 5 patients, 2 sous-graphes sont non-spécifiques alors que pour 4 patients il n'y en a qu'un.

Au début de la crise, un seul sous-graphe est actif pour 1 patient, deux sous-graphes sont actifs pour 4 patients, 3 sous-graphes sont actifs pour 3 patients et 5 sous-graphes sont actifs pour 1 patient. Pour 6 des patients, soit 24 crises au total, le début de la crise déterminé par analyse visuelle coïncide avec le réseau divulgué par la méthode BTND. Pour ces patients, les régions corticales de la zone de début de crise déterminées par analyse visuelle, sont incluses dans les sous-graphes de début de crise. Cependant, les sous-graphes de début de crises peuvent inclure aussi d'autres régions qui peuvent être externes à la zone déclenchant la crise. Pour l'un de ces 6 patients (Pt 2), le début de la crise implique soit le lobe temporal médial gauche, soit le lobe temporal médial droit. Les sous-graphes de début de crise sont différents pour chaque type de latéralisation, et sont concordants avec l'analyse visuelle. Pour 3 crises chez un patient (Pt 9), le sous-graphe de début de crise était clairement discordant avec l'observation visuelle. Pour ce patient, la zone de début de crise implique le lobe temporal médial droit ou le lobe temporal médial gauche selon la crise. Les sous-graphes de début de crise pour ce patient latéralisent incorrectement le lobe temporal droit ou gauche.

Chez tous les patients, au cours de la propagation des crises, il y a toujours une correspondance spatiale étroite entre les sous-graphes activés durant cette période et l'analyse visuelle. Pour les 27 crises des 9 patients, les régions cérébrales impliquées dans la propagation des crises sont incluses dans les sous-graphes activés. Cependant, la congruence entre les sous-graphes activés et les régions révélées par l'analyse visuelle n'est pas toujours parfaite : quelques régions sont parfois révélées par la méthode BTND mais pas détectées par l'analyse visuelle.

Pendant la fin de la crise, il y a aussi une forte correspondance entre les sous-graphes actifs durant cette période et l'analyse visuelle. Pour 23 crises, les régions cérébrales impliquées dans la fin de la crise sont incluses dans les sous-graphes actifs. Pour 4 crises chez un même patient (Pt 4), l'inspection visuelle divulgue plus de régions actives que la méthode BTND.

Les résultats détaillés sont maintenant présentés pour deux patients (Pt 1 et Pt 2)

Cas du patient 1

Pt 1 est un homme âgé de 49 ans. Des examens pré-chirurgicaux non invasifs ont suggéré une épilepsie du lobe temporal gauche, mais certaines caractéristiques radiologiques ont été considérées comme atypiques pour ce type d'épilepsie du lobe temporal mésial. L'enregistrement sEEG cible plusieurs régions à l'intérieur du lobe temporal gauche (hippocampe antérieur, hippocampe postérieur, amygdale, pôle temporal, néocortex temporal antérieur, lobe temporal postérieur), du cortex frontal orbito gauche et du lobe temporal droit (amygdale droite, néocortex temporal antérieur droit).

3 crises ont pu être enregistrées de cette manière. L'activité initiale de la crise se développe dans l'hippocampe antérieur et postérieur gauche avec une implication secondaire du pôle temporal, du noyau amygdalien et du néocortex temporal antérieur gauche à la fin des crises.

La méthode BTND appliquée aux trois crises a permis d'inférer 6 sous-graphes. La Figure 7.2 montre l'enregistrement de deux crises (crise 1 et 2) du patient 1 pour les électrodes sélectionnées. Sous chaque enregistrement, on fournit les profils d'activation spécifique de tous les sous-graphes obtenus par la BTND. En haut, sont représentées les principales structures cérébrales ciblées par les électrodes

intracrâniennes pour ce patient. La Figure 7.3 montre les 6 sous-graphes FC révélés.

Un sous-graphe était actif avant et durant la crise. Ce sous-graphe était principalement composé de connexions locales dans le lobe temporal (principalement dans l'hippocampe antérieur, l'hippocampe postérieur, l'amygdale). Pendant les premières secondes de la crise, il y a une activation reproductible d'un sous-graphe, qui implique principalement des connexions entre l'hippocampe antérieur et l'amygdale, l'hippocampe postérieur et le néocortex temporal postérieur. Quelques secondes plus tard, une forte activation d'un autre sous-graphe est observée impliquant principalement des connexions entre le lobe temporal médial et le pôle temporal. Durant l'étape de propagation, il y a une activation cohérente de trois autres sous-graphes, leurs activations sont très similaires entre les crises. Ces sous-graphes contiennent principalement des connexions entre le néocortex temporal latéral antérieur, le néocortex temporal postérieur et le pôle temporal. Dans l'ensemble, le schéma des activations est très similaire pour les trois crises. Ce schéma est très cohérent avec l'analyse visuelle des crises. Le patient a subi une lobectomie antérieure gauche qui a entraîné une suppression des crises après plus de 24 mois de suivi.

Cas du patient 2

Pt 2 est une femme de 37 ans. Des examens pré-chirurgicaux non invasifs ont suggéré que les deux lobes temporaux pouvaient déclencher la crise d'épilepsie. Les enregistrements sEEG sont donc nécessaires dans ce cas pour évaluer l'épileptogénicité intrinsèque de chaque lobe temporal. Les électrodes sEEG intracrâniennes ciblent principalement les lobes temporaux médiaux et latéraux (hippocampe antérieur gauche et droit, pôle temporal gauche et droit, amygdale droite, néocortex temporal antérieur gauche et droit, néocortex temporal postérieur gauche et droit, insula gauche et droite), mais aussi gauche et le cortex orbitofrontal droit.

3 crises ont été enregistrées pour ce patient. La crise 1, se caractérise par une décharge rapide dans l'hippocampe antérieur et postérieur gauche avec une propagation secondaire de cette activité haute fréquence au cortex frontal latéral gauche et orbitofrontal gauche. À la fin de la propagation de la crise, cette activité est transférée au lobe temporal droit (hippocampe droit et amygdale droite). Pour deux crises (crises 2 et 3), l'activité initiale de début de crise se développe dans l'hippocampe droit, l'amygdale droite et le cortex entorhinal droit avec une propagation ictale secondaire vers le cortex temporal antérieur droit et avec une propagation vers le lobe temporal gauche à la fin de la crise.

La méthode BTND appliquée aux trois crises a permis d'inférer 7 sous-graphes. La Figure 7.4 montre l'enregistrement de deux crises (crise 1 et 2) du patient 2 pour les électrodes sélectionnées. Sous chaque enregistrement, on fournit les profils d'activation spécifiques de tous les sous-graphes obtenus. En haut, sont représentées les principales structures cérébrales ciblées par les électrodes intracrâniennes pour ce patient. La Figure 7.5 montre les 7 sous-graphes FC révélés.

Un sous-graphe est principalement composé de connexions au sein du lobe temporal médial gauche et du cortex orbitofrontal gauche. Un autre sous-graphe est principalement composé de connexions à l'intérieur du néocortex temporal antérieur gauche. Les cinq sous-graphes restants sont constitués de régions reliant principalement des structures temporales médiales droites (hippocampe, amygdale, cortex entorhinal) et du néocortex temporal latéral droit. L'évolution dans le temps de l'activation de ces sous-graphes est étroitement liée à l'implication ictale des deux lobes temporaux révélés par l'analyse visuelle des crises : l'atteinte du lobe temporal médian gauche a été accompagnée d'une activation des sous-graphes 1 et 2 ciblant cette région en temps opportun. Au contraire, lorsque le lobe temporal médian droit est atteint en début de crise, on observe une activation des sous-graphes 3 et 4 qui cible des régions de l'hémisphère droit. Comme l'étude de la sEEG a révélée une épileptogénicité intrinsèque des deux lobes temporaux, la résection chirurgicale était contre-indiquée.

7.3 Conclusion

Pour cette étude, la méthode joinNMF a été intégrée à une chaîne de traitement d'enregistrement sEEG afin de décrire les structures de connectivité des crises d'épilepsie.

Dans l'ensemble, on a constaté que cette méthode produit plusieurs sous-graphes de connexions, avec leurs profils d'activation temporelle, qui correspondent étroitement au modèle visuel de propagation des crises. Dans 24 des 27 crises, un ou deux sous-graphes correspondent clairement à la zone de

début de crise. Cela suggère que cette méthode peut aider à localiser la zone de début de crise si une augmentation de la synchronie entre ces régions est révélée par la mesure de CF. Cette découverte confirme plusieurs études démontrant que les modulations focales de la synchronie aident à localiser la zone d'apparition des crises à réséquer chirurgicalement [192], [21]. De plus, grâce aux contraintes de parcimonie et de compacité, la méthode révèle de manière intelligible comment la connectivité se propage aux différentes régions du cerveau. Dans l'ensemble, la méthode résume une grande quantité d'interactions complexes mais facilement interprétables. On pense que la BTND est pertinente dans les études cliniques axées sur la connectivité dans le cerveau de patients épileptiques. Par exemple, cela peut aider à démêler les bases neurales de la sémiologie clinique des crises, qui sont souvent liées à l'étude de la propagation de la crise dans des zones d'intérêts du cerveau. Aussi, l'interprétation clinique des signaux sEEG est principalement axée sur des changements de puissance dans les principales bandes de fréquences révélées par l'inspection visuelle. Généralement, le début de la crise est associé aux hautes fréquences au-dessus de 20 Hz, tandis que la propagation est plutôt étudiée à partir des bandes delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz). La méthode BTND est capable de détecter les étapes de propagations bien que les mesures de CFs soient uniquement appliquées sur les hautes fréquences au-dessus de 20 Hz. À cet égard, on pense que la méthode peut ainsi contribuer à apporter dans la pratique clinique des mesures complémentaires pour l'interprétation visuelle des crises.

Cependant, plusieurs limitations doivent être considérées. Premièrement, la méthode joinNMF possède 3 paramètres qui impactent la qualité des sous-graphes et des activations obtenues. On a exposé une procédure simple pour choisir chaque paramètre qui produit des résultats pertinents pour chacune des 27 crises. Cependant, la méthode de sélection de ces paramètres peut dépendre de la mesure de CF et doit potentiellement être adaptée pour d'autres applications cliniques. Deuxièmement, pour un patient, la méthode BTND a identifié des sous-graphes qui étaient discordants avec la zone de début de crise. Pour ce patient, au début de la crise, il y a une décharge rapide dans la zone de début de crise (sans aucun changement focal de synchronie) et une activité de pointe dans l'hémisphère controlatéral (accompagnée d'une augmentation de la synchronie). L'augmentation de la CF évaluée par la valeur de PLV ne permet pas de révéler correctement cette décharge et cette activité de pointe. Cela souligne que la méthode BTND dépend fortement de la mesure de CF utilisée pour inférer le réseau dynamique multi-crisis. Ainsi, l'utilisation d'une autre mesure de CFs pourrait être envisagée pour ce patient. Troisièmement, plusieurs études ont déjà montré que l'apparition des crises est souvent marquée par une diminution considérable de la connectivité parmi les structures cérébrales enregistrées, tandis que la synchronie augmente progressivement au cours de la crise [199], [107]. On pourrait donc s'attendre à ce que le début de crise puisse être caractérisé par une diminution de l'activation des sous-graphes situés dans la zone de début de crise. Il est important de considérer que la méthode BTND ne met en évidence que les sous-graphes fonctionnels associés à des valeurs élevées de connectivité fonctionnelle. Ainsi, cette méthode permet seulement de montrer les activations de synchronie entre différentes régions du cerveau. Une extension intéressante serait de permettre à la méthode BTND de révéler les désactivations qui peuvent se produire dans le cerveau au début de la crise lorsque les activités sEEG de différentes zones du cerveau sont soudainement décorréliées. Finalement, une validation directe de la méthode est hors de portée car il n'existe pas d'étalon parfait pour estimer le schéma de connectivité des patients épileptiques. On a choisi de faire une correspondance entre la propagation révélée par la méthode BNTD et l'interprétation visuelle du clinicien. Il s'agit donc d'une étude qualitative. Cependant, il existe des modèles neuronaux d'activité épileptique qui simulent la propagation d'une crise [200]. Ces modèles pourraient être utilisés pour pouvoir mieux quantifier les performances de la méthode proposée.

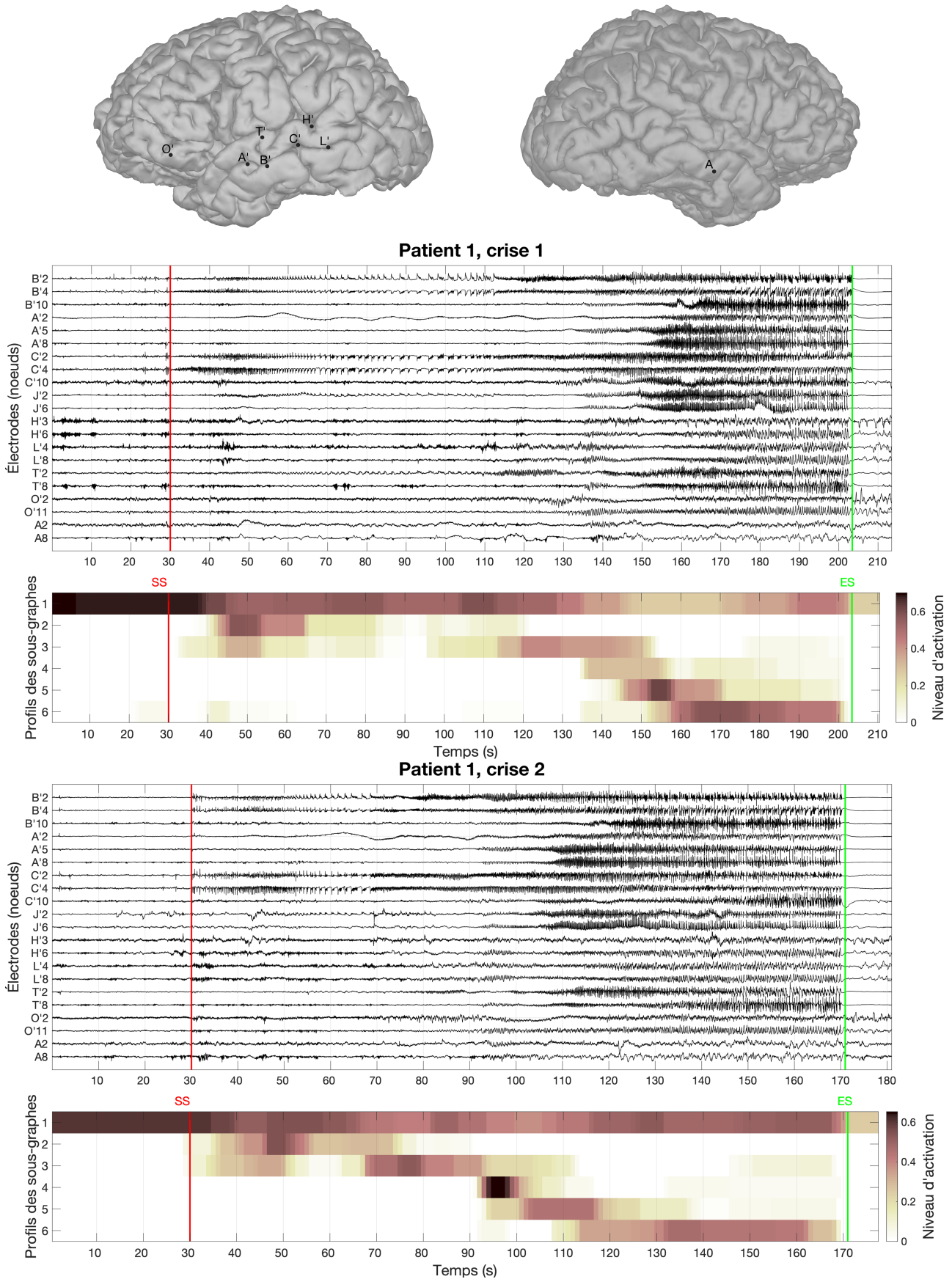
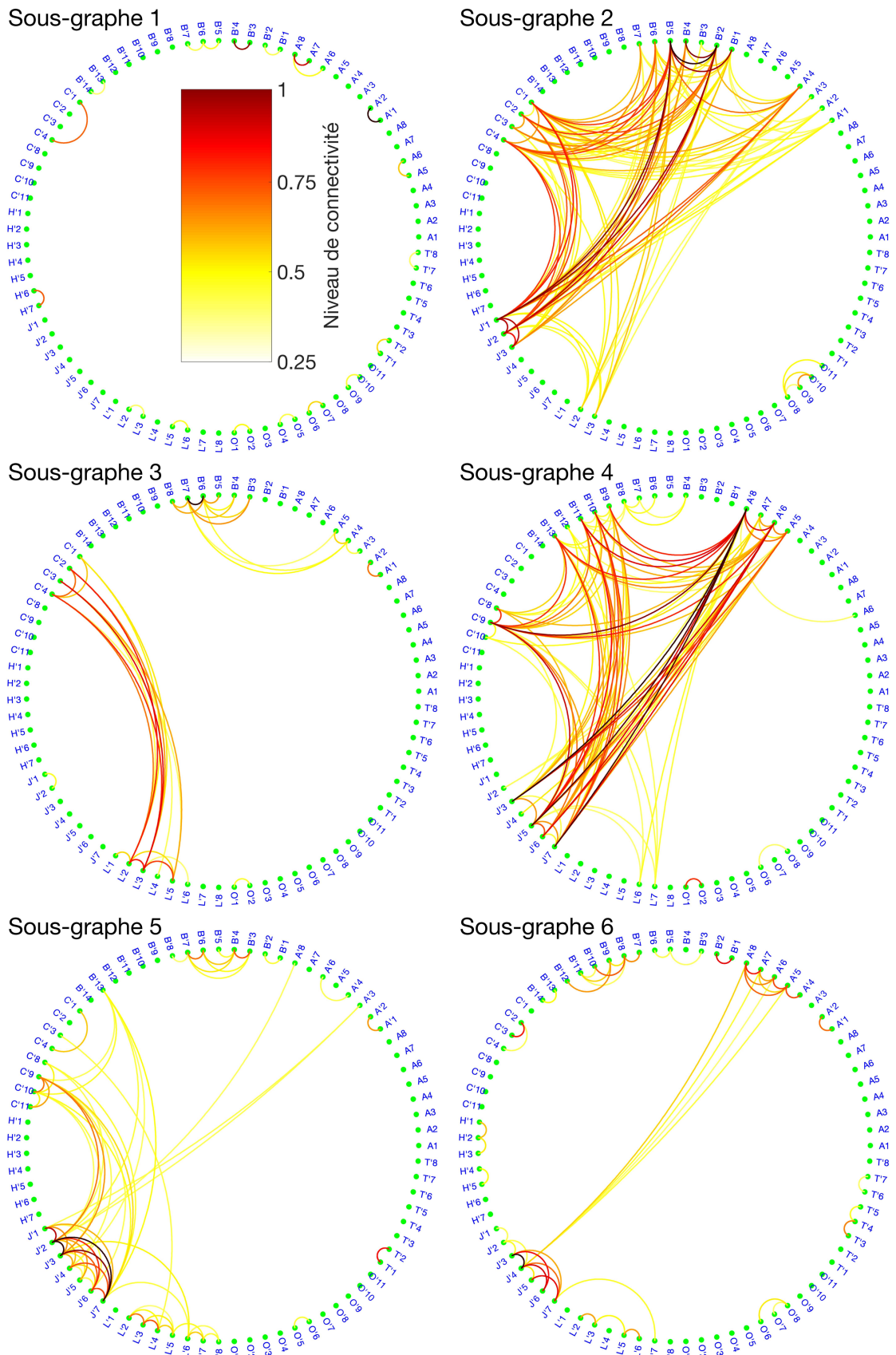


FIGURE 7.2 – En haut, sont représentées les principales structures cérébrales ciblées par les électrodes intracrâniennes pour le patient 1. Ensuite, on montre l'enregistrement de deux crises (crise 1 et 2) pour un sous-ensemble d'électrodes sélectionnées. Sous chaque enregistrement, on fournit les profils d'activation spécifiques de tous les sous-graphes obtenus. Les lignes rouges et vertes correspondent respectivement aux instants du début et de fin de crise.



Legend ; B'2 : hippocampe gauche B'4 : hippocampe gauche B'10 : hippocampe gauche A'2 : amygdale gauche A'5 : amygdale gauche A'8 : gyrus temporal moyen gauche
 C'2 : hippocampe postérieur gauche C'4 : hippocampe postérieur gauche J'2 : gyrus temporal antérieur gauche J'6 : pôle temporal gauche J'6 : pôle temporal gauche
 H'3 : insula postérieure gauche H'6 : opercule temporal gauche L'4 : néocortex temporal postérieur gauche L'8 : néocortex temporal postérieur gauche T'2 : insula antérieure gauche
 T'8 : gyrus temporal supérieur gauche O'2 : cortex orbitofrontal gauche O'11 : orbitale gauche cortex frontal
 A'2 : amygdale droite A'8 : gyrus temporal moyen droit.

FIGURE 7.3 – Les 6 sous-graphes révélés par la méthode BTND pour le patient 1.

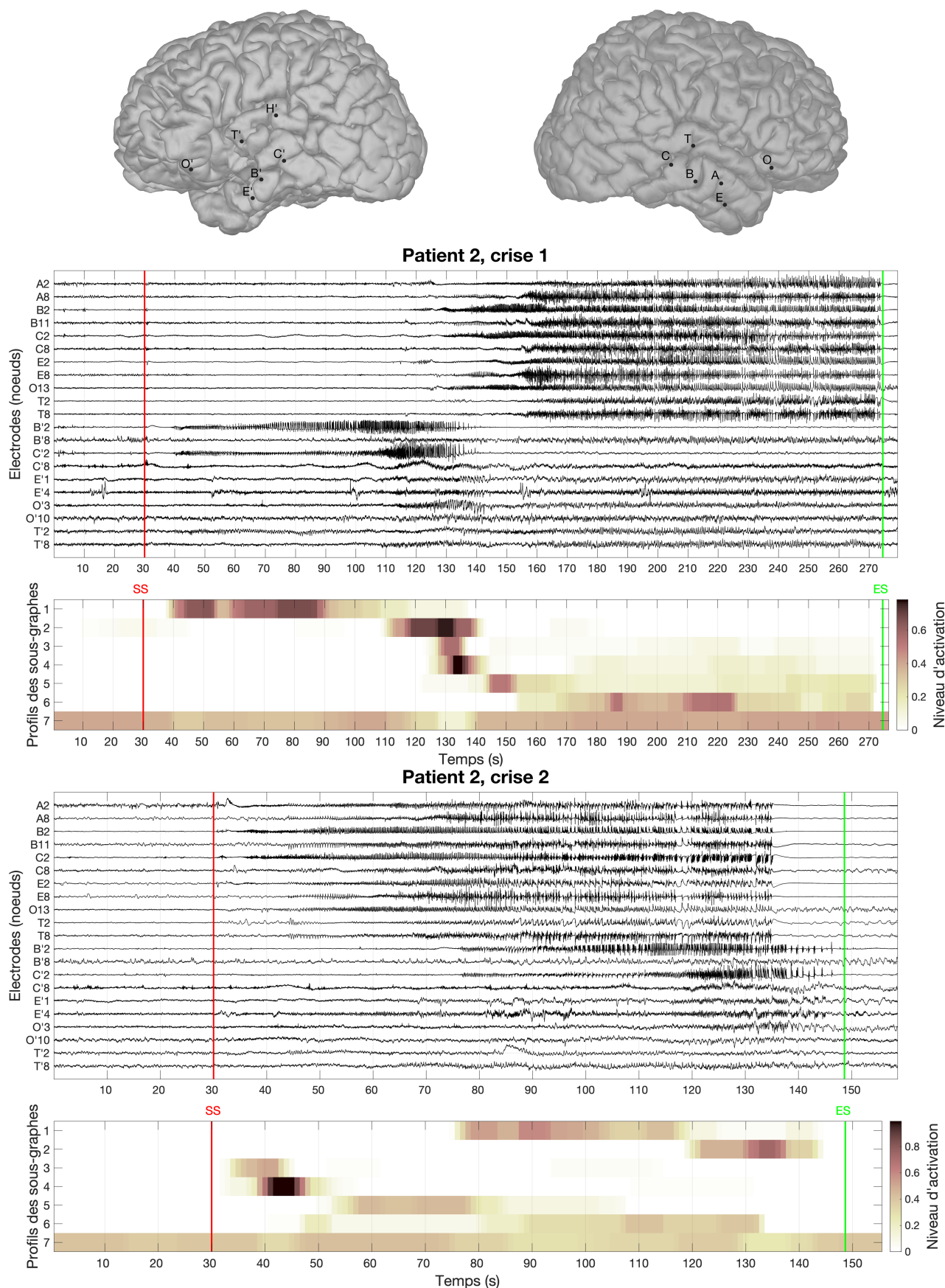


FIGURE 7.4 – En haut, sont représentées les principales structures cérébrales ciblées par les électrodes intracrâniennes pour le patient 2. Ensuite, on montre l'enregistrement de deux crises (crise 1 et 2) pour un sous-ensemble d'électrodes sélectionnées. Sous chaque enregistrement, on fournit les profils d'activation spécifiques de tous les sous-graphes obtenus. Les lignes rouges et vertes correspondent respectivement aux instants du début et de fin de crise.

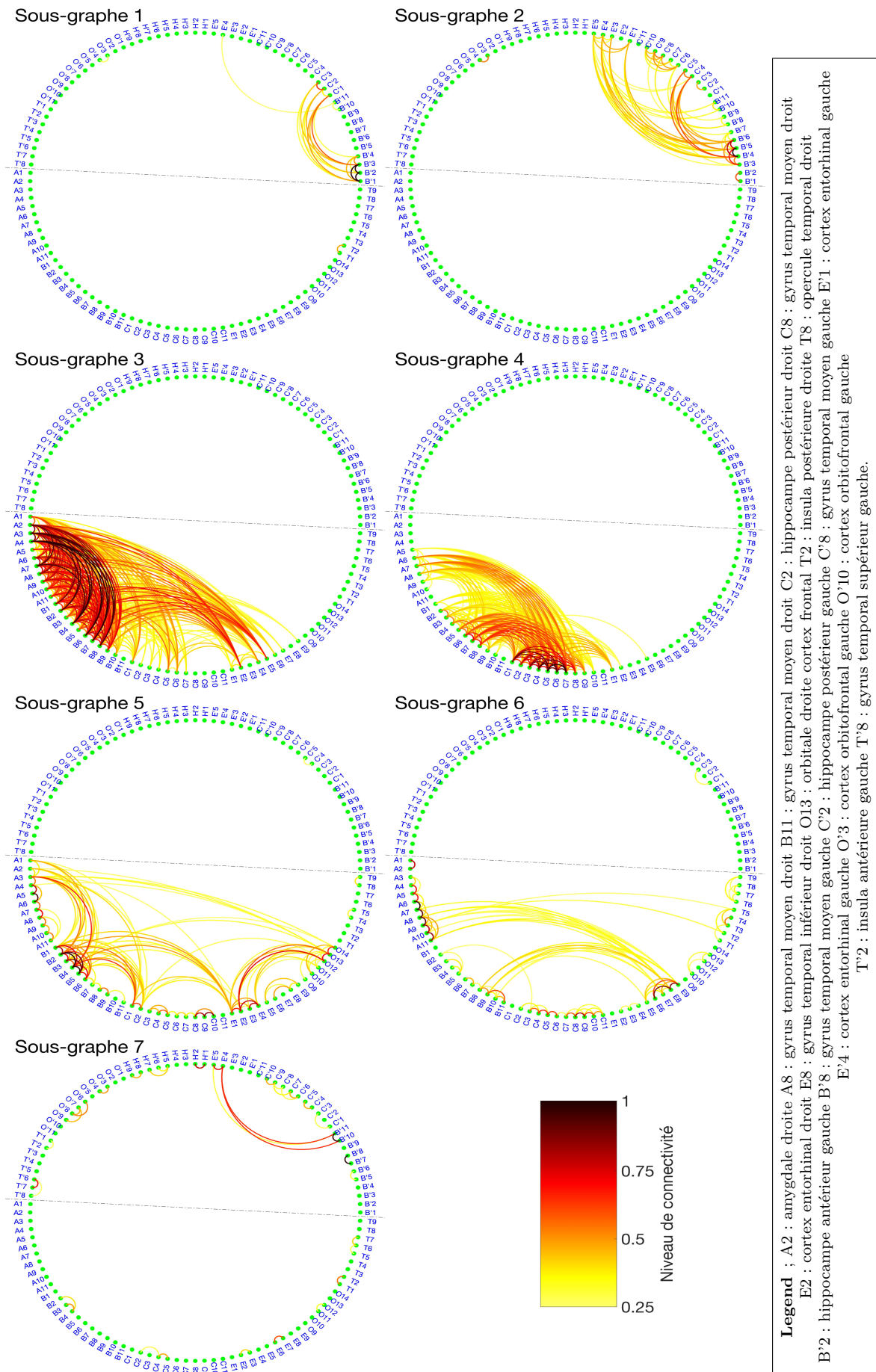


FIGURE 7.5 – Les 7 sous-graphes révélés par la méthode BTND pour le patient 2. La ligne pointillée sépare les électrodes situées à l'hémisphère droit et à l'hémisphère gauche du cerveau.

		DÉBUT CRISE	PROPAGATION	FIN CRISE
Patient 1	Crise 1	Clinique :L ANT HIPPOC + L POST HIPPOC Méthode : subg3 + (subg1)	Clinique : L ANT HIPPOC + L POST HIPPOC + L TEMP POLE Méthode : subg3 + subg2 + (subg1)	Clinique : L ANT HIPPOC + L POST HIPPOC + L TEMP POLE + L ANT TEMP NEOCORTEX Méthode : subg4 + subg5 + subg6 + (subg1)
	Crise 2	Clinique :L ANT HIPPOC + L POST HIPPOC Méthode : subg3 + (subg1)	Clinique : L ANT HIPPOC + L POST HIPPOC + L TEMP POLE Méthode : subg3 + subg2 + (subg1)	Clinique : L ANT HIPPOC + L POST HIPPOC + L TEMP POLE + L ANT TEMP NEOCORTEX Méthode : subg4 + subg5 + subg6 + (subg1)
	Crise 3	Clinique :L ANT HIPPOC + L POST HIPPOC Méthode : subg3 + (subg1)	Clinique : L ANT HIPPOC + L POST HIPPOC + L TEMP POLE Méthode : subg3 + subg2 + (subg1)	Clinique : L ANT HIPPOC + L POST HIPPOC + L TEMP POLE + L ANT TEMP NEOCORTEX Méthode : subg4 + subg5 + subg6 + (subg1)
	Sous-graphes	subg1 : L ANT HIPPOC + L AMYG, subg2 : L ANT HIPPOC + L POST HIPPOC + L TEMPORAL POLE + L AMYG, subg3 : L ANT HIPPOC + L POST HIPPOC + L POST TEMPORAL NEOCORTEX + L AMYG, subg4 : L POST TEMPORAL NEOCORTEX + L TEMPORAL POLE + L ANT HIPPOC + L POST HIPPOC subg5 : L ANT TEMPORAL NEOCORTEX + L TEMP POLE, subg6 : L TEMPORAL POLE + L ANT TEMPORAL NEOCORTEX.		
Patient 2	Crise 1	Clinique : L ANT HIPPOC + L POST HIPPOC Méthode : subg1 + (subg7)	Clinique : L ANT HIPPOC + L POST HIPPOC + L ANT TEMPORAL NEOCORTEX + L ORBITO FRONTAL NEOCORTEX Méthode : subg2 + (subg7)	Clinique : R ANT HIPPOC + R POST HIPPOC + R AMYG Méthode :subg3 + subg4 + subg5 + subg6 + (subg7)
	Crise 2	Clinique : R ANT HIPPOC + R POST HIPPOC + R AMYG + R ENTORHINAL CORTEX Méthode : subg3 + subg4 + (subg7)	Clinique : R ANT HIPPOC + R POST HIPPOC + R AMYG + R ENTORHINAL CORTEX + R ANT TEMPORAL NEOCORTEX Méthode : subg5 + (subg7)	Clinique : L ANT HIPPOC + L POST HIPPOC Méthode : subg1 + subg2 + subg6 + (subg7)
	Crise 3	Clinique : R ANT HIPPOC + R POST HIPPOC + R AMYG + R ENTORHINAL CORTEX Méthode : subg3 + subg4 + (subg7)	Clinique : R ANT HIPPOC + R POST HIPPOC + R AMYG + R ENTORHINAL CORTEX + R ANT TEMPORAL NEOCORTEX Méthode : subg5 + (subg7)	Clinique : L ANT HIPPOC + L POST HIPPOC Méthode : subg1 + subg5 + subg6 + (subg7)
	Sous-graphes	subg1 : L ANT HIPPOC + L POST HIPPOC, subg2 : L ANT HIPPOC + L POST HIPPOC + L ENTORHINAL CORTEX + L ORBITO FRONTAL CORTEX, subg3 : R ANT HIPPOC + R POST HIPOC + R AMYG + R ENTORHINAL CORTEX, subg4 : R ANT HIPPOC + R POST HIPOC + R AMYG + R ENTORHINAL CORTEX, subg5 : R ANT HIPPOC + R POST HIPPOC + R ENTORHINAL CORTEX + R ORBITO FRONTAL CORTEX + R ANT TEMPORAL NEOCORTEX, subg6 : R ANT TEMPORAL NEOCORTEX + R POST TEMPORAL NEOCORTEX, subg7 : L ANT TEMPORAL NEOCORTEX		
Legend ; subg : sous-graphe, Crise : crise, L : droit, R : gauche, ANT HIPPOC : hippocampe antérieure, POST HIPPOC : hippocampe postérieure, ANT TEMPORAL NEOCORTEX : néocortex temporal antérieur, POST TEMPORAL NEOCORTEX : néocortex temporal postérieur, AMYG : amygdala.				

Tableau 7.3 – Comparaison qualitative entre l'ensemble des structures activées, déterminées par analyse visuelle et la méthode BTND pour les patients 1 et 2

	DÉBUT CRISE	PROPAGATION	FIN CRISE
Patient 3	Crise 1	Clinique : R ANT HIPPOC Méthode : (subg2)	Clinique : R ANT HIPPOC + R POST HIPPOC + R ENT CX + R TEMPORAL LATERAL + R OCCIPITAL CX + R PARIETAL CX + R POST CENTRAL OPERC Méthode : subg3 + subg4 + subg5 + subg6
	Crise 2	Clinique : R ANT HIPPOC + R POST HIPPOC Méthode : (subg2) + subg3 + subg5	Clinique : R ANT HIPPOC + R POST HIPPOC + R ENT CX + R TEMPORAL LATERAL + R OCCIPITAL CX + R PARIETAL CX + R POST CENTRAL OPERC Méthode : subg3 + subg4 + subg5 + subg6
	Crise 3	Clinique : R ANT HIPPOC + R POST HIPPOC Méthode : (subg2) + subg3 + subg5	Clinique : R ANT HIPPOC + R POST HIPPOC + R ENT CX + R TEMPORAL LATERAL + R OCCIPITAL CX + R PARIETAL CX + R POST CENTRAL OPERC Méthode : (subg2) + subg4
	Sous-graphes	subg1 : R ANT HIPPOC + R POST HIPPOC + R POST HIPPOC + R ENT CX, subg2 : R ANT HIPPOC, subg3 : R PARIETAL CX + R OCCIPITAL CX + R TEMP LATERAL, subg5 : R TEMP LATERAL + R POST CENTRAL OPERC, subg6 : R TEMP LATERAL,	
Patient 4	Crise 1	Clinique : L ANT HIPPOC + L POST HIPPOC + L TEMP POLE Méthode : subg2 + (subg1)	Clinique : L ANT HIPPOC + L POST HIPPOC + L TEMP POLE + L POST TEMP NEOCORTEX + R AMYG + R ANT HIPPOC Méthode : subg3 + subg4 + subg5
	Crise 2	Clinique : L ANT HIPPOC + L POST HIPPOC + L TEMP POLE Méthode : subg2 + (subg1)	Clinique : L ANT HIPPOC + L POST HIPPOC + L TEMP POLE + L POST TEMP NEOCORTEX + R AMYG + R ANT HIPPOC Méthode : subg3 + subg4 + subg5
	Crise 3	Clinique : L ANT HIPPOC + L POST HIPPOC + L TEMP POLE Méthode : subg2 + (subg1)	Clinique : L ANT HIPPOC + L POST HIPPOC + L TEMP POLE + L POST TEMP NEOCORTEX + R AMYG + R ANT HIPPOC Méthode : subg3 + subg4 + subg5
	Crise 4	Clinique : L ANT HIPPOC + L POST HIPPOC + L TEMP POLE Méthode : subg2 + (subg1)	Clinique : L ANT HIPPOC + L POST HIPPOC + L TEMP POLE + L POST TEMP NEOCORTEX + R AMYG + R ANT HIPPOC Méthode : subg3 + subg4 + subg5
	Sous-graphes	subg1 : L ANT TEMP NEOCORTEX + R ANT HIPPOC, subg2 : L ANT HIPPOC + L POST HIPPOC, subg3 : L POST TEMPORAL NEOCORTEX + L AMYGDALA + L ANTERIOR CINGULATE, subg4 : R AMYGD + R ANT HIPPOC + L ANT HIPPOC + L AMYG, subg5 : R AMYGD + R ANT HIPPOC LATERAL + L TEMP POST	
Patient 5	Crise 1	Clinique : R ANT HIPPOC + R POST HIPPOC + R TEMP POLE + R AMYG Méthode : subg3 + (subg1)	Clinique : R ANT HIPPOC + R POST HIPPOC + R TEMP POLE + R AMYG Méthode : subg3 + (subg4) + (subg1)
	Crise 2	Clinique : R ANT HIPPOC + R POST HIPPOC + R TEMP POLE + R AMYG Méthode : subg3 + (subg1)	Clinique : R ANT HIPPOC + R POST HIPPOC + R TEMP POLE + R AMYG + R POST TEMP NEOCORTEX Méthode : subg3 + (subg4) + (subg1)
	Crise 3	Clinique : R ANT HIPPOC + R POST HIPPOC + R TEMP POLE + R AMYG Méthode : subg3 + (subg1)	Clinique : R ANT HIPPOC + R POST HIPPOC + R TEMP POLE + R AMYG + R POST TEMP NEOCORTEX R ORBITO FRONTAL CX Méthode : subg3 + (subg4) + (subg1)
	Crise 4	Clinique : R ANT HIPPOC + R POST HIPPOC + R TEMP POLE + R AMYG Méthode : subg3 + (subg1)	Clinique : R ANT HIPPOC + R POST HIPPOC + R TEMP POLE + R AMYG + R POST TEMP NEOCORTEX Méthode : subg3 + (subg4) + (subg1)
	Sous-graphes	subg1 : R LAT TEMP NEOCORTEX + R LAT FRONTAL CORTEX, subg2 : R ANT HIPPOC + R POST HIPPOC + R AMY + R TEMPORAL POLE, subg3 : R ANT HIPPOC + R POST HIPPOC + R POST HIPPOC + R ORBITO FRONTAL CX + R TEMPORAL POST NEOCORTEX	

Tableau 7.4 – Comparaison qualitative entre l'ensemble des structures activées, déterminées par analyse visuelle et la méthode BTND pour le patient 3, 4, et 5 (voir légende dans la Tableau 7.3)

	DÉBUT CRISE	PROPAGATION	FIN CRISE
Patient 6	Crise 1	Clinique : L PRECENTRAL OPERCULUM + L POST CENTRAL OPERCULUM + L FRONTAL CORTEX Méthode : subg2 + subg3 + subg4 + (subg1) (subg5)	Clinique : L PRECENTRAL OPERCULUM + L POST CENTRAL OPERCULUM + L FRONTAL CORTEX Méthode : subg4 + (subg1) (subg5)
	Crise 2	Clinique : L PRECENTRAL OPERCULUM + L POST CENTRAL OPERCULUM + L FRONTAL CORTEX Méthode : subg2 + subg3 + subg4 + (subg1) (subg5)	Clinique : L PRECENTRAL OPERCULUM + L POST CENTRAL OPERCULUM + L FRONTAL CORTEX Méthode : subg4 + (subg1) (subg5)
	Crise 3	Clinique : L PRECENTRAL OPERCULUM + L POST CENTRAL OPERCULUM + L FRONTAL CORTEX Méthode : subg2 + subg3 + subg4 + (subg1) (subg5)	Clinique : L PRECENTRAL OPERCULUM + L POST CENTRAL OPERCULUM + L FRONTAL CORTEX Méthode : subg4 + (subg1) (subg5)
Patient 7	Sous-graphes	subg1 : L PRECENTRAL OPERCULUM + L POST CENTRAL OPERCULUM + L PARIETAL CORTEX, subg3 : L POST CENTRAL OPERCULUM + L PARIETAL CORTEX, subg4 : L PRECENTRAL OPERCULUM + L POST CENTRAL OPERCULUM + L POST PARIETAL CORTEX Méthode : subg2 + subg3 + subg4 + (subg1) (subg5)	Clinique : L ANT HIPPOC + L POST HIPPOC + L AMYGD + L ANT TEMPORAL NEOCORTEX Méthode : subg4 + subg5 + subg6 (+subg1)
	Crise 1	Clinique : L ANT HIPPOC + L POST HIPPOC + L AMYGD + L ANT TEMPORAL NEOCORTEX Méthode : subg2 + (subg1) subg1 : L ANT TEMPORAL NEOCORTEX, subg2 : L ANT HIPPOC + L POST HIPPOC + L AMYGD, subg3 : L ANT HIPPOC + L POST HIPPOC + L AMYGD, subg4 : L ANT TEMPORAL NEOCORTEX, subg5 : L ANT TEMPORAL NEOCORTEX, subg6 : L ANT HIPPOC + L POST HIPPOC + L AMYGD.	Clinique : L ANT HIPPOC + L POST HIPPOC + L AMYGD + L ANT TEMPORAL NEOCORTEX Méthode : subg4 + subg5 + subg6 (+subg1)
Patient 8	Crise 1	Clinique : R ANT HIPPOC + R POST HIPPOC + R AMYG + R ENTORINAL CORTEX + R TEMP POLE Méthode : subg3 + subg6 + (subg2)	Clinique : R ANT HIPPOC + R POST HIPPOC + R AMYG + R ENTORINAL CORTEX + R TEMP POLE Méthode : subg1 + subg4 + subg5 + (subg2)
	Crise 2	Clinique : R ANT HIPPOC + R POST HIPPOC + R AMYG + R ENTORINAL CORTEX + R TEMP POLE Méthode : subg3 + subg6 + (subg2)	Clinique : R ANT HIPPOC + R POST HIPPOC + R AMYG + R ENTORINAL CORTEX + R TEMP POLE Méthode : subg1 + subg4 + subg5 + (subg2)
	Crise 3	Clinique : R ANT HIPPOC + R POST HIPPOC + R AMYG + R ENTORINAL CORTEX + R TEMP POLE Méthode : subg3 + subg6 + (subg2)	Clinique : R ANT HIPPOC + R POST HIPPOC + R AMYG + R ENTORINAL CORTEX + R TEMP POLE Méthode : subg3 + subg6 + (subg2)
Patient 9	Sous-graphes	subg1 : R ANT TEMPORAL NEOCORTEX + R POST TEMPORAL NEOCORTEX, subg2 : R POST HIPPOC + R ANT TEMPORAL NEOCORTEX + R ENTORINAL CORTEX + R TEMP POLE + R AMYG, subg3 : R ANT HIPPOC + R POST HIPPOC + R AMYG, subg4 : R ENTORINAL CORTEX + R TEMP POLE + R AMYG, subg5 : R ANT HIPPOC + R POST HIPPOC + R AMYG, subg6 : R ANT HIPPOC + R AMYG	Clinique : R ANT HIPPOC + R POST HIPPOC + R AMYG + R ENTORINAL CORTEX + R TEMP POLE + L TEMP POLE Méthode : subg3 + subg4 + subg5 + subg6
	Crise 1	Clinique : R ANT HIPPOC + R POST HIPPOC + R AMYG + R ENTORINAL CORTEX + R TEMP POLE Méthode : subg2 + (subg1)	Clinique : R ANT HIPPOC + R POST HIPPOC + R AMYG + R ENTORINAL CORTEX + R TEMP POLE + L TEMP POLE Méthode : subg3 + subg4 + subg5 + subg6
	Crise 2	Clinique : L ANT HIPPOC + L TEMP POLE Méthode : subg3 + (subg5) + (subg1)	Clinique : R ANT HIPPOC + R POST HIPPOC + R AMYG + R ENTORINAL CORTEX + R TEMP POLE + L TEMP POLE Méthode : subg4 + subg5 + (subg1)
Patient 9	Crise 2	Clinique : L ANT HIPPOC + L TEMP POLE Méthode : subg3 + (subg5) + (subg1)	Clinique : R ANT HIPPOC + R POST HIPPOC + R AMYG + R ENTORINAL CORTEX + R TEMP POLE + L TEMP POLE Méthode : subg4 + subg5 + (subg1)
	Sous-graphes	subg1 : L POST TEMPORAL NEOCORTEX + L ANT TEMPORAL NEOCORTEX + L ANT TEMPORAL NEOCORTEX, subg2 : L ANT HIPPOC + L ANT TEMPORAL NEOCORTEX, subg3 : R ANT HIPPOC + R POST HIPPOC + R AMYG, subg4 : L ANT HIPPOC + L ANT TEMPORAL NEOCORTEX, subg5 : + L POST TEMPORAL NEOCORTEX + L ANT TEMPORAL NEOCORTEX, subg6 : R ANT TEMPORAL NEOCORTEX + R POST TEMPORAL NEOCORTEX	Clinique : R ANT HIPPOC + R POST HIPPOC + R AMYG + R ENTORINAL CORTEX + R TEMP POLE + L TEMP POLE Méthode : subg4 + subg5 + (subg6)

Tableau 7.5 – Comparaison qualitative entre l'ensemble des structures activées, déterminées par analyse visuelle et la méthode BTND pour le patient 6, 7, 8 et 9. (voir légende dans le Tableau 7.3)

Conclusion

Résumé des contributions

Dans cette thèse, on a proposé des méthodes d'inférence et de décomposition de graphes dynamiques dans le contexte des neurosciences. Pour répondre à cet objectif :

Dans le chapitre 2, on a présenté une nouvelle expression des moments d'ordres supérieurs entre la phase et l'amplitude de deux signaux gaussiens, en fonction des éléments de leur matrice de covariance analytique. Cette expression a permis de montrer que, dans le cas gaussien, les mesures de CFs comme la PLV ou l'AEC entre deux signaux, sont des fonctions monotones de leur corrélation et sont donc des mesures similaires. De plus, l'expression des moments d'ordres supérieurs entre la phase et l'amplitude est généralisée au cas multivarié. Cette expression a permis de proposer des mesures de PLV ou d'AEC conditionnées, qui sont dans le cas gaussien des fonctions monotones de la corrélation partielle. Cependant, on a montré dans ce chapitre que ces formules de PLV ou d'AEC conditionnées étaient fonction de la matrice de précision analytique. C'est-à-dire qu'elles restent des mesures différentes de la corrélation partielle dans un contexte général. On fait présumer alors qu'elles peuvent être une alternative intéressante aux mesures de PLV ou d'AEC conditionnées déjà existantes.

Dans le chapitre 3, on a proposé une nouvelle mesure de CF pour directement inférer des graphes dynamiques robustes et facilement interprétables. Pour être plus précis, on a dans un premier temps proposé deux extensions de la PLV, qui sont des mesures de PLV conditionnelle et régularisée. On conditionne les mesures de PLV afin de prendre en compte l'existence des autres signaux analysés : la première se base sur la mesure usuelle de PLV non paramétrique, la seconde sur la mesure de PLV conditionnelle que l'on a proposée dans le chapitre précédent, supposant que le signal suit une distribution gaussienne multivariée. Comme seconde contribution, on a explicité des contraintes utiles, pour obtenir des graphes de CFs pertinents dans le cadre d'enregistrement de crise d'épilepsie. Les régularisations sont choisies pour obtenir des graphes de CFs parcimonieux, qui permettent de mettre en avant uniquement les CFs les plus pertinentes pour caractériser la crise. Elles sont aussi choisies ainsi, car on s'attend à ce que les estimations des réseaux entre deux instants consécutifs varient peu. C'est pourquoi, on a spécifiquement proposé d'imposer une contrainte de lissage temporel dans les deux procédures d'estimation de la PLV conditionnée. Cela se traduit par un problème de lasso graphique variant dans le temps. La résolution de ce problème utilisant la méthode de direction alternée des multiplicateurs (ADMM) est détaillée. Finalement, on a comparé les différentes mesures de PLV conditionnée et régularisée avec la corrélation partielle classique sur un modèle de signaux synchrones réalistes, générés par un oscillateur de Roessler. Les résultats de cette étude ont montré qu'une seule mesure de PLV conditionnée et régularisée est satisfaisante, dans le sens où elle n'est jamais impactée par l'amplitude des signaux modélisés : il s'agit de l'extension de la PLV non paramétrique. Sur un jeu de données réelles, la PLV conditionnée et régularisée, la PLV conditionnée non régularisée et la PLV classique ont été comparées. On a ainsi pu montrer que la PLV conditionnée et régularisée infère des CFs qui ont tendance à être plus robustes sur l'ensemble des crises d'un même patient.

Dans le chapitre 5 et 6, on a proposé des méthodes afin de décomposer un graphe dynamique de CFs en un ensemble de sous-graphes contenant une information neuro-physiologique pertinente.

Afin d'obtenir des résultats plus robustes, on considère les différentes crises d'un même patient. Si les crises sont de durées similaires, on propose une décomposition tensorielle nommée HOSMR basée sur un algorithme déjà existant, la HOOI, pouvant être utilisée pour sélectionner des facteurs de CFs représentatifs de l'ensemble des crises. La réduction dimensionnelle qui a été proposée est une décomposition tensorielle originale et spécifique au problème, imposant des contraintes structurelles à la solution : il s'agit de rendre simultanément les facteurs de CFs et les profils d'activation parcimonieux. On a opté pour imposer une contrainte Elastic-net sur les profils d'activation. Cette contrainte s'est montrée efficace pour obtenir des graphes de CFs qui ont des profils d'activation localisés et donc spécifiques à certaines étapes de la crise.

Pour le cas où les crises ont des durées variables, on propose plusieurs alternatives à la HOSMR. En particulier, une décomposition jointe en matrices non négatives jointes, nommée joinNMF, qui permet d'inférer des facteurs de CFs communs à toutes les crises mais avec des profils d'activation spécifiques à chaque crise. Pour cette décomposition, on a ajouté une régularisation de parcimonie sur les facteurs de CFs, mais aussi une contrainte fused-lasso sur les profils d'activation spécifiques à chaque crise obtenue. Cette contrainte permet à la fois d'obtenir des profils parcimonieux et compacts dans le temps qui s'activent durant une étape de la crise.

Une variation connue de la décomposition jointe en matrices non négatives a aussi été étudiée. Celle-ci permet d'obtenir des facteurs de CFs plus souples qui contiennent certaines spécificités de chaque crise. On a montré que cette décomposition peut être modifiée pour exhiber une régularisation de Thikonov sur le mode des crises. Cette nouvelle décomposition se nomme graphNMF.

On a comparé ces différentes méthodes entre elles sur des graphes dynamiques de CFs simulées. On a ainsi pu montrer que les décompositions jointes en matrices non négatives sous contraintes proposées permettaient d'obtenir des résultats globalement aussi bons que la HOSMR, mais sur des crises de durées différentes. Enfin, on a appliqué les différentes approches proposées sur les crises de deux patients différents. Le premier ayant des crises de durées similaires, tandis que le second subit des crises de durées hétérogènes. L'intérêt des décompositions jointes en matrices non négatives sous contraintes est rendu évident sur le patient qui endure des crises de durée hétérogène ; ce sont les seules méthodes qui fournissent des résultats cohérents avec l'observation clinique.

Dans le chapitre 7, on a proposé une procédure globale, nommé BNTD, afin de caractériser les réseaux pathologiques de CFs de patients épileptiques à partir d'enregistrement sEEG. Elle se base sur la méthode joinNMF présentée dans le chapitre 6. La stratégie proposée a été appliquée à 27 crises provenant de 9 patients différents qui souffrent d'épilepsie focale résistante aux médicaments. Une étude clinique a été réalisée et les décompositions obtenues ont été comparées dans chaque cas à l'interprétation du clinicien. Pour chaque patient, les contraintes imposées à la méthode multNMF ont globalement permis d'obtenir des sous-graphes de CFs spatialement localisés dans le cerveau avec des profils d'activation indiquant des périodes temporelles de la crise. Ces résultats ont donc pu être facilement interprétés cliniquement. Pour 8 des 9 patients, les décompositions obtenues correspondaient à l'interprétation du clinicien. Pour le patient avec une mauvaise correspondance entre la décomposition et l'interprétation clinique, on a remarqué que la mesure de connectivité fonctionnelle utilisée (la PLV) ne permettait pas de souligner les CFs importantes caractérisant le début de la crise. On pense donc que la méthode proposée peut être utilisée pour enrichir l'interprétation visuelle du clinicien.

Perspectives

Les différentes stratégies proposées dans cette thèse ouvrent de nombreuses perspectives. Le chapitre 2 compare deux mesures de CFs dans le contexte de signaux qui suivent une loi gaussienne multivariée. Cela pourrait être étendu à d'autres hypothèses sur la distribution des signaux. Cependant, il est vraisemblable que pour d'autres distributions, il n'existe pas d'expression analytique de ces mesures de CFs en fonction des paramètres de cette distribution. Il serait alors envisageable d'estimer la fonction liant les paramètres de la distribution à la mesure de CFs. Ainsi on pourrait obtenir des mesures

alternatives de CFs via les valeurs estimées des paramètres d'une loi de probabilité quelconque faite sur les signaux.

À partir du chapitre 3, qui reste un travail exploratoire, de nombreuses pistes de recherche ont pu être soulevées. Dans un premier temps, on a remarqué que si les signaux utilisés étaient spatialement corrélés, alors, les mesures conditionnelles de PLV ou de corrélation semblaient donner des résultats moins pertinents. En effet, lorsque toutes les électrodes de l'enregistrement sEEG sont utilisées, les mesures de PLV conditionnées révèlent uniquement des adjacences spatiales entre les électrodes. De nombreuses CFs qui semblent pertinentes pour caractériser la crise d'épilepsie disparaissent. Devant ce constat, deux approches sont envisageables. La première est une modification de la définition de PLV conditionnée, permettant de ne pas conditionner le signal avec les signaux des électrodes spatialement trop proches. La seconde approche est de calculer des mesures de PLV régularisées simples, sans conditionnement préalable des signaux ; il s'agit alors d'un problème beaucoup plus complexe à implémenter car non convexe.

Le calcul de la PLV conditionnée et régularisée se base sur le modèle du lasso graphique, dont le critère est la divergence de Kullback Leibler régularisée. Cependant, ce critère est avant tout adapté pour calculer la matrice de précision d'un processus gaussien. D'autres modèles pourraient être proposés comme le critère CLIME [43], [11]. Cela pourrait faire l'objet d'une comparaison de différentes options pour régulariser la PLV conditionnée.

Dans ce chapitre, on fait l'hypothèse que les mesures des CFs doivent être lisses dans le temps. Cependant, comme une crise d'épilepsie est composée d'étapes et qu'il existe des modifications brutales de l'état cortical du cerveau entre deux étapes successives, il pourrait être plus pertinent d'utiliser la régularisation des variations totales [56]. En effet, cette régularisation permet d'obtenir des mesures de CFs constantes par morceaux, mais avec potentiellement de fortes différences entre deux étapes successives d'une crise.

Enfin, l'algorithme ADMM, bien que fonctionnel, n'est peut-être pas le meilleur choix d'implémentation en terme de vitesse de convergence. Une comparaison des performances obtenues pour inférer des PLV conditionnées et régularisées avec un algorithme de type Forward-Backward, pourrait être envisagée [186].

Dans le chapitre 6, on a présenté une méthode de décomposition jointe en matrice non-négative avec régularisation de Thikonov appliquée sur le mode des crises, nommée graphNMF. Il s'agit d'une nouvelle méthode qui correspond à une généralisation d'une décomposition déjà proposée dans [136]. La méthode graphNMF pourrait faire l'objet d'analyses complémentaires. En effet, la méthode graphNMF peut être intéressante pour des cas cliniques complexes. Pour certains patients, les activations de CFs peuvent être similaires pour beaucoup de crises, mais avoir un comportement atypique pour certaines crises spécifiques. Alors il serait intéressant de prendre en compte la similarité des différentes crises dans la matrice laplacienne de la méthode graphNMF. Cela permettrait d'obtenir des décompositions plus robustes pour ces patients. Par conséquent, il est nécessaire de trouver une mesure de similarité entre les différentes crises d'un patient. On peut par exemple comparer les différentes matrices de CFs obtenues pour chaque crise de durées variables. Cela implique d'utiliser des mesures de distance entre des espaces de différentes dimensions. On pourra alors penser à des mesures comme la distance de Grassman [206]. Celle-ci, doit cependant être modifiée, car en pratique elle ne permet pas de bien discriminer des crises avec des comportements différents.

Aussi, on a montré dans l'annexe du chapitre 6, comment les méthodes multNMF et graphNMF peuvent être associées à un problème d'analyse canonique des corrélations. Ces méthodes peuvent donc être utilisées pour plonger les matrices de CFs de durées différentes dans un espace où elles ont toutes la même durée. Ainsi, la liste de matrice de CFs de durée différente peut être transformée en un tenseur de CFs et une méthode tensorielle. Cette dernière représentation permet d'appliquer des décompositions tensorielles comme la HOSMR.

Enfin, comme il n'est pas possible de parfaitement contrôler les contraintes appliquées sur Parafac2,

on pourrait s'aider d'un relâchement de cette méthode proposée dans [51] afin d'ajouter les contraintes désirées sur les profils d'activation.

Un des relâchements de Parafac2 proposée dans [51] afin d'obtenir des composantes non négatives pourrait être une bonne approche pour ajouter les contraintes désirées sur les profils d'activation.

À partir du chapitre 7, on a présenté une procédure, la Brain-Wide Time-varying Network Decomposition BNTD, qui permettait de capturer des réseaux de CF pathologiques interprétables cliniquement. Cependant, cette méthode souffre de la dépendance à 3 hyperparamètres impactant la qualité des décompositions obtenues. Une procédure simple a été employée pour sélectionner automatiquement les paramètres pour les 9 patients. À chaque fois, la décomposition obtenue était pertinente. Cependant, il est possible que pour un contexte clinique différent, ou une mesure de CFs différente de la PLV, la procédure proposée ne donne pas lieu à l'identification des paramètres permettant d'obtenir une décomposition pertinente. Il est alors nécessaire de mieux étudier la relation entre ces 3 paramètres.

On a montré que pour un patient, la décomposition obtenue via la méthode BTND ne correspondait pas avec l'interprétation du clinicien. Dans ce cas, la mesure de CFs utilisée, qui est la PLV, n'arrive pas à souligner les connections pertinentes du début de crise. C'est d'autant plus intéressant dans ce cas, car la visualisation des signaux montre clairement une dépendance entre eux. Cependant, peu de mesures de CFs arrivent à exhiber leurs couplages. Cela montre l'importance de la mesure de CFs utilisée pour créer le réseau dynamique de CFs. Une étude des performances de différentes mesures de CFs, dans le contexte de l'épilepsie et dans l'objectif d'appliquer la méthode BTND, pourrait être proposée.

Il est important de noter que la méthode BNTD n'extrait que des graphes de CFs associés à de fortes synchronies entre des régions du cerveau. Cependant, des travaux ont montré que le début de la crise était aussi caractérisé par d'importantes désynchronisations entre certaines régions clés de la crise d'épilepsie [199]. Alors, il serait pertinent que la méthode BNTD puisse aussi montrer des sous-graphes de désynchronie (c'est à dire associer à des mesures de CFs qui deviennent faibles durant une période de la crise).

Finalement, une validation directe de la méthode est hors de portée car il n'existe pas d'étalon absolu pour estimer le schéma de connectivité chez les patients épileptiques. On a choisi de faire une correspondance entre la propagation révélée par la méthode BNTD et l'interprétation visuelle du clinicien. Il s'agit donc d'une étude qualitative. Cependant, il existe des modèles neuronaux simulant la propagation d'une crise. Ces modèles pourraient être utilisés pour pouvoir mieux quantifier les performances de la méthode proposée.

Bibliographie

- [1] Prox Repository, <http://proximity-operator.net/>.
- [2] Wolfram|Alpha : Making the world's knowledge computable, <https://www.wolframalpha.com>.
- [3] Emmanuel Abbe. Community detection and stochastic block models : recent developments. *The Journal of Machine Learning Research*, 18(1) :6446–6531, 2017.
- [4] Ardavan Afshar, Ioakeim Perros, Evangelos E. Papalexakis, Elizabeth Searles, Joyce Ho, and Jimeng Sun. COPA : Constrained PARAFAC2 for sparse & large datasets. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 793–802, 2018.
- [5] Frédéric Alexandre. Comprendre les réseaux cérébraux. 2013.
- [6] Genevera Allen. Sparse higher-order principal components analysis. In *Artificial Intelligence and Statistics*, pages 27–36, 2012.
- [7] Joan F. Alonso, Miguel A. Mañanas, Sergio Romero, Dirk Hoyer, Jordi Riba, and Manel J. Barbanoj. Drug effect on EEG connectivity assessed by linear and nonlinear couplings. *Human brain mapping*, 31(3) :487–497, 2010.
- [8] G. D. Anderson, M. K. Vamanamurthy, and M. Vuorinen. Hypergeometric functions and elliptic integrals. *Current topics in analytic function theory*, pages 48–85, 1992.
- [9] Claus A. Andersson and Rasmus Bro. The N-way toolbox for MATLAB. *Chemometrics and intelligent laboratory systems*, 52(1) :1–4, 2000.
- [10] David Arthur and Sergei Vassilvitskii. k-means++ : The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.
- [11] Karina Ashurbekova, Sophie Achard, and Florence Forbes. Robust structure learning using multivariate T-distributions. 2018.
- [12] Sergul Aydore, Dimitrios Pantazis, and Richard M. Leahy. Phase synchrony in multivariate Gaussian data with applications to cortical networks. In *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1547–1550. IEEE, 2012.
- [13] Sergul Aydore, Dimitrios Pantazis, and Richard M. Leahy. A note on the phase locking value and its properties. *Neuroimage*, 74 :231–244, 2013.
- [14] Lacoma Ayoubian, H. Lacoma, and J. Gotman. Automatic seizure detection in SEEG using high frequency activities in wavelet domain. *Medical engineering & physics*, 35(3) :319–328, 2013.
- [15] Francis Bach, Julien Mairal, and Jean Ponce. Convex sparse matrix factorizations. *arXiv preprint arXiv :0812.1869*, 2008.
- [16] Roland Badeau and Rémy Boyer. Fast multilinear singular value decomposition for structured tensors. *SIAM Journal on Matrix Analysis and Applications*, 30(3) :1008–1021, 2008.
- [17] Brett W. Bader and Tamara G. Kolda. Matlab tensor toolbox version 2.5. *Available online, January, 7, 2012*.
- [18] Giulia Barbati, Camillo Porcaro, Filippo Zappasodi, Paolo Maria Rossini, and Franca Tecchio. Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals. *Clinical Neurophysiology*, 115(5) :1220–1232, 2004.

- [19] Ole Barndorff-Nielsen. Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of statistics*, pages 151–157, 1978.
- [20] Fabrice Bartolomei, Patrick Chauvel, and Fabrice Wendling. Epileptogenicity of brain structures in human temporal lobe epilepsy : a quantified study from intracerebral EEG. *Brain : A Journal of Neurology*, 131(Pt 7) :1818–1830, July 2008. Number : Pt 7.
- [21] Fabrice Bartolomei, Stanislas Lagarde, Fabrice Wendling, Aileen McGonigal, Viktor Jirsa, Maxime Guye, and Christian Bénar. Defining epileptogenic networks : Contribution of SEEG and signal analysis. *Epilepsia*, 58(7) :1131–1147, 2017. Number : 7.
- [22] Danielle Smith Bassett and E. D. Bullmore. Small-world brain networks. *The neuroscientist*, 12(6) :512–523, 2006.
- [23] Federico Battiston, Vincenzo Nicosia, and Vito Latora. Structural measures for multiplex networks. *Physical Review E*, 89(3) :032804, 2014.
- [24] Christian Bauckhage. K-means clustering is matrix factorization. *arXiv preprint arXiv :1512.07548*, 2015.
- [25] Heinz H. Bauschke and Patrick L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*, volume 408. Springer, 2011.
- [26] Dimitri Bayle. *Traitement cérébral de l'expression faciale de peur : vision périphérique et effet de l'attention*. PhD thesis, 2009.
- [27] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1) :183–202, 2009.
- [28] Hanna Becker. *Denoising, separation and localization of EEG sources in the context of epilepsy*. PhD thesis, 2014.
- [29] Hanna Becker, Laurent Albera, Pierre Comon, Martin Haardt, Gwénaél Birot, Fabrice Wendling, Martine Gavaret, Christian G. Bénar, and Isabelle Merlet. EEG extended source localization : tensor-based vs. conventional methods. *NeuroImage*, 96 :143–157, 2014.
- [30] Hanna Becker, Laurent Albera, Pierre Comon, Amar Kachenoura, and Isabelle Merlet. A penalized semialgebraic deflation ICA algorithm for the efficient extraction of interictal epileptic signals. *IEEE journal of biomedical and health informatics*, 21(1) :94–104, 2015.
- [31] Hanna Becker, Pierre Comon, and Laurent Albera. Tensor-based preprocessing of combined EEG/MEG data. In *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 275–279. IEEE, 2012.
- [32] Edward Bedrosian. A product theorem for Hilbert transforms. 1962. Publisher : Rand Corporation.
- [33] Asa Ben-Hur and Isabelle Guyon. Detecting stable clusters using principal component analysis. In *Functional genomics*, pages 159–182. Springer, 2003.
- [34] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. Pearson correlation coefficient. In *Noise reduction in speech processing*, pages 1–4. Springer, 2009.
- [35] Peter J. Bickel and Aiyu Chen. A nonparametric view of network models and Newman–Girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50) :21068–21073, 2009.
- [36] Stephen Boyd, Stephen P. Boyd, and Lieven Vandenbergh. *Convex optimization*. Cambridge university press, 2004.
- [37] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1) :1–122, 2011.
- [38] Urs Braun, Axel Schäfer, Henrik Walter, Susanne Erk, Nina Romanczuk-Seiferth, Leila Haddad, Janina I. Schweiger, Oliver Grimm, Andreas Heinz, and Heike Tost. Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proceedings of the National Academy of Sciences*, 112(37) :11678–11683, 2015.

- [39] Lev M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3) :200–217, 1967. ISBN : 0041-5553 Publisher : Elsevier.
- [40] Steven L. Bressler and Anil K. Seth. Wiener–Granger causality : a well established methodology. *Neuroimage*, 58(2) :323–329, 2011.
- [41] Ed Bullmore and Olaf Sporns. Complex brain networks : graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, 10(3) :186–198, 2009.
- [42] Samuel P. Burns, Sabato Santaniello, Robert B. Yaffe, Christophe C. Jouny, Nathan E. Crone, Gregory K. Bergey, William S. Anderson, and Sridevi V. Sarma. Network dynamics of the brain and influence of the epileptic seizure onset zone. *Proceedings of the National Academy of Sciences*, 111(49) :E5321–E5330, 2014.
- [43] Tony Cai, Weidong Liu, and Xi Luo. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494) :594–607, 2011. ISBN : 0162-1459 Publisher : Taylor & Francis.
- [44] Lucy R. Chai, Ankit N. Khambhati, Rastko Ciric, Tyler M. Moore, Ruben C. Gur, Raquel E. Gur, Theodore D. Satterthwaite, and Danielle S. Bassett. Evolution of brain network dynamics in neurodevelopment. *Network Neuroscience*, 1(1) :14–30, 2017.
- [45] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1-2) :89–97, 2004.
- [46] Patrick Chauvel, Sylvain Rheims, Aileen McGonigal, and Philippe Kahane. *French guidelines on stereoelectroencephalography (SEEG) : Editorial comment*. Elsevier, 2017.
- [47] Jia Chen, Gang Wang, and Georgios B. Giannakis. Graph multiview canonical correlation analysis. *IEEE Transactions on Signal Processing*, 67(11) :2826–2838, 2019. ISBN : 1053-587X Publisher : IEEE.
- [48] Andrzej Cichocki. Tensor decompositions : a new concept in brain data analysis? *arXiv preprint arXiv :1305.0395*, 2013.
- [49] Andrzej Cichocki and Anh-Huy Phan. Fast local algorithms for large scale nonnegative matrix and tensor factorizations. *IEICE transactions on fundamentals of electronics, communications and computer sciences*, 92(3) :708–721, 2009.
- [50] Andrzej Cichocki, Rafal Zdunek, Seungjin Choi, Robert Plemmons, and Shun-Ichi Amari. Non-negative tensor factorization using alpha and beta divergences. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, volume 3, pages III–1393–III–1396. IEEE, 2007.
- [51] Jeremy E. Cohen and Rasmus Bro. Nonnegative PARAFAC2 : a flexible coupling approach. In *International Conference on Latent Variable Analysis and Signal Separation*, pages 89–98. Springer, 2018.
- [52] Giles L. Colclough, Mark W. Woolrich, P. K. Tewarie, Matthew J. Brookes, Andrew J. Quinn, and Stephen M. Smith. How reliable are MEG resting-state connectivity metrics? *Neuroimage*, 138 :284–293, 2016.
- [53] D. Louis Collins, Peter Neelin, Terrence M. Peters, and Alan C. Evans. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of computer assisted tomography*, 18(2) :192–205, 1994. ISBN : 0363-8715.
- [54] Pierre Comon. Tensors : a brief introduction. *IEEE Signal Processing Magazine*, 31(3) :44–53, 2014.
- [55] Pierre Comon, Xavier Luciani, and André LF De Almeida. Tensor decompositions, alternating least squares and other tales. *Journal of Chemometrics : A Journal of the Chemometrics Society*, 23(7-8) :393–405, 2009.
- [56] Laurent Condat. A direct algorithm for 1-D total variation denoising. *IEEE Signal Processing Letters*, 20(11) :1054–1057, 2013.

- [57] Fengyu Cong, Qiu-Hua Lin, Li-Dan Kuang, Xiao-Feng Gong, Piia Astikainen, and Tapani Ristaniemi. Tensor decomposition of EEG signals : a brief review. *Journal of neuroscience methods*, 248 :59–69, 2015.
- [58] Aude Costard. *Estimation de la structure d'indépendance conditionnelle d'un réseau de capteurs : application à l'imagerie médicale*. PhD thesis, 2014.
- [59] Joachim Dahl, Lieven Vandenberghe, and Vwani Roychowdhury. Covariance selection for non-chordal graphs via chordal embedding. *Optimization Methods & Software*, 23(4) :501–520, 2008. ISBN : 1055-6788 Publisher : Taylor & Francis.
- [60] Patrick Danaher, Pei Wang, and Daniela M. Witten. The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 76(2) :373–397, 2014.
- [61] Lieven De Lathauwer. Decompositions of a higher-order tensor in block terms—Part I : Lemmas for partitioned matrices. *SIAM Journal on Matrix Analysis and Applications*, 30(3) :1022–1032, 2008.
- [62] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, 21(4) :1253–1278, 2000.
- [63] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. On the best rank-1 and rank-(r_1, r_2, \dots, r_n) approximation of higher-order tensors. *SIAM journal on Matrix Analysis and Applications*, 21(4) :1324–1342, 2000.
- [64] Chris Ding and Xiaofeng He. K-means clustering via principal component analysis. In *Proceedings of the twenty-first international conference on Machine learning*, page 29. ACM, 2004.
- [65] David Donoho and Victoria Stodden. When does non-negative matrix factorization give a correct decomposition into parts ? In *Advances in neural information processing systems*, pages 1141–1148, 2004.
- [66] Jonathan Eckstein and Dimitri P. Bertsekas. On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1-3) :293–318, 1992.
- [67] Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. *The Annals of statistics*, 32(2) :407–499, 2004.
- [68] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.
- [69] Dan Feldman, Melanie Schmidt, and Christian Sohler. Turning big data into tiny data : Constant-size coresets for k-means, pca and projective clustering. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pages 1434–1453. Society for Industrial and Applied Mathematics, 2013.
- [70] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3-5) :75–174, 2010.
- [71] Santo Fortunato and Darko Hric. Community detection in networks : A user guide. *Physics reports*, 659 :1–44, 2016.
- [72] Brendan J. Frey and Delbert Dueck. Clustering by passing messages between data points. *science*, 315(5814) :972–976, 2007.
- [73] Jerome Friedman, Trevor Hastie, Holger Höfling, and Robert Tibshirani. Pathwise coordinate optimization. *The annals of applied statistics*, 1(2) :302–332, 2007.
- [74] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3) :432–441, 2008.
- [75] Karl J. Friston. Functional and effective connectivity : a review. *Brain connectivity*, 1(1) :13–36, 2011.
- [76] Gaëtan Frusque, Julien Jung, Pierre Borgnat, and Paulo Gonçalves. Sparse tensor dimensionality reduction with application to clustering of functional connectivity. In *Wavelets and Sparsity XVIII*, volume 11138, page 111380N. International Society for Optics and Photonics, 2019.

- [77] Gaëtan Michel Frusque, Julien Jung, Pierre Borgnat, and Paulo Gonçalves. Multiplex network inference with sparse tensor decomposition for functional connectivity. *IEEE transactions on Signal and Information Processing over Networks*, 2020. ISBN : 2373-776X Publisher : IEEE.
- [78] Matteo Garofalo, Thierry Nieus, Paolo Massobrio, and Sergio Martinoia. Evaluation of the performance of information theory-based methods and cross-correlation to estimate the functional connectivity in cortical networks. *PloS one*, 4(8), 2009.
- [79] Laetitia Gauvin, André Panisson, and Ciro Cattuto. Detecting the community structure and activity patterns of temporal networks : a non-negative tensor factorization approach. *PloS one*, 9(1), 2014.
- [80] Linda Geerligs, Natasha M. Maurits, Remco J. Renken, and Monicque M. Lorst. Reduced specificity of functional connectivity in the aging brain during task performance. *Human brain mapping*, 35(1) :319–330, 2014.
- [81] Christoph H. Gierull and Ishuwa C. Sikaneta. Estimating the effective number of looks in interferometric SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 40(8) :1733–1742, 2002.
- [82] Pontus Giselsson and Stephen Boyd. Diagonal scaling in Douglas-Rachford splitting and ADMM. In *53rd IEEE Conference on Decision and Control*, pages 5033–5039. IEEE, 2014.
- [83] Pontus Giselsson and Stephen Boyd. Linear convergence and metric selection for Douglas-Rachford splitting and ADMM. *IEEE Transactions on Automatic Control*, 62(2) :532–544, 2016. ISBN : 0018-9286 Publisher : IEEE.
- [84] Vadym Gnatkovsky, Stefano Francione, Francesco Cardinale, Roberto Mai, Laura Tassi, Giorgio Lo Russo, and Marco De Curtis. Identification of reproducible ictal patterns based on quantified frequency analysis of intracranial EEG signals. *Epilepsia*, 52(3) :477–488, 2011.
- [85] Javier Gonzalez-Castillo, Colin W. Hoy, Daniel A. Handwerker, Meghan E. Robinson, Laura C. Buchanan, Ziad S. Saad, and Peter A. Bandettini. Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. *Proceedings of the National Academy of Sciences*, 112(28) :8762–8767, 2015.
- [86] Izrail Solomonovich Gradshteyn and Iosif Moiseevich Ryzhik. *Table of integrals, series, and products*. Academic press, 2014.
- [87] Michael D. Greicius, Kaustubh Supekar, Vinod Menon, and Robert F. Dougherty. Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral cortex*, 19(1) :72–78, 2009.
- [88] Olesya Grinenko, Jian Li, John C. Mosher, Irene Z. Wang, Juan C. Bulacio, Jorge Gonzalez-Martinez, Dileep Nair, Imad Najm, Richard M. Leahy, and Patrick Chauvel. A fingerprint of the epileptogenic zone in human epilepsies. *Brain*, 141(1) :117–131, 2018.
- [89] Marc Guenet, Jean Isnard, Philippe Ryvlin, Catherine Fischer, Karine Ostrowsky, François Mauguere, and Marc Sindou. Neurophysiological monitoring for epilepsy surgery : the Talairach SEEG method. *Stereotactic and functional neurosurgery*, 77(1-4) :29–32, 2001.
- [90] David Hallac, Youngsuk Park, Stephen Boyd, and Jure Leskovec. Network inference via the time-varying graphical lasso. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 205–213. ACM, 2017.
- [91] Mark Hallett, Willem de Haan, Gustavo Deco, Reinhard Dengler, Riccardo Di Iorio, Cecile Gallea, Christian Gerloff, Christian Grefkes, Rick C. Helmich, Morten L. Kringelbach, Francesca Miraglia, Ivan Rektor, Ondřej Strýček, Fabrizio Vecchio, Lukas J. Volz, Tao Wu, and Paolo M. Rossini. Human brain connectivity : Clinical applications for clinical neurophysiology. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, 131(7) :1621–1651, July 2020. Number : 7.
- [92] Ronan Hamon, Pierre Borgnat, Patrick Flandrin, and Celine Robardet. Extraction of temporal network structures from graph-based signals. *IEEE Transactions on Signal and Information Processing over Networks*, 2(2) :215–226, 2016.

- [93] Ronan Hamon, Pierre Borgnat, Patrick Flandrin, and Céline Robardet. Nonnegative matrix factorization to find features in temporal networks. In *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 1065–1069. IEEE, 2014.
- [94] Trevor Hastie, Robert Tibshirani, and Martin Wainwright. *Statistical learning with sparsity : the lasso and generalizations*. Chapman and Hall/CRC, 2015.
- [95] Anant Hegde, Deniz Erdogmus, and Jose C. Principe. Spatio-temporal clustering of epileptic ECOG. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 4199–4202. IEEE, 2006.
- [96] Ahmad Hany Hossny, Terry Moschuo, Grant Osborne, Lewis Mitchell, and Nick Lothian. Enhancing keyword correlation for event detection in social networks using SVD and k-means : Twitter case study. *Social Network Analysis and Mining*, 8(1) :49, 2018.
- [97] Patrik O. Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of machine learning research*, 5(Nov) :1457–1469, 2004.
- [98] Weiyu Huang, Thomas AW Bolton, John D. Medaglia, Danielle S. Bassett, Alejandro Ribeiro, and Dimitri Van De Ville. Graph signal processing of human brain imaging data. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 980–984. IEEE, 2018.
- [99] Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of classification*, 2(1) :193–218, 1985.
- [100] Borbála Hunyadi, Patrick Dupont, Wim Van Paesschen, and Sabine Van Huffel. Tensor decompositions and data fusion in epileptic electroencephalography and functional magnetic resonance imaging data. *Wiley Interdisciplinary Reviews : Data Mining and Knowledge Discovery*, 7(1) :e1197, 2017.
- [101] Asif Iqbal, Mohamed Nait-Meziane, Abd-Krim Seghouane, and Karim Abed-Meraim. Adaptive complex-valued dictionary learning : Application to fMRI data analysis. *Signal Processing*, 166 :107263, 2020. ISBN : 0165-1684 Publisher : Elsevier.
- [102] Jean Isnard, Delphine Taussig, Fabrice Bartolomei, Pierre Bourdillon, Hélène Catenoix, Francine Chassoux, Mathilde Chipaux, Stéphane Clemenceau, Sophie Colnat-Coulbois, and Marie Denuelle. French guidelines on stereoelectroencephalography (SEEG). *Neurophysiologie Clinique*, 48(1) :5–13, 2018.
- [103] Anil K. Jain. Data clustering : 50 years beyond K-means. *Pattern recognition letters*, 31(8) :651–666, 2010.
- [104] S. Rao Jammalamadaka and Ambar Sengupta. *Topics in circular statistics*, volume 5. world scientific, 2001.
- [105] Jana Janková and Sara van de Geer. Inference in high-dimensional graphical models. *arXiv preprint arXiv :1801.08512*, 2018.
- [106] Xingpeng Jiang, Xiaohua Hu, and Weiwei Xu. Microbiome data representation by joint nonnegative matrix factorization with laplacian regularization. *IEEE/ACM transactions on computational biology and bioinformatics*, 14(2) :353–359, 2015.
- [107] Premysl Jiruska, Marco de Curtis, John G. R. Jefferys, Catherine A. Schevon, Steven J. Schiff, and Kaspar Schindler. Synchronization and desynchronization in epilepsy : controversies and hypotheses. *The Journal of Physiology*, 591(4) :787–797, February 2013. Number : 4.
- [108] Alexander Jung, Gabor Hannak, and Norbert Goertz. Graphical lasso based model selection for time series. *IEEE Signal Processing Letters*, 22(10) :1781–1785, 2015.
- [109] Mahdi M. Kalayeh, Haroon Idrees, and Mubarak Shah. NMF-KNN : Image annotation using weighted multi-view non-negative matrix factorization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 184–191, 2014.
- [110] Keita Kamijo, Yuji Takeda, and Charles H. Hillman. The relation of physical activity to functional connectivity between brain regions. *Clinical Neurophysiology*, 122(1) :81–89, 2011.

- [111] Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical review E*, 83(1) :016107, 2011.
- [112] Matthew SD Kerr, Samuel P. Burns, John Gale, Jorge Gonzalez-Martinez, Juan Bulacio, and Sridevi V. Sarma. Multivariate analysis of SEEG signals during seizure. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 8279–8282. IEEE, 2011.
- [113] Jon R. Kettenring. Canonical analysis of several sets of variables. *Biometrika*, 58(3) :433–451, 1971. ISBN : 1464-3510 Publisher : Oxford University Press.
- [114] Ankit N. Khambhati, Danielle S. Bassett, Brian S. Oommen, Stephanie H. Chen, Timothy H. Lucas, Kathryn A. Davis, and Brian Litt. Recurring functional interactions predict network architecture of interictal and ictal states in neocortical epilepsy. *eneuro*, 4(1), 2017.
- [115] Ankit N. Khambhati, Kathryn A. Davis, Brian S. Oommen, Stephanie H. Chen, Timothy H. Lucas, Brian Litt, and Danielle S. Bassett. Dynamic network drivers of seizure generation, propagation and termination in human neocortical epilepsy. *PLoS computational biology*, 11(12) :e1004608, 2015.
- [116] Ankit N. Khambhati, Ann E. Sizemore, Richard F. Betzel, and Danielle S. Bassett. Modeling and interpreting mesoscale network dynamics. *NeuroImage*, 180 :337–349, 2018.
- [117] Henk AL Kiers, Jos MF Ten Berge, and Rasmus Bro. PARAFAC2—Part I. A direct fitting algorithm for the PARAFAC2 model. *Journal of Chemometrics : A Journal of the Chemometrics Society*, 13(3-4) :275–294, 1999.
- [118] Hannah Kim, Jaegul Choo, Jingu Kim, Chandan K. Reddy, and Haesun Park. Simultaneous discovery of common and discriminative topics via joint nonnegative matrix factorization. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 567–576, 2015.
- [119] Yong-Deok Kim and Seungjin Choi. Nonnegative tucker decomposition. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.
- [120] Frederick W. King. *Hilbert transforms*, volume 2. Cambridge University Press Cambridge, 2009.
- [121] Trupti M. Kodinariya and Prashant R. Makwana. Review on determining number of Cluster in K-Means Clustering. *International Journal*, 1(6) :90–95, 2013.
- [122] T. Kolda and B. Bader. Tensor Decompositions and Applications. *SIAM Review*, 51(3) :455–500, August 2009.
- [123] Ken Kreutz-Delgado. The complex gradient operator and the CR-calculus. *arXiv preprint arXiv :0906.4835*, 2009.
- [124] Jean-Philippe Lachaux, Eugenio Rodriguez, Jacques Martinerie, and Francisco J. Varela. Measuring phase synchrony in brain signals. *Human brain mapping*, 8(4) :194–208, 1999.
- [125] Preethi Lahoti, Kiran Garimella, and Aristides Gionis. Joint non-negative matrix factorization for learning ideological leaning on twitter. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 351–359, 2018.
- [126] Zhihui Lai, Yong Xu, Qingcai Chen, Jian Yang, and David Zhang. Multilinear sparse principal component analysis. *IEEE transactions on neural networks and learning systems*, 25(10) :1942–1950, 2014.
- [127] Michel Le Van Quyen, Jason Soss, Vincent Navarro, Richard Robertson, Mario Chavez, Michel Baulac, and Jacques Martinerie. Preictal state identification by synchronization changes in long-term intracranial EEG recordings. *Clinical Neurophysiology*, 116(3) :559–568, 2005.
- [128] Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pages 556–562, 2001.
- [129] Nora Leonardi and Dimitri Van De Ville. Identifying network correlates of brain states using tensor decompositions of whole-brain dynamic functional connectivity. In *2013 International Workshop on Pattern Recognition in Neuroimaging*, pages 74–77. IEEE, 2013.

- [130] Na Li, Stefan Kindermann, and Carmeliza Navasca. Some convergence results on the regularized alternating least-squares method for tensor decomposition. *Linear Algebra and its Applications*, 438(2) :796–812, 2013.
- [131] Xuelong Li, Guosheng Cui, and Yongsheng Dong. Graph regularized non-negative low-rank matrix factorization for image clustering. *IEEE transactions on cybernetics*, 47(11) :3840–3853, 2017.
- [132] Lek-Heng Lim and Pierre Comon. Nonnegative approximations of nonnegative tensors. *Journal of Chemometrics : A Journal of the Chemometrics Society*, 23(7-8) :432–441, 2009.
- [133] Chih-Jen Lin. Projected gradient methods for nonnegative matrix factorization. *Neural computation*, 19(10) :2756–2779, 2007.
- [134] Georg Lindgren. *Stationary stochastic processes : theory and applications*. Chapman and Hall/CRC, 2012.
- [135] Feng Liu, Yifeng Wang, Meiling Li, Wenqin Wang, Rong Li, Zhiqiang Zhang, Guangming Lu, and Huaifu Chen. Dynamic functional network connectivity in idiopathic generalized epilepsy with generalized tonic–clonic seizure. *Human brain mapping*, 38(2) :957–973, 2017.
- [136] Jialu Liu, Chi Wang, Jing Gao, and Jiawei Han. Multi-view clustering via joint nonnegative matrix factorization. In *Proceedings of the 2013 SIAM International Conference on Data Mining*, pages 252–260. SIAM, 2013.
- [137] Shuangzhe Liu and Götz Trenkler. Hadamard, Khatri-Rao, Kronecker and other matrix products. *International Journal of Information and Systems Sciences*, 4(1) :160–177, 2008.
- [138] Yuanyuan Ma, Xiaohua Hu, Tingting He, and Xingpeng Jiang. Multi-view clustering microbiome data by joint symmetric nonnegative matrix factorization with Laplacian regularization. In *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 625–630. IEEE, 2016.
- [139] Arash Golibagh Mahyari, David M. Zoltowski, Edward M. Bernat, and Selin Aviyente. A Tensor Decomposition-Based Approach for Detecting Dynamic Network States From EEG. *IEEE Transactions on Biomedical Engineering*, 64(1) :225–237, 2017.
- [140] Julien Mairal, F. Bach, J. Ponce, G. Sapiro, R. Jenatton, and G. Obozinski. SPAMS : A SPArse Modeling Software, v2. 3. URL <http://spams-devel.gforge.inria.fr/downloads.html>, 2014.
- [141] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, 11(Jan) :19–60, 2010.
- [142] Kenneth S. Miller. Complex gaussian processes. *Siam Review*, 11(4) :544–567, 1969.
- [143] Peter J. Mucha, Thomas Richardson, Kevin Macon, Mason A. Porter, and Jukka-Pekka Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *science*, 328(5980) :876–878, 2010.
- [144] Andrew C. Murphy, Shi Gu, Ankit N. Khambhati, Nicholas F. Wymbs, Scott T. Grafton, Theodore D. Satterthwaite, and Danielle S. Bassett. Explicitly linking regional activation and function connectivity : community structure of weighted networks with continuous annotation. *arXiv preprint arXiv :1611.07962*, 2016.
- [145] Mathilde Ménoret, Nicolas Farrugia, Bastien Padeloup, and Vincent Gripon. Evaluating graph signal processing for neuroimaging through classification and dimensionality reduction. In *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 618–622. IEEE, 2017.
- [146] Morten Mørup, Lars Kai Hansen, and Sidse M. Arnfred. Algorithms for sparse nonnegative Tucker decompositions. *Neural computation*, 20(8) :2112–2131, 2008.
- [147] Alexandre KW Navarro, Jes Frellsen, and Richard E. Turner. The multivariate generalised von Mises distribution : inference and applications. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
- [148] Mark Newman. *Networks*. Oxford university press, 2018.

- [149] Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2) :026113, 2004.
- [150] James Ng, Richard Prager, Nick Kingsbury, Graham Treece, and Andrew Gee. Modeling ultrasound imaging as a linear, shift-variant system. *ieee transactions on ultrasonics, ferroelectrics, and frequency control*, 53(3) :549–563, 2006.
- [151] Guido Nolte, Ou Bai, Lewis Wheaton, Zoltan Mari, Sherry Vorbach, and Mark Hallett. Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clinical neurophysiology*, 115(10) :2292–2307, 2004.
- [152] Paul L. Nunez and Ramesh Srinivasan. *Electric fields of the brain : the neurophysics of EEG*. Oxford University Press, USA, 2006.
- [153] Paul L. Nunez, Ramesh Srinivasan, Andrew F. Westdorp, Ranjith S. Wijesinghe, Don M. Tucker, Richard B. Silberstein, and Peter J. Cadusch. EEG coherency : I : statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalography and clinical neurophysiology*, 103(5) :499–515, 1997.
- [154] Khumaisa Nur'aini, Ibtisami Najahaty, Lina Hidayati, Hendri Murfi, and Siti Nurrohmah. Combination of singular value decomposition and K-means clustering methods for topic detection on Twitter. In *Advanced Computer Science and Information Systems (ICACSIS), 2015 International Conference on*, pages 123–128. IEEE, 2015.
- [155] Esa Ollila. Direction of arrival estimation using robust complex Lasso. In *2016 10th European Conference on Antennas and Propagation (EuCAP)*, pages 1–5. IEEE, 2016.
- [156] George C. O'Neill, Prejaas Tewarie, Diego Vidaurre, Lucrezia Liuzzi, Mark W. Woolrich, and Matthew J. Brookes. Dynamics of large-scale electrophysiological networks : A technical review. *Neuroimage*, 180 :559–576, 2018.
- [157] Guillermo Ortiz-Jiménez, Mario Coutino, Sundeep Prabhakar Chepuri, and Geert Leus. Sampling and Reconstruction of Signals on Product Graphs. In *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 713–717. IEEE, 2018.
- [158] Alp Ozdemir, Edward M. Bernat, and Selin Aiyente. Recursive tensor subspace tracking for dynamic brain network analysis. *IEEE Transactions on Signal and Information Processing over Networks*, 3(4) :669–682, 2017.
- [159] George C. O'Neill, Eleanor L. Barratt, Benjamin AE Hunt, Prejaas K. Tewarie, and Matthew J. Brookes. Measuring electrophysiological connectivity by power envelope correlation : a technical review on MEG methods. *Physics in Medicine & Biology*, 60(21) :R271, 2015.
- [160] J. Matias Palva, Sheng H. Wang, Satu Palva, Alexander Zhigalov, Simo Monto, Matthew J. Brookes, Jan-Mathijs Schoffelen, and Karim Jerbi. Ghost interactions in MEG/EEG source space : a note of caution on inter-areal coupling measures. *Neuroimage*, 173 :632–643, 2018.
- [161] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3) :127–239, 2014.
- [162] Bernard Picinbono. On instantaneous amplitude and phase of signals. *IEEE Transactions on signal processing*, 45(3) :552–560, 1997.
- [163] Francesco Pisana, Thomas Henzler, Stefan Schönberg, Ernst Klotz, Bernhard Schmidt, and Marc Kachelrieß. High quality high spatial resolution functional classification in low dose dynamic CT perfusion using singular value decomposition (SVD) and k-means clustering. In *Medical Imaging 2017 : Physics of Medical Imaging*, volume 10132, page 101320M. International Society for Optics and Photonics, 2017.
- [164] Nishant Rai, Sumit Negi, Santanu Chaudhury, and Om Deshmukh. Partial multi-view clustering using graph regularized NMF. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 2192–2197. IEEE, 2016.
- [165] Pradeep Ravikumar, Martin J. Wainwright, Garvesh Raskutti, and Bin Yu. High-dimensional covariance estimation by minimizing ℓ_1 -penalized log-determinant divergence. *Electronic Journal of Statistics*, 5 :935–980, 2011. ISBN : 1935-7524 Publisher : The Institute of Mathematical Statistics and the Bernoulli Society.

- [166] Reinhold Remmert. *Theory of complex functions*, volume 122. Springer Science & Business Media, 1991.
- [167] Jennifer Rizkallah, Hassan Amoud, Matteo Frascini, Fabrice Wendling, and Mahmoud Hassan. Exploring the Correlation Between M/EEG Source–Space and fMRI Networks at Rest. *Brain Topography*, pages 1–10, 2020.
- [168] Mikail Rubinov, Stuart A. Knock, Cornelis J. Stam, Sifis Micheloyannis, Anthony WF Harris, Leanne M. Williams, and Michael Breakspear. Small-world properties of nonlinear brain activity in schizophrenia. *Human brain mapping*, 30(2) :403–416, 2009.
- [169] Jorge M. Santos and Mark Embrechts. On the use of the adjusted rand index as a metric for evaluating supervised classification. In *International Conference on Artificial Neural Networks*, pages 175–184. Springer, 2009.
- [170] Björn Schelter, Matthias Winterhalder, Rainer Dahlhaus, Jürgen Kurths, and Jens Timmer. Partial phase synchronization for multivariate synchronizing systems. *Physical review letters*, 96(20) :208103, 2006.
- [171] Thomas Schreiber. Measuring information transfer. *Physical review letters*, 85(2) :461, 2000.
- [172] Andrey A. Shabalin and Andrew B. Nobel. Reconstruction of a low-rank matrix in the presence of Gaussian noise. *Journal of Multivariate Analysis*, 118 :67–76, 2013.
- [173] Fatemeh Sheikholeslami and Georgios B. Giannakis. Overlapping community detection via constrained parafac : A divide and conquer approach. In *2017 IEEE International Conference on Data Mining (ICDM)*, pages 127–136. IEEE, 2017.
- [174] Yanning Shen, Brian Baingana, and Georgios B. Giannakis. Tensor decompositions for identifying directed graph topologies and tracking dynamic networks. *IEEE Transactions on Signal Processing*, 65(14) :3675–3687, 2017.
- [175] Xianjin Shi, Wanwan Wang, and Chongsheng Zhang. An Empirical Comparison of Latest Data Clustering Algorithms with State-of-the-Art. *Indonesian Journal of Electrical Engineering and Computer Science*, 5(2) :410–415, 2017.
- [176] Nicholas D. Sidiropoulos, Lieven De Lathauwer, Xiao Fu, Kejun Huang, Evangelos E. Papalexakis, and Christos Faloutsos. Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing*, 65(13) :3551–3582, 2017.
- [177] Laurent Sorber, Marc Van Barel, and Lieven De Lathauwer. Optimization-based algorithms for tensor decompositions : Canonical polyadic decomposition, decomposition in rank-($L_r, L_r, 1$) terms, and a new generalization. *SIAM Journal on Optimization*, 23(2) :695–720, 2013.
- [178] Scott R. Sponheim, Kathryn A. McGuire, Seung Suk Kang, Nicholas D. Davenport, Selin Aviyente, Edward M. Bernat, and Kelvin O. Lim. Evidence of disrupted functional connectivity in the brain after combat-related blast injury. *Neuroimage*, 54 :S21–S29, 2011.
- [179] Loukianos Spyrou, Mario Parra, and Javier Escudero. Complex tensor factorization with PARAFAC2 for the estimation of brain connectivity from the EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(1) :1–12, 2018.
- [180] Cornelis J. Stam, Guido Nolte, and Andreas Daffertshofer. Phase lag index : assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Human brain mapping*, 28(11) :1178–1193, 2007.
- [181] Petre Stoica and Girish Ganesan. Linear regression constrained to a ball. *Digital Signal Processing*, 11(1) :80–90, 2001.
- [182] Liang Sun, Shuiwang Ji, Shipeng Yu, and Jieping Ye. On the equivalence between canonical correlation analysis and orthonormalized partial least squares. In *Twenty-First International Joint Conference on Artificial Intelligence*, 2009.
- [183] Arthur Tenenhaus and Michel Tenenhaus. Regularized generalized canonical correlation analysis. *Psychometrika*, 76(2) :257, 2011.
- [184] Robert Tibshirani. Regression shrinkage and selection via the lasso : a retrospective. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 73(3) :273–282, 2011.

- [185] Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 67(1) :91–108, 2005.
- [186] Federico Tomasi, Veronica Tozzo, Alessandro Verri, and Saverio Salzo. Forward-Backward Splitting for Time-Varying Graphical Models. In *International Conference on Probabilistic Graphical Models*, pages 475–486, 2018.
- [187] Jitendra K. Tugnait. Graphical LASSO for high-dimensional complex gaussian graphical model selection. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2952–2956. IEEE, 2019.
- [188] Viivi Uurtio, João M. Monteiro, Jaz Kandola, John Shawe-Taylor, Delmiro Fernandez-Reyes, and Juho Rousu. A tutorial on canonical correlation methods. *ACM Computing Surveys (CSUR)*, 50(6) :1–33, 2017.
- [189] Martijn P. van den Heuvel and Olaf Sporns. A cross-disorder connectome landscape of brain dysconnectivity. *Nature Reviews. Neuroscience*, 20(7) :435–446, 2019. Number : 7.
- [190] Charles F. Van Loan. The ubiquitous Kronecker product. *Journal of computational and applied mathematics*, 123(1-2) :85–100, 2000.
- [191] Pieter van Mierlo, Yvonne Höller, Niels K. Focke, and Serge Vulliemoz. Network Perspectives on Epilepsy Using EEG/MEG Source Connectivity. *Frontiers in Neurology*, 10 :721, 2019.
- [192] Pieter van Mierlo, Margarita Papadopoulou, Evelien Carrette, Paul Boon, Stefaan Vandenberghe, Kristl Vonck, and Daniele Marinazzo. Functional brain connectivity from EEG in epilepsy : Seizure prediction and epileptogenic focus localization. *Progress in neurobiology*, 121 :19–35, 2014.
- [193] Marisel Villafane-Delgado and Selin Aviyente. Dynamic graph fourier transform on temporal functional connectivity networks. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 949–953. IEEE, 2017.
- [194] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4) :395–416, 2007.
- [195] Michael E. Wall, Andreas Rechtsteiner, and Luis M. Rocha. Singular value decomposition and principal component analysis. In *A practical approach to microarray data analysis*, pages 91–109. Springer, 2003.
- [196] Hong-Qiang Wang, Chun-Hou Zheng, and Xing-Ming Zhao. j NMFMA : a joint non-negative matrix factorization meta-analysis of transcriptomics data. *Bioinformatics*, 31(4) :572–580, 2015.
- [197] Jeremy Watt, Reza Borhani, and Aggelos K. Katsaggelos. *Machine learning refined : foundations, algorithms, and applications*. Cambridge University Press, 2016.
- [198] F. Wendling, F. Bartolomei, J. J. Bellanger, and P. Chauvel. Identification of epileptogenic networks from modeling and nonlinear analysis of SEEG signals. *Neurophysiologie clinique=Clinical neurophysiology*, 31(3) :139–151, 2001.
- [199] Fabrice Wendling, Fabrice Bartolomei, Jean-Jacques Bellanger, Jérôme Bourien, and Patrick Chauvel. Epileptic fast intracerebral EEG activity : evidence for spatial decorrelation at seizure onset. *Brain*, 126(6) :1449–1459, 2003.
- [200] Fabrice Wendling, Jean-Jacques Bellanger, Fabrice Bartolomei, and Patrick Chauvel. Relevance of nonlinear lumped-parameter models in the analysis of depth-EEG epileptic signals. *Biological cybernetics*, 83(4) :367–378, 2000. ISBN : 0340-1200 Publisher : Springer.
- [201] Joe Whittaker. *Graphical models in applied multivariate statistics*. Wiley Publishing, 2009.
- [202] Junjie Wu, Hui Xiong, and Jian Chen. Adapting the right measures for k-means clustering. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 877–886. ACM, 2009.
- [203] Yangyang Xu. On the convergence of higher-order orthogonal iteration. *Linear and Multilinear Algebra*, 66(11) :2247–2265, 2018.

- [204] Dan Yang, Zongming Ma, and Andreas Buja. A sparse singular value decomposition method for high-dimensional data. *Journal of Computational and Graphical Statistics*, 23(4) :923–942, 2014.
- [205] Jaewon Yang and Jure Leskovec. Overlapping community detection at scale : a nonnegative matrix factorization approach. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 587–596. ACM, 2013.
- [206] Ke Ye and Lek-Heng Lim. Distance between subspaces of different dimensions. *arXiv preprint arXiv :1407.0900*, 4, 2014.
- [207] Anru Zhang and Dong Xia. Tensor SVD : Statistical and Computational Limits. *IEEE Transactions on Information Theory*, 2018.
- [208] Qian Zhao, Deyu Meng, Zongben Xu, Wangmeng Zuo, and Yan Yan. L1-norm low-rank matrix factorization by variational Bayesian method. *IEEE Trans. Neural Netw. Learning Syst.*, 26(4) :825–839, 2015.
- [209] Qinpei Zhao. Cluster validity in clustering methods. *Publications of the University of Eastern Finland*, 2012.
- [210] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the royal statistical society : series B (statistical methodology)*, 67(2) :301–320, 2005.