

MODÈLES DE MARKOV CACHÉS MULTICANAUX POUR LA MODÉLISATION DE SIGNAUX EEG ET MEG

PROJET DE THÈSE

Ecole doctorale : Ecole doctorale de Mathématiques et Informatique, ED 184

Unité de recherches : Laboratoire d'Analyse, Topologie et Probabilités (LATP), UMR CNRS 6632, équipe *probabilités et statistique*, groupe de traitement du signal.

Localisation : Centre de Mathématiques et Informatique, 39 rue Joliot-Curie, 13453 Marseille Cedex 13.

Directeur de thèse : Bruno TORRÉSANI

Courriel : Bruno.Torresani@cmi.univ-mrs.fr

Titre de la thèse : Modèles de Markov cachés multicanaux pour la modélisation de signaux EEG et MEG.

Description du projet : L'objectif de ce projet est de développer des modèles de chaînes de Markov cachées, et éventuellement des généralisations, dans le cadre de l'analyse de signaux d'électroencéphalographie et magnétoencéphalographie, et plus précisément l'analyse par essai unique. Cette thématique se développe depuis deux ans au Laboratoire d'Analyse, Topologie et Probabilités de Marseille, dans le cadre d'un programme de recherches impliquant plusieurs laboratoires Marseillais de neurosciences.

Les signaux EEG et MEG sont des signaux multicateurs (c'est à dire à valeurs vectorielles), et généralement multi-essais (c'est à dire issus de mesures répétées). Ils sont souvent analysés via une moyenne sur les essais, qui ne permet pas de faire apparaître la variabilité inter-essais, qui est cependant souvent importante. Il existe très peu d'approches permettant d'appréhender simultanément l'aspect multi-capteur et l'aspect multi-essai. Ce projet se place dans le cadre de telles approches.

L'une des caractéristiques des signaux EEG et MEG est leur non-stationnarité, ce qui justifie l'utilisation de modèles de Markov cachés. De tels modèles ont déjà été utilisés à des fins de classification (Súrastný *et al* 2002), beaucoup moins dans un but d'analyse. On se propose de développer des modèles de chaînes de Markov cachées multicanaux intégrant divers types d'écart à la stationnarité, associés à des sources biologiquement plausibles.

Une seconde difficulté inhérente aux signaux EEG et MEG est la présence d'artefacts (comme ceux résultant de clignement d'yeux ou de bruits non physiologiques), ou de signaux considérés comme parasites, car gênants pour la caractérisation des signaux d'intérêt (comme par exemple les ondes alpha, qui apparaissent dans certaines phases de perte d'attention du sujet). On se propose de développer des techniques permettant la suppression de ces signaux parasites.

A l'heure actuelle, le projet est basé sur deux types de modèles de Markov cachés. D'une part les modèles dynamiques décrits dans (Penny et Roberts, 1999), dans lesquels les observations sont engendrées par un modèle autorégressif multivarié (MAR). Ces modèles permettent une analyse spectrale non-stationnaire, susceptible de caractériser des régions de comportement différent dans les signaux. D'autre part, les modèles de Markov cachés harmoniques développés dans (Molla et Torrèsani 2005) dans un contexte d'analyse de signaux audiophoniques. Ces modèles ont l'avantage

de permettre des synthèses partielles de différentes composantes, ce qui peut avoir son intérêt pour la suppression de certains types d'artefacts ou signaux indésirables.

Il s'agira dans un premier temps de développer des versions "multi-capteur" de ces modèles, et d'écrire les problèmes d'estimation correspondants : estimation de paramètres et estimation des états cachés. Dans le cadre du modèle de Markov caché harmonique, on s'intéressera également aux problèmes de resynthèse partielle du signal, plus précisément de synthèse de certaines composantes spécifiques (ou couches).

On s'intéressera également aux modèles faisant intervenir des notions de *parcimonie* de la décomposition du signal dans une base bien choisie, dans l'esprit de l'approche développée dans (Kowalski et Torrèsani 2008). Il s'agira là encore d'approfondir ce type d'approche dans un cadre de signaux multi-canaux, puis de clarifier les relations avec les modèles Markoviens, ainsi qu'avec les méthodes de régression parcimonieuses structurées introduites notamment dans (Fornasier et Rauhut 2008) et (Kowalski et Torrèsani 2008b).

Le projet comprend également un volet "applications numériques". Les méthodes développées seront implémentées (en utilisant, au moins dans un premier temps, l'environnement Matlab, ce qui permettra d'exploiter des codes déjà développés dans l'équipe, puis éventuellement en C/C++). Elles seront également testées et validées sur des jeux de données synthétiques et réels, fournis par les collaborateurs neurobiologistes.

Ce projet de recherches, qui se développera au sien du groupe de traitement du signal du Laboratoire d'Analyse, Topologie et Probabilités de Marseille, se place dans le cadre de collaborations avec plusieurs groupes de neurosciences, en particulier le Laboratoire de Neurobiologie de la Cognition (Marseille, UMR CNRS 6155), le Laboratoire de Neurophysiologie et de Neuropsychologie, Service de Neurophysiologie Clinique (Marseille, unité INSERM U.751), et le projet ODYSSEE (INRIA, Sophia Antipolis). Les jeux de données (simulés et réels) seront fournis par ces collaborations.

Références :

- (Penny et Roberts 1999) : William D. Penny et Stephen J. Roberts, Dynamic Models for Nonstationary Signal Segmentation, *Computers and Biomedical Research* **32** :6 (1999), pp. 483-502
- (Štrastný *et al* 2002) : J. Štrastný, P. Sovka, et A. Stancák, EEG signal classification and segmentation by means of hidden Markov models, Proceedings of 16th Biennial International EURASIP Conference on *Analysis of Biomedical Signals and Images* (BIOSIGNAL '02), Brno, Czech Republic (2002), pp. 415-417
- (Molla et Torrèsani 2005) : Stéphane Molla et Bruno Torrèsani, An hybrid audio scheme using hidden Markov models of waveforms, *Applied and Computational Harmonic Analysis* **18** :2 (2005), pp.137-166.
- (Kowalski et Torrèsani 2008) : Matthieu Kowalski et Bruno Torrèsani, Random Models for Sparse Signals Expansion on Unions of Bases with Application to Audio Signals, *IEEE Transactions on Information Theory*, à paraître (2008).
- (Fornasier et Rauhut 2008) : Massimo Fornasier et Holger Rauhut, Recovery algorithms for vector valued data with joint sparsity constraints, *SIAM Journal on Numerical Analysis* (2008), à paraître.
- (Kowalski et Torrèsani 2008b) : Matthieu Kowalski et Bruno Torrèsani, Sparse and persistent multi-channel expansions, with application to EEG signals, soumis à EUSIPCO'08 (the *European Signal Processing Conference* (2008).

Connaissances et compétences requises : Le candidat devra avoir validé un master en mathématiques, et avoir des connaissances en probabilités et statistique, notamment sur les chaînes de Markov. Il devra également maîtriser des éléments de programmation (matlab et/ou C/C++).