

# Modèles auto-regressifs avancés pour le traitement de données MEG/EEG

**Directeur :** Théo Papadopoulo (Theodore.Papadopoulo@inria.fr)

**Co-Encadrant :** Rachid Deriche (Rachid.Deriche@inria.fr)

**Lieu :** Équipe Athena, INRIA Sophia Antipolis

## Cadre général

L'électro-encéphalographie (EEG) et la magnéto-encéphalographie (MEG) sont deux modalités qui permettent la mesure de façon passive à l'extérieur de la tête du potentiel électrique et du champ magnétique dûs principalement à l'activité électrique du cerveau. Ces modalités sont utilisées pour caractériser différentes pathologies du cerveau telles que, par exemple, l'épilepsie. Les mesures obtenues aux capteurs à l'instant  $t$  sont des combinaisons linéaires des activités électriques à ce même instant  $t$  d'un ensemble de "sources" au niveau du cortex cérébral (modèle direct). Ces sources dépendent dans le temps les une des autres suivant un modèle biophysique complexe mal connu (car il dépend de nombreux paramètres) et qui n'est pas donc aisé à établir.

Pour certaines applications (par exemple pour la planification chirurgicale), il est nécessaire de remonter aux sources électriques au niveau du cortex (problème inverse de reconstruction de sources). Cette approche est intellectuellement satisfaisante, mais nécessite l'utilisation d'un modèle adapté au sujet pour être totalement pertinente, ce qui est à la fois couteux et compliqué à obtenir. Or, il est souvent possible d'exploiter directement les mesures EEG ou MEG au niveau des capteurs. Par exemple, de nombreux systèmes d'interfaces cerveau-ordinateur se servent des seules données capteurs pour classifier l'activité du cerveau : la classe détectée sert alors à déterminer la tâche que doit effectuer l'ordinateur. Il existe donc au niveau des capteurs des "traces" de l'activité au niveau des sources qui sont exploitables sans avoir besoin de modèles de têtes compliqués à obtenir.

## Sujet de la thèse

Comme indiqué ci-dessus, la dépendance temporelle entre les sources corticales reste mal connue et difficile à obtenir, du moins si l'on cherche à coller de près aux modèles biophysiques. Il existe néanmoins des modèles simples qui, même s'ils ne sont pas complètement justifiés par la biologie, sont suffisamment généraux pour être utilisés avec les mesures EEG/MEG : les modèles auto-regressifs [BP20]. Ces modèles permettent de prédire l'activité au temps  $t$  à partir des activités aux instants précédents. Ce sujet vise donc à aborder le problème général de l'utilisation de modèles auto-regressifs pour l'exploitation des données EEG/MEG au niveau des capteurs (contrairement au travail [BP20] qui considère des modèles auto-regressifs au niveau des sources). Le parti pris de travailler au niveau des capteurs se justifie par la moindre complexité de travailler à ce niveau et par le fait que tout ce qui peut être appris au niveau des capteurs pourra être ensuite utilisé pour contraindre des modèles au niveau des sources avec l'espoir d'en réduire la complexité. Dans ce cadre, plusieurs problèmes intéressants se posent pour des modèles auto-régressifs au niveau des capteurs :

1. **Segmentation des données EEG/MEG :** L'existence même d'un modèle auto-régressif est conditionnée par le fait que les sources suivent un modèle stationnaire. Or dans le cerveau, le lien entre les sources change avec la tâche (et même au cours d'une même tâche pour les plus complexes d'entre elles). Comment segmenter les données acquises en tronçons pour lesquels on peut obtenir un modèle auto-regressif stable ? Alternativement, on peut estimer un modèle auto-régressif variable dans le temps via l'utilisation de fenêtres de signal.
2. **Comment calculer efficacement ces modèles auto-regressifs :** Il existe des méthodes bien établies pour estimer des modèles auto-regressifs (Algorithmes de Yule-Walker, algorithme de Burg). Comment fonctionnent ils avec des données EEG/MEG qui sont très bruitées ? Pour palier à ce bruit, une même expérience est répétée plusieurs fois (plusieurs essais) que l'on moyenne pour augmenter le rapport signal à bruit. Peut on exploiter ces essais mieux que regarder leur seule moyenne (par exemple en exploitant des idées de type dictionary learning [HCS<sup>+</sup>17, PHP17]) ? Peut on remonter à de la variabilité inter essais ? Dans le cadre des interfaces cerveau-ordinateur, on a besoin d'algorithmes qui fonctionnent en temps réel et en essai par essai (donc sans possibilité de faire une moyenne sur plusieurs essais), que peut-on faire avec des modèles auto-régressifs dans ce cadre ?

3. **Lien entre un modèle au niveau des sources et un autre au niveau des capteurs :** Que peut on apprendre à partir des données capteurs sur les sources ? Quelles caractéristiques du modèle sont invariantes lors de la transformation par le modèle direct. Sous certaines hypothèses, l'ordre du modèle auto-regressif reste le même. Y en a-t-il d'autres ? On regardera via des simulations quelles propriétés sont conservées en pratique.
4. **Lien avec les modèles temps-fréquence :** Une méthode d'analyse classique des données EEG/MEG est de regarder les données via des diagrammes temps-fréquence. Par comparaison, un modèle auto-regressif estime un nombre fini de fréquences qui permettent de modéliser le signal. Il existe un lien fort entre les modèles auto-regressifs, leur interprétation dans le domaine fréquentiel et la causalité de Granger. Même si l'utilisation de ces notions de causalité n'est pas extrêmement pertinente pour des analyses au niveau des capteurs, on peut comme pour la question précédente se poser des questions sur ce qui peut être appris au niveau des capteurs sur les sources...
5. **Classification de l'activité sur la base de modèles auto-regressifs :** Quelles métriques peut on utiliser sur l'espace des modèles auto-régressifs pour classifier des activités ? Avec quelles garanties ? Ces problèmes ont un intérêt direct pour des applications aussi bien cognitive (comprendre le fonctionnement du cerveau) que pour des applications aux interfaces cerveau-ordinateur. On pourra notamment explorer l'apport des techniques récentes d'apprentissage machine pour ce problème de classification.

Bien entendu, nous n'aborderons pas tous ces sujets dans le cadre de la thèse. On s'intéressera principalement aux trois premières questions avec la possibilité d'explorer l'une des deux autres en fonction des appétences du/de la candidat.e ou de ce qui semblera prometteur au vu de l'étude des deux premiers points...

Ces travaux se baseront sur l'utilisation de données disponibles dans l'équipe ATHENA, ou de logiciels tels qu'OpenMEEG ou mne-python. Les nouveaux développements se feront en langage python ou C++ (si il devient nécessaire de rendre les calculs plus efficaces).

## Profil recherché

Un Master M2 en Maths Applis/Informatique/Machine Learning/Imagerie Médicale avec de bonnes, et de préférence solides, connaissances en traitement du signal et apprentissage statistique ainsi qu'une bonne connaissance de l'anglais et de la programmation en Python ou C++.

## References

- [Bel17] Brahim Belaoucha. *Using diffusion MR information to reconstruct networks of brain activations from MEG and EEG measurements*. Theses, Université Côte d'Azur, June 2017.
- [BP20] Brahim Belaoucha and Théodore Papadopoulos. Structural connectivity to reconstruct brain activation and effective connectivity between brain regions. *Journal of Neural Engineering*, 17(3):035006, June 2020.
- [HCS<sup>+</sup>17] S. Hitziger, M. Clerc, S. Sallet, C. Bénar, and T. Papadopoulos. Adaptive waveform learning: A framework for modeling variability in neurophysiological signals. *IEEE Transactions on Signal Processing*, 65(16):4324–4338, April 2017.
- [Hit14] Sebastian Hitziger. *Modeling the Variability of Electrical Activity in the Brain*. PhD thesis, University of Nice-Sophia Antipolis, April 2014.
- [PHP17] Christos Papageorgakis, Sebastian Hitziger, and Théodore Papadopoulos. Dictionary learning for multidimensional data. In *Proceedings of GRETSI 2017*, Juan-Les-Pins, September 2017.