Advanced auto-regressive models for processing EEG/MEG data

Supervisor : Théo Papadopoulo (Theodore.Papadopoulo@inria.fr)
Co-Supervisor : Rachid Deriche (Rachid.Deriche@inria.fr)
Place : Athena team, INRIA Sophia Antipolis

General framework

Electroencephalography (EEG) and magnetoencephalography (MEG) are two modalities that allow the passive measurement outside the head of the electric potential and the magnetic field arising mainly from the electrical activity of the brain. These modalities are used to characterize different brain pathologies such as, for example, epilepsy. The measurements obtained at sensors at the time t are linear combinations of the electrical activities at this same time t of a set of "sources" located at the level of the cerebral cortex (direct model). These sources are time dependent on each other according to a complex biophysical model that is poorly understood (because it depends on many parameters) and which is therefore not easy to establish.

For some applications (for example for surgical planning), it is necessary to recover the cortical electrical sources (inverse problem of source reconstruction). This approach is intellectually satisfying, but requires the use of a model adapted to the subject to be fully relevant, which is both expensive and complicated to get. However, it is often possible to directly use the EEG or MEG measurements at the level of the sensors. For example, many brain-computer interface systems use sensor data alone to classify brain activity: the class detected is then used to determine the task to be performed by the computer. There are therefore at the level of the sensors "traces" of activity at the source level which are exploitable without the need for complicated to obtain head models.

Ph.D. subject

As indicated above, the temporal dependence between cortical sources remains poorly understood and difficult to obtain, at least if we are trying to stick closely to biophysical models. There are, however, simple models which, even if they are not completely justified by biology, are general enough to be used with EEG/MEG measurements: auto-regressive models [BP20]. These models make it possible to predict the activity at time t from the activities at the previous time instants. This subject therefore aims to address the general problem of using auto-regressive models for modeling and analyzing EEG/MEG data at the level of the sensors (unlike in the work [BP20] which considers auto-regressive models at the level of sources). The decision to work at the level of sensors is justified by the lesser complexity of working at this level and by the fact that everything that can be learned at the level of sensors can be then used to constrain models at the source level with the hope of reducing their complexity. In this context, several interesting problems arise for auto-regressive models at the level of sensors:

1. **Segmentation of EEG/MEG data :** The very existence of an auto-regressive model is conditioned by the fact that sources follow a stationary model. However, in the brain, the link between sources changes with the task (and even during the same task for the most complex ones). How to segment the acquired data into sections for which a stable autoregressive model can be obtained? Alternatively, can we estimate an auto-regressive model varying in time through the use of signal windows?

2. **How to estimate efficiently these auto-regressive models:** There are well-established methods for estimating auto-regressive models (Yule-Walker algorithms, Burg's algorithm). How do they work with EEG/MEG data which is very noisy? To overcome this noise, the same experiment is repeated several times (several trials) that are averaged to increase the signal to noise ratio. Can we use these trials better than looking at their average alone (for example by using ideas of dictionary learning type [HCS+17, PHP17])? Can we recover to inter-trial variability? In the context of brain-computer interfaces, we need algorithms that operate in real time and on a trial-by-trial basis (therefore without the possibility of taking an average over several trials). What can be done with autoregressive models in this framework?

3. **Link between models at the source and sensor level:** What can we learn from sensor data on sources? Which characteristics of the model are invariant to the transformation by the direct model. Under certain assumptions, the order of the auto-regressive model remains the same. Are there others invariants? We will look through simulations which properties are preserved in practice.

4. **Link with time-frequency analysis:** A classic method for analyzing EEG/MEG data is to look at the data via time-frequency diagrams. By comparison, an auto-regressive model estimates a finite number of frequencies that allow to model the signal. There is a strong link between auto-regressive models, their interpretation in the frequency domain and Granger causality. Even if the use of these notions of causality is not extremely relevant for analyzes at the level of sensors, we can as in the previous point ask questions about what can be learned on sources from studying things at the level of sensors...

5. **Classification of brain activity based on auto-regressive models:** What metrics can be used on the space of auto-regressive models for classifying activities? With what guarantees? These problems are of direct interest to cognitive applications (understanding how the brain works) and applications to brain-computer interfaces. We will explore in particular the contribution of recent machine learning techniques for this classification problem.

Obviously, all these subjects cannit be tackled within the framework of a single Ph.D. thesis. We will mainly focus on the first three questions with the possibility of exploring one of the other two according to the interests of the candidate or to what will seem promising in view of the study of the first two points ...

This work will be based on the use of data available in the ATHENA team, and on software such as that OpenMEEG or mne-python. New developments will be done in python or C++ language (if it becomes necessary to make calculations more efficient).

Profile of the candidate

A Master M2 in Applied Maths/Computer Science/Machine Learning/Medical Imaging with good, and preferably solid, knowledge in signal processing and statistical learning as well as a good knowledge of English and programming in Python or 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 Papadopoulo. Structural connectivity to reconstruct brain activation and effective connectivity between brain regions. *Journal of Neural Engineering*, 17(3):035006, June 2020.

[HCS+17] S. Hitziger, M. Clerc, S. Saillet, C. Bénar, and T. Papadopoulo. 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 Papadopoulo. Dictionary learning for multidimensional data. In *Proceedings of GRETSI 2017*, Juan-Les-Pins, September 2017.