Multi-resolution analysis of EEG signals through Hierarchical Independent Component Analysis

Piercesare Secchi[1], **Simone Vantini**[1], **Paolo Zanini**[1]
[1] MOX – Department of Mathematics, Politecnico di Milano, Italy

The analysis of high-dimensional and complex dataset is nowadays a very hot topic in statistics. In particular two problems are often considered: dimension reduction and a concise description of data through a phenomenological interpretation of the problem under study. We look for a reduced space where to represent data through a basis whose elements highlight the most important features of the phenomenon analyzed. A very important characteristic in this sense is multi-resolution, since some of the main features may involve a great number of the original variables while others may be restricted only to a few. Multi-resolution is a crucial property in several applicative scenarios. In particular in this work we analyze subjects affected by alcoholism through their EEG signals. For each patient in the study, measurements from 61 electrodes placed at standard sites on the scalp are available. For each electrode, the recorded signal measures the electrode electric potential with respect to some reference electrode and describes the electrical activity of the brain in the neighborhood of the electrode across time. In particular we analyze the brain signals related to one patient which was exposed to two stimuli: the patient was shown two identical pictures in a 1 second time lag. For each electrode, we observe the signal at 256 equally spaced instants. We consider that time indexes the sample of observations of a random vector whose components are indexed by electrodes. Hence we consider $n = 256$ realizations of a random vector $\mathbf{X}$ in $\mathbb{R}^p$, whith $p = 61$. The model we take into account in this analysis for the component X_{ij} of $\mathbf{X}$ (i.e., for the signal of electrode j at time instant i) reads as follows:

$$X_{ij} = S_{i1}a_{j1} + ... + S_{iK}a_{jK},$$

where $S_{i1}, ..., S_{iK}$ are the components of a latent source vector $\mathbf{S} \in \mathbb{R}^K$, while $a_{j1}, ..., a_{jK}$ represent unknown real coefficients. This model fits in the well known Blind Source Separation (BSS) framework. Now consider the rows of the $n \times p$ matrix $\mathbb{X}$ represent n observed realizations $\mathbf{x}_1, ..., \mathbf{x}_n \in \mathbb{R}^p$ of the random vector $\mathbf{X}$, while the rows of the $n \times K$ matrix $\mathbb{S}$ gather the corresponding unobserved realizations of the latent random vector $\mathbf{S}$, a BSS problem can be written as

$$\mathbb{X} = \mathbb{S}A^T, \tag{1}$$

where A is a $p \times K$ matrix of unknown coefficients. A BSS problem consists in the estimate of A and $\mathbb{S}$, given $\mathbb{X}$. Referring to the EEG analysis, the columns of A are spatial maps of the brain associated to some reference elements generating the whole signal and in this context multi-resolution represents a very interesting property. Indeed, some brain processes could involve the whole brain, while others activities involve only a specific part of brain. Imposing multi-resolution property on the elements of A implies the potential for the identification of a wide range of different behaviors characterizing the brain activity. In this work we solve the BSS through Hierarchical Independent Component Analysis (HICA), an innovative method for the construction of a multi-resolution data-driven basis obtained merging the idea of Treelets (see Lee et al. (2008) for the details) and Independent Component Analysis (see Hyvarinen and Oja (2000) for the details). The HICA algorithm is a hierarchical procedure that at each level l of the tree provides a multi-resolution non-orthogonal data-driven basis (i.e., an estimate $\widehat{A}^{(l)}$ for the basis matrix A) and an estimate for the source matrix $\mathbb{S}$, say $\widehat{\mathbb{S}}^{(l)}$. A detailed description of the algorithm, strategies for choosing the optimal estimates $\widehat{A}^{(l)}$ and $\widehat{\mathbb{S}}^{(l)}$ as well as selecting the appropriate dimension K for the source space, are discussed in Secchi et al. (2014), where we also prove some consistency properties of the HICA algorithm.

In the EEG application we compare the results obtained by HICA with those provided by other BSS techniques, as Treelets, ICA (exploiting the fastICA algorithm), and PCA. We show how the properties characterizing the HICA solution, allow to obtain noticeable improvements in terms of phenomenological interpretation. In particular we show that HICA is the only method able to split the occipital cerebral hemisphere into two separate components. It highlights both the primary visual cortex, responsible for the stimulus analysis in terms of shape and pattern recognition, and the internal area of the occipital region, which associates to the stimulus specific features like color, direction and origin.

Keywords: Blind Source Separation, Independent Component Analysis, multi-resolution, EEG.

References:

Hyvarinen, A., Oja, E. (2000) Independent Component Analysis: Algorithms and Applications. *Neural Networks 13*, 411-430.

Lee, A. B., Nadler, B., Wasserman L. (2008) Treelets - an adaptive multi-scale basis for sparse unordered data *The Annals of Applied Statistics 2*, 435-471.

Secchi, P., Vantini, S., Zanini, P. (2014). Hierarchical Independent Component Analysis: a multi- resolution non-orthogonal data-driven basis *MOX report 01/2014, MOX – Department of Mathematics, Politecnico di Milano*.