Multivariate approach for brain decomposable connectivity networks

Florent Chatelain, Sophie Achard, Olivier J.J. Michel, Cedric Gouy-Pailler

MULTIVARIATE APPROACH FOR BRAIN DECOMPOSABLE CONNECTIVITY NETWORKS

F. Chatelain, S. Achard, O. Michel*

GIPSA-lab, CNRS, Université de Grenoble,
38000 Grenoble, FRANCE

C. Gouy-Pailler

CEA, LIST,
Data Analysis Tools Laboratory,
91191 Gif-sur-Yvette CEDEX, FRANCE

ABSTRACT

This paper deals with the analysis of brain functional network using fMRI data. It recapitulates the concept of decomposable connectivity graph. Graphs are a usual tool to represent complex systems behavior, although edge strength estimation issues have not yet received a universally adopted solution. In the framework of linear Gaussian instantaneous exchanges, the well known partial correlation is usually introduced. However its estimation remains a challenge for highly connected or dense systems. Here, we propose to combine a wavelet decomposition and a graphical Gaussian model approach relying on decomposable graphs. This is shown to improve the estimations of brain function networks in the presence of long range dependence; the results are compared to those obtained with classical partial correlation estimators.

Index Terms— functional MRI, brain connectivity, partial correlation, graphical Gaussian model, maximum likelihood, decomposable graphs

1. INTRODUCTION

The observation and description of the living brain has attracted a lot of research during the last centuries. First, it has been shown that brain areas can be associated to a specific function, which is called *phrenology*. However, during the 19th century, a *connectionist* approach was introduced in order to model the brain as a global complex function. Since the development of functional Magnetic Resonance Imaging (fMRI), it is possible to observe the whole living brain during several minutes with a high spatial resolution (voxels up to one mm^3) and a temporal resolution of one or two seconds. Consequently, some authors [?] attempt to model and characterize brain functional connectivity networks.

Indeed, neuroimagery provides us with time series, associated with voxels, which correspond to the information processed by a brain area in time. These data have brought to light the dynamic nature of the brain, which follows complex time patterns. The complexity of the signals obtained from fMRI prevents from using classical statistical techniques to analyse the dependence between time series. Prior works [?, ?, ?] have shown that brain dynamics are typically long memory with fractal scaling properties, and that brain function networks have a complex, modular, small-world organization which is topologically conserved over multiple frequency scales.

The approach introduced in [?] consisted in using wavelets to take into account the long memory property of the fMRI time series after averaging the voxels for a given anatomical brain region.

Then, the brain function networks were extracted using pairwise correlations of the wavelet coefficients. Thanks to this approach, the fMRI time series can be assumed to be Gaussian, and we propose to combine a wavelet decomposition and a graphical Gaussian model approach [?] in order to improve the estimations of brain function networks.

Section 2 presents the concept of conditional dependence, and its relation to correlation and partial correlation estimators. In section 3, we recall the maximum likelihood estimator for graphical Gaussian models. Finally, in section 4, the two approaches are compared on simulated multivariate data with long memory properties and a fixed model of connections extracted from real fMRI data.

In the sequel of the paper, all the estimations will be based on the wavelet coefficients extracted at a given scale. We will denote $(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^n)$, the wavelet coefficients obtained at a specific scale. In the context of maximum likelihood estimations, $(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^n)$ is assumed to be independent which is a reasonable hypothesis since it has been shown that for long memory time series the wavelet coefficients are stationary and asymptotically decorrelated [?].

2. INSTANTANEOUS CONDITIONAL DEPENDENCE

Let x_i^t, $t \in \{1, \ldots, n\}$ be the time series associated to the voxel of index i, $i \in V = \{1, \ldots, p\}$. Connectivity between voxels will be assessed by studying dependences between the associated time series. For practical issues, dependence analysis is often restricted to linear dependence between time series [?], and the analysis is summarized by the variance or covariance matrix of the p dimensional multivariate process $\boldsymbol{x}^t = (x_1^t, \ldots, x_p^t)^T$. An appealing approach is to relate high correlation value (eventually normalized) to the existence of an 'information exchange' between the time series. This leads to a major pitfall: two series may seem dependent of each other and exhibit a high correlation, although they are not 'connected'. This situation appears if both voxels receive information from a common third one, whilst they are independent conditionally to that voxel. The importance of conditioning was already quoted in Granger's seminal work [?], and was recently revisited in a non Gaussian framework, for directional dependences [?].

In this paper, linear dependence model and Gaussian processes are considered. Therefore, covariance provides an exhaustive representation of the available information. Furthermore, only instantaneous dependence is considered here. All possible conditioning in the instantaneous dependence between two series (x_i^t, x_j^t) is taken into account by evaluating the correlations between the innovations

processes ε_i^t and ε_j^t defined below. Let

$$\mathcal{H}_k^{V \setminus \{i,j\}} = \operatorname{span}\left\{ x_l^k, k \in V \setminus \{i,j\} \right\}$$

be the Hilbert subspace spanned all the observations but x_i^t and x_j^t. The best (MS sense) linear estimator of x_i^t, given $\mathcal{H}_t^{V \setminus \{i,j\}}$ writes

$$\hat{x}_i^t = \mathcal{P}\left(x_i^t | \mathcal{H}_t^{V \setminus \{i,j\}} \right)$$

where $\mathcal{P}(x|\mathcal{H})$ denotes the orthogonal projection of x on $\mathcal{H}$. The corresponding innovation is given by $\varepsilon_i^t = \hat{x}_i^t - x_i^t$. The *partial correlation* $p_{i,j}$ between the two series is defined as

$$\begin{cases} k_{i,j} = -\operatorname{cov}[\varepsilon_i^t, \varepsilon_j^t] \\ p_{i,j} = \frac{k_{i,j}}{\sqrt{k_{i,i} k_{j,j}}} \end{cases} \quad \text{for all } 1 \leq i, j \leq p \quad (1)$$

Equation (1) is an important result from multivariate analysis. It states that the components of $[P]_{i,j} = p_{i,j}$ for the partial correlation matrix are directly derived from the $k_{i,j} = [K]_{i,j} = -[\Sigma^{-1}]_{i,j}$ for $i \neq j$, where Σ^{-1} is the inverse of the covariance matrix of $\boldsymbol{x}^t$ [?]. Note that the partial correlation matrix P satisfies $-1 \leq p_{i,j} \leq 1$, $\forall \{i,j\} \in V^2$. Estimating P may be tackled in two ways; either it is inferred by computing the inverse of the sample covariance matrix Γ or it is directly estimated. In the former case, the estimator will be noted $\widehat{P}_{\text{emp}}$. This paper focuses on the latter approach, where a robust estimation of P based on decomposable graph representation is derived in the context of long range dependent process.

3. MAXIMUM LIKELIHOOD ESTIMATION FOR GRAPHICAL MODELS

Graphical models are classically used to characterise conditional independence constraints between random variables. Let $G = (V, E)$ be an undirected graph, where $V = \{1, ..., p\}$ denotes the set of nodes, while E is the set of undirected edges $(i,j) \in V \times V$. Then, a multivariate model indexed by V is said to be Markov with respect to the graph G if for any edge (i,j) in E, the ith and jth variables are conditionally independent given all the other variables (see [?] for more details on graphical models).

3.1. Graphical Gaussian Models

Let $\boldsymbol{x}$ be a p-dimensional centered Gaussian vector $\mathcal{N}_p(0, \Sigma)$ whose elements are indexed by V. Define $K = \Sigma^{-1}$ to be the so-called precision matrix of $\boldsymbol{x}$. A fundamental result about graphical Gaussian models is that the Markov property induced by the graph G is equivalent to impose some sparsity constraints on $K = (k_{i,j})_{1 \leq i, j \leq p}$:

$$k_{i,j} = 0 \text{ for all } (i,j) \notin E. \quad (2)$$

Consider now a sample $(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^n)$ from the graphical Gaussian model associated with G. From the density expression of a Gaussian vector, the likelihood is expressed as

$$p(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^n; K) \propto \det(K)^{n/2} \exp\left[-\frac{n}{2} \operatorname{trace}\left(K \pi_G(\Gamma) \right) \right], \quad (3)$$

where $\Gamma = \frac{1}{n} \sum_{i=1}^n (\boldsymbol{x}^i)(\boldsymbol{x}^i)^T$ is the sample covariance matrix, $\pi_G(\cdot)$ denotes the projection operator associated with the topology of G:

$$\pi_G(A) = \begin{cases} a_{ij} & \text{if } (i,j) \in E, \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

and K belongs to the convex cone $\mathcal{P}_G$ of symmetric definite matrix satisfying the linear constraints (2). Expression (3) shows that the graphical Gaussian model describes an exponential family with canonical parameter $K \in \mathcal{P}_G$ and sufficient statistic with the incomplete sample covariance $\pi_G(\Gamma)$. As a consequence, the maximum likelihood estimator (MLE) $\widehat{K}$ of the precision matrix K is well defined and exists with probability one for large enough sample size n. Numerical iterative methods, such as Iterative Proportional Scaling [?, p. 134] have been derived to evaluate this MLE. More recently, covariance selection methods, with ℓ_1 penalization to enforce sparsity, have received a great attention [?, ?, ?]. These last methods aim at estimating jointly the entries of the covariance (or precision) matrix and the conditional independence structure. However, even when this structure is known, it is not possible to derive a tractable analytic expression of the MLE in the general case.

3.2. Decomposable Graphical Gaussian Models

First, we recall that the cliques of a graph correspond to the maximal complete subgraphs, i.e. the fully connected subgraphs that are not included in another fully connected subgraphs. A graph is said to be decomposable if it admits a perfect ordering of its cliques $(C_1, \ldots, C_k)$: for $i = 2, \ldots, n$, it exists $l < i$ such that

$$S_i = C_i \bigcap \left[\bigcup_{j=1}^{i-1} C_j \right] \subseteq C_l.$$

The sets S_i are called the separators. In this case, the likelihood of the graphical Gaussian model specified by G can be factorized

$$p(\boldsymbol{x}; K) = \frac{\prod_{i=1}^k p(\boldsymbol{x}_{C_i}; \overline{K}_{C_i})}{\prod_{i=2}^k p(\boldsymbol{x}_{S_i}; \overline{K}_{S_i})}, \quad (5)$$

where $\boldsymbol{x}_A$ denotes the subvector of $\boldsymbol{x}$ corresponding to the elements indexed by $A \subset V$, while $\overline{K}_A$ stands for the inverse of the covariance submatrix Σ_A corresponding to the entries of Σ indexed by $A \times A \subset V \times V$. Note that a necessary and sufficient condition for decomposability is that the graph does not contain a cycle of length greater than or equal to four as an induced subgraph (decomposable graphs are also called triangulated or chordal graphs).

This hypothesis is in agreement with the concepts of hierarchy, modularity and "nearly-decomposability" (strong interactions between the elements of a subsystem, here the elements of a clique, and weaker interactions among the subsystems, here conditional independence of the cliques given a set of separators) introduced by Simon [?] to characterize complex systems.

A consequence of property (5) is that the MLE of K admits the following closed form expression

$$\widehat{K}_{\text{MLE}} = \sum_{i=1}^k \left[\Gamma_{C_i}^{-1} \right]^0 - \sum_{i=2}^k \left[\Gamma_{S_i}^{-1} \right]^0, \quad (6)$$

where Γ is the sample covariance matrix, and $[M_A]^0$, for $A \in V$, denotes the completion on V of the submatrix M_A with zero values. Equation (6) ensures that the MLE exists with probability one if and only if the sample size n is greater than or equal to the cardinal of the clique with maximal size $n_C = \max_{1 \leq i \leq k} |C_i|$.

Furthermore, based on the MLE expression, the minimal variance unbiased estimator (MVUE) has been recently derived in [?]

$$\widehat{K}_{\text{MVUE}} = \sum_{i=1}^k \frac{n - |C_i| - 1}{n} \left[\Gamma_{C_i}^{-1} \right]^0 - \frac{n - |S_i| - 1}{n} \sum_{i=2}^k \left[\Gamma_{S_i}^{-1} \right]^0.$$

Henceforth, we will focus on the estimation of the partial correlation matrix (1). Partial correlations present the advantage to be invariant from some scale changes that occurs in fMRI experiments. By functional invariance principle, the MLE of P denoted as $\widehat{P}_{\text{MLE}}$ is obtained directly by plugging $\widehat{K}_{\text{MLE}}$ in (1).

We will also consider the plug-in MVUE denoted as $\widehat{P}_{\text{PMVUE}}$ derived by plugging $\widehat{K}_{\text{MVUE}}$ in (1). Note that this estimator of P may be biased, even if it is based on the MVUE of K.

4. SIMULATION RESULTS

4.1. Data generation

A surrogate data generation approach has been adopted to allow for realistic comparisons between the aforementioned estimators. First a real fMRI dataset is considered as a basis for computing a realistic sparse precision matrix K, hence a graph of dependencies. In order to achieve this step, the fMRI time series are first gathered into ninety ($p = 90$) distinct components according to spatial considerations. The sample correlation matrix is estimated on wavelet decomposition on the frequency band 0.02-0.04 Hz (see [?] for details).

The resulting observed sample correlation matrix is then thresholded (by keeping the highest ones among the absolute pairwise correlations) such that the precision matrix K only contains sparse nonzero entries. Among the $\frac{p(p-1)}{2} = 4005$ available off-diagonal entries of the precision matrix, only 700 have been retained.

In the next step the artificial data generation is achieved by controlling three crucial characteristics of the $p = 90$ times series, namely the correlation matrix, the Hurst exponents, and the spectrum of the series. An R[1] [?] implementation of the algorithm described in [?] by the authors allowed the generation of series of arbitrary length, from 256 to 4096 points. The correlation matrix of the artificial multivariate series is set to the one extracted from real data. Hurst exponents of the p series were set to 0.7. The rationale behind this generation procedure is that clear long range memory effects (Hurst exponents strictly greater than 0.5) have been observed in real datasets, which results in biased correlation estimations mainly influenced by low-frequency components [?].

In order to alleviate long range dependence, a wavelet transform is used to decompose the signal into successive scale-specific components. In the sequel of the paper, only coefficients of the wavelet transform at scale three are retained. Therefore the resulting time series used for the estimation are composed of various length between 32 and 512 points.

4.2. Graph triangulation

The realistic precision matrix used in the simulation does not yield a decomposable graph. In this case, it is quite natural to embed it in a decomposable graph by adding some edges. The resulting graph is a so-called triangulation of the input graph: the induced cycles whose length are greater than 4 are triangulated by adding new edges. Finding the minimal width triangulation, i.e. that minimizes the number of edges to be added, is NP-hard. However suboptimal solution can be derived efficiently by node elimination heuristics. In this simulation, a minimum degree heuristic is used for the triangulation [?]. It results in a decomposable graph derived by adding 141 edges to the original graph. The triangulation and factorization tasks have been

---

[1]Parts of the procedure are made available in the R package called *brainwaver*

performed by using the java JPSGCS package [?]. The sparsity patterns of the corresponding precision matrices are depicted in fig. 1. It is important to note that even if a few conditional dependences are artificially added in the model, the estimates of their magnitude should be quite small.
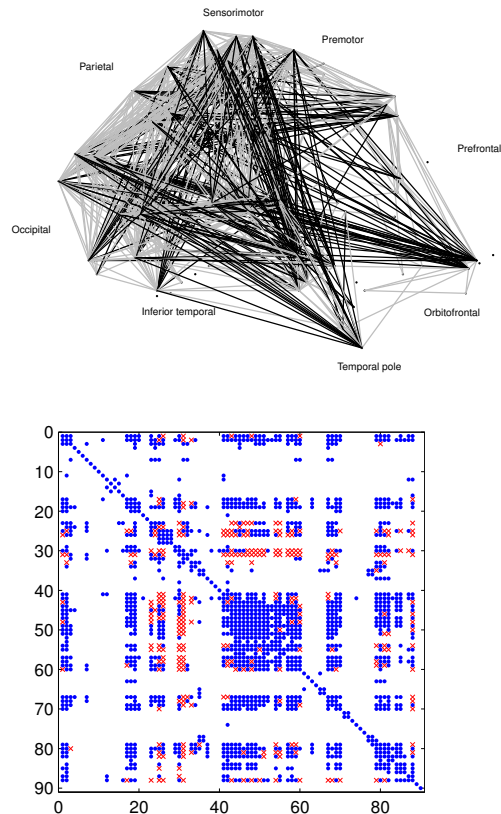

Fig. 1. Top: Anatomical representation of the 700 selected connections where short, < 7.5cm, (resp. long) euclidean distances are grey (resp. black). Bottom: Pattern of the nonzero entries in the precision matrix. Marker •: before triangulation, marker ×: additional edges due to the triangulation.

In this simulation, the maximal clique size of the decomposable graph is $n_c = 28$. Therefore, the MLE derived from the decomposable graphical Gaussian model is defined with probability one when the sample size n is greater than or equal to 28.

4.3. Estimation performance

In order to appreciate the performance of the considered partial correlation estimators, this section studies their mean square errors (MSEs). Several experiments have been conducted corresponding to different sizes of the wavelet coefficient time series, namely $n = 32, 64, 128, 256$ or 512. For each experiment, the number of Monte-Carlo runs is 1000.

In the simulations, the MSE are computed as

$$\text{MSE}\left(\widehat{P}\right) = \mathsf{E}\left[||\widehat{P} - P||^2\right], \qquad (7)$$
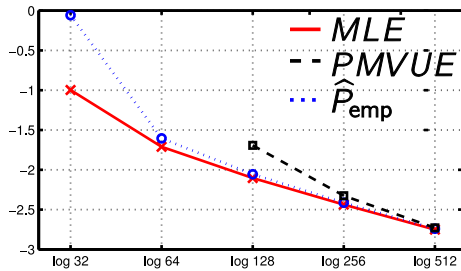
where $||A|| = \sqrt{\text{trace}\left(AA^T\right)}$ is the classical Frobenius norm. It is important to note that, by construction, the MSEs of the graphical

estimators $\widehat{P}_{\text{MLE}}$ and $\widehat{P}_{\text{PMVUE}}$ equal zero for all the entries indexed by $(i, j) \notin E$. This is not the case for the regression estimator $\widehat{P}_{\text{emp}}$, for which there is no prior on the covariance structure of the model. Thus, to perform a fair comparison, only the errors corresponding to the non fixed to zero entries in the precision matrix K are taken into account. It leads to consider $\text{MSE}\left(\pi_G(\widehat{P})\right)$. Finally the MSEs are normalized by dividing by the number of nonzero entries in the considered matrix. Thus, it corresponds to the average of the MSEs associated with each nonzero entry of the partial correlation estimator.

Fig. 2(a) displays the decimal logarithm of the normalized MSEs of $\pi_G(\widehat{P})$ as a function of the decimal logarithm of the sample size. It shows that all the estimators have similar performance for large enough samples. However, for small samples, the MLE appears well behaved and clearly outperforms the regression estimator when it exists, i.e. when $n \geq 90$. More surprisingly, the MLE appears to be also more accurate than the one derived from the MVUE of the precision matrix.

Finally, fig. 2(b) displays the normalized MSEs associated with the set of edges T added during the triangulation step. It is interesting to note that the estimators behave quite similarly than on the whole graph G. This result also illustrates that although the constraint of equality to zero has been relaxed, the maximum likelihood estimates of the partial correlation associated with these edges are very close to zero.

(a) Average MSEs over all the decomposable graph edges
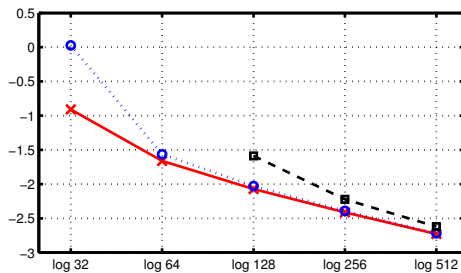
(b) Average MSEs over the triangulated edges

Fig. 2. $\log_{10}$ normalized MSEs of the partial correlation estimators vs the $\log_{10}$ of the sample size. Red solid line: MLE, blue dotted line: PMVUE, black dashed line: $\widehat{P}_{\text{emp}}$ Estimator.

fMRI data exhibiting long range dependences, such graphs are constructed on wavelet transformed series. Although extracted graphs from real data sets do usually not satisfy the decomposability property, it may be forced by triangulation. It is shown that ML estimation of partial correlation on the forced decomposable graph exhibit good performances, even for small sample size. A future work will focus on model selection for brain connectivity networks, as suggested by Dempster in [**?**], with this new estimation scheme.

5. CONCLUSION

In this paper, we illustrated the potential of ML approaches for estimating partial correlation on decomposable graphs. In the case of