CONFIDENTIAL. Limited circulation. For review only.

Sparse plus low-rank autoregressive identification in neuroimaging time series*

Raphaël Liégeois[1], Bamdev Mishra[1], Mattia Zorzi[2], and Rodolphe Sepulchre[1,3]

***Abstract*— This paper considers the problem of identifying multivariate autoregressive (AR) sparse plus low-rank graphical models. Based on the corresponding problem formulation recently presented, we use the alternating direction method of multipliers (ADMM) to efficiently solve it and scale it to sizes encountered in neuroimaging applications. We apply this decomposition on synthetic and real neuroimaging datasets with a specific focus on the information encoded in the low-rank structure of our model. In particular, we illustrate that this information captures the *spatio-temporal structure* of the original data, generalizing classical component analysis approaches.**

I. Introduction

Identifying the links between the variables in multivariate datasets is a fundamental and recurrent problem in many engineering applications. To this end, the use of graphs especially in neuroimaging applications has become very popular because they allow to study and represent the interactions between variables in a concise manner [1].

A particular class of graphs, called *graphical models*, encodes information about dependence between the variables conditioned on all the other variables, or *conditional dependence* [2]. For *static* models this information is contained in the inverse of the covariance matrix also known as the *precision* matrix, and can be estimated by solving the covariance selection problem [3]. Additionally, sparsity and/or low-rank structural constraints can be imposed to the precision matrix estimation. The sparsity constraint results from the parsimony principle in model fitting, i.e., one assumes few direct interactions between the variables, and is obtained through $l1$-norm regularizers [4]. The low-rank structure, induced through nuclear-norm regularizers, models the presence of *latent variables* that are not observed but generate a common behavior in all the observed variables [5]. The low-rank modeling is inspired from what is done in classical component analysis techniques, and leads to models that are simpler and more interpretable [6], [7]. An example of graphical model is given in Fig. 1

*This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Program, initiated by the Science Policy Office of the Belgian State. The scientific responsibility rests with its authors. BM is supported as an FRS-FNRS research fellow (Belgian Fund for Scientific Research).

[1]RL, BM and RS are with the Department of Electrical Engineering and Computer Science, University of Liège, Liège, Belgium. {R.Liegeois,B.Mishra}@ulg.ac.be

[2]MZ is with the Department of Information Engineering, University of Padova, 35131 Padova, Italy. ZorziMat@dei.unipd.it

[3]RS is with the Department of Engineering, Trumpington Street, University of Cambridge, Cambridge CB2 1PZ, United Kingdom. R.Sepulchre@eng.cam.ac.uk

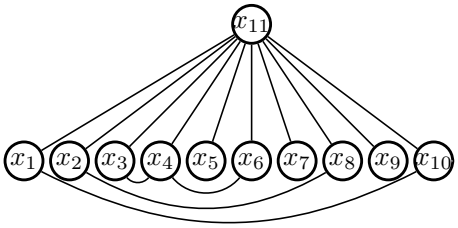

Fig. 1. Ten observed variables $x_i, i \in [1 \dots 10]$ with few interactions among them (sparsity) and one latent variable x_{11} (low-rank structure).

In *dynamical* models, the additional information of the ordering of the data is taken into account and datasets are seen as time series. A widely used class of models encoding this information are autoregressive (AR) models which are characterized by their power spectral density, the dynamic equivalent of the covariance matrix [8]. As in the static case, it has been shown that a zero in the inverse power spectral density corresponds to conditional independence between two variables [9]. In the dynamic case the (inverse) power spectral density is encoded in a block Toeplitz matrix. Because of this particular structure the classical $l1$-norm can not be used to induce sparsity in the inverse power spectral density. This problem is solved by introducing an alternate regularization proposed in [10]. Finally, [11] presents a unifying framework allowing sparse plus low-rank identification of inverse power spectral densities in multivariate time series.

In this paper we adapt the problem formulated in [11] to the alternating direction method of multipliers (ADMM) framework of [12] in order to scale it with larger datasets for which the CVX Matlab toolbox of [13] is computationally expensive. In particular, we exploit separability of constraints of the ADMM framework to decouple the sparsity and the low-rank constraints. The first update is a projective gradient update similar to the one proposed in [10] and the second update is a well known projection onto the cone of positive semidefinite matrices. In the numerical examples we show the efficacy of our proposed algorithm on a real neuroimaging dataset. We also provide further insight into the information encoded in the low-rank structure of our model by applying the proposed algorithm to datasets with different spatio-temporal structures, which is shown to be at least partially recovered in the latent variables.

The paper is organized as follows. We present the optimization problem leading to this sparse plus low-rank decomposition in Section II and we explain how we use ADMM to efficiently solve it in Section III. We then show the results of our approach on synthetic and real data in Section IV and conclude.

Preprint submitted to 54th IEEE Conference on Decision and Control.
Received March 24, 2015.

II. Problem formulation

We first introduce some basic notions, explain the motivation of the sparse plus low-rank (S+L) graphical models and then formally deduce the corresponding optimization problem. Finally, we define *latent components* that we use in the numerical examples in order to characterize information encoded in the low-rank part of this decomposition.

Consider a q-dimensional autoregressive (AR) gaussian process $x = [x_1(t) \dots x_q(t)]^T$ of order p

$$x(t) = \sum_{i=1}^{p} A_i x(t-i) + \epsilon(t),$$

where $x(t) \in \mathbb{R}^q$, $A_i \in \mathbb{R}^{q\times q}$, $i = 1 \dots p$ and $\epsilon(t)$ is white gaussian noise with covariance matrix Σ. x is completely characterized by its spectral density $\Phi(e^{j\theta})$ which encodes information about dependence relations between the q variables [8]. On the contrary the *inverse* power spectral density $\Phi_x(e^{j\theta})^{-1}$ encodes *conditional* dependence relations between variables [9], [14]. That is, two variables x_k and x_l are independent, conditionally on the other $q-2$ variables of x over $\{t \in \mathbb{Z}\}$, if and only if

$$(\Phi(e^{j\theta})^{-1})_{kl} = 0 \quad \forall \theta \in [0, 2\pi], \tag{1}$$

The nodes of the corresponding graphical model are the q variables of x and there is no edge between the two nodes x_k and x_l if and only if (1) is satisfied.

A. S+L graphical models

Assume that $x(t) = [(x^m(t))^T \ (x^l(t))^T]^T$ where $x^m(t) = [x_1(t) \dots x_m(t)]^T \in \mathbb{R}^m$ contains *manifest* variables, that is variables accessible to observations, and $x^l(t) = [x_{m+1}(t) \dots x_{m+l}(t)]^T \in \mathbb{R}^l$ contains *latent* variables, not accessible to observations. The power spectrum of x can be expressed using the following block decomposition

$$\Phi_x = \begin{bmatrix} \Phi_m & \Phi_{ml}^* \\ \Phi_{ml} & \Phi_l \end{bmatrix}, \quad \Phi_x^{-1} = \begin{bmatrix} \Upsilon_m & \Upsilon_{ml}^* \\ \Upsilon_{ml} & \Upsilon_l \end{bmatrix}, \tag{2}$$

where the * denotes the conjugate transpose operation.

In order to better characterize the conditional dependence relations between the manifest variables, from (2) we obtain the following decomposition of Φ_m^{-1} using the Schur complement [15]

$$\Phi_m^{-1} = \Upsilon_m - \Upsilon_{ml}^* \Upsilon_l \Upsilon_{ml} \ . \tag{3}$$

The main modeling assumption here is that $l \ll m$ and that the conditional dependencies relations among the m manifest variables encoded in Φ_m^{-1} can be largely explained through few latent variables. The corresponding graphical model has few edges between the manifest variables and few latent nodes, as in Fig. 1. This leads to a S+L structure for Φ_m^{-1} following (3): $\Phi_m^{-1} = \Sigma - \Lambda$, where $\Sigma = \Upsilon_m$ is sparse because it encodes conditional dependence relations between the manifest variables and $\Lambda = \Upsilon_{ml}^* \Upsilon_l \Upsilon_{ml}$ is low rank. Since x is an AR process of order p, we can assume that Σ and Λ belong to the family of matrix pseudo-polynomials

$$\mathcal{Q}_{m,p} = \{ \sum_{j=-p}^{p} e^{-ij\theta} R_j : R_j = R_{-j}^T \in \mathbb{R}^{m\times m} \}.$$

Following [11] we further rewrite Σ and Λ as

$$\begin{aligned} \Sigma - \Lambda &= \Delta X \Delta^*, \\ \Lambda &= \Delta L \Delta^*, \end{aligned} \tag{4}$$

where Δ is a shift operator $\Delta(e^{i\theta}) := [I \quad e^{i\theta} I \ \dots \ e^{ip\theta} I]$ and X and L are now matrices belonging to $Q_{m(p+1)}$ which is the set of symmetric matrices of size $m(p+1) \times m(p+1)$.

Finally, $M_{m,p}$ is the vector space of matrices $W := [W_0 \ W_1 \dots W_p]$ with $W_0 \in Q_m$ and $W_1 \dots W_p \in \mathbb{R}^{m\times m}$. The linear mapping $\mathcal{T} : M_{m,p} \to Q_{m(p+1)}$ outputs a symmetric block Toeplitz matrix from the blocks of W as

$$\mathcal{T}(W) = \begin{pmatrix} W_0 & W_1 & \cdots & W_p \\ W_1^T & W_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & W_1 \\ W_n^T & \cdots & W_1^T & W_0 \end{pmatrix}.$$

The adjoint operator of $\mathcal{T}$ is the linear mapping $\mathcal{D} : Q_{m(p+1)} \to M_{m,p}$ defined for a matrix $X \in Q_{m(p+1)}$ partitioned in square blocks of size $m \times m$ as

$$X = \begin{pmatrix} X_{00} & X_{01} & \cdots & X_{0p} \\ X_{00}^T & X_{11} & \cdots & X_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{0p}^T & X_{1p}^T & \cdots & X_{pp} \end{pmatrix}. \tag{5}$$

Following this partition, $W = \mathcal{D}(X) \in M_{m,p}$ is given by

$$\begin{cases} W_0(X) &= \sum_{h=0}^{p} X_{hh}, \\ W_j(X) &= 2\sum_{h=0}^{p-j} X_{h,h+j} \quad \forall j = 1 \dots p. \end{cases} \tag{6}$$

B. S+L identification problem formulation

Assume that we have a finite length realization $x^m(1) \dots x^m(N)$ of the manifest process x^m. It should be emphasized that no data is available regarding the latent process x^l, and its dimension is not even known. Our goal is to recover the S+L model defined in the previous section which best explains the collected data. Therefore, one estimate of X and L (hence of Σ and Λ) is given by solving the regularized maximum entropy problem [16], [17] for which the primal is

$$\begin{aligned} \min_{X,L\in Q_{m(p+1)}} \quad & -\log\det X_{00} + \langle C, X\rangle \\ & + \lambda\gamma h(\mathcal{D}(X+L)) + \lambda \ \mathrm{trace}(L) \\ \text{subject to} \quad & X \succeq 0, X_{00} \succ 0, L \succeq 0, \end{aligned} \tag{7}$$

where

- $\gamma > 0$ and $\lambda > 0$ are weighting parameters leading to a sparse Σ and a low-rank Λ,
- $C = \mathcal{T}(\hat{R})$ where $\hat{R} \in M_{m,p}$ are the $p+1$ first sample covariance lags $\hat{R}_k$ [8],
- $\langle ., . \rangle$ is the inner product associated with $\mathbb{R}^{m\times m}$,

- h is the following function chosen to encourage a structured sparse solution $\mathcal{D}(X+L)$

$$h(Y) = \sum_{j>i} \max\{|(Y_0)_{ij}|, \max_{k=1,\dots,p} |(Y_k)_{ij}|, \max_{k=1,\dots,p} |(Y_k)_{ji}|\},$$

which is convex but non smooth. This limitation was contoured in similar works [10] by solving the corresponding dual problem. In the present case, it can be shown based on [10] and [11] that the dual of (7) is

$$\begin{array}{ll} \max\limits_{Z\in M_{m,p}} & \Psi(C+\mathcal{T}(Z)) \\ \text{subject to} & \sum_{k=0}^{p}(|(Z_k)_{ij}|+|(Z_k)_{ji}|) \le \gamma\lambda,\ i\neq j \\ & \textbf{diag}(Z_k)=0,\ k=0,...,p \\ & \lambda I+\mathcal{T}(Z)\succeq 0, \end{array} \tag{8}$$

where $\Psi : Q_{m(p+1)} \to \mathbb{R}$ is defined as

$$\Psi(V) = -\log\det(V_{00} - V_{1:p,0}^T V_{1:p,1:p}^{-1} V_{1:p,0}) - m$$

and $V \in Q_{m(p+1)}$ is partitioned as in (5).

C. Definition of latent components

The optimal primal variables (X^{opt},L^{opt}) are recovered from the optimal solution Z^{opt} following [11] from which Σ^{opt} and Λ^{opt} are computed by (4). Λ^{opt} is a square matrix function of size m, defined over the unit circle and for which we consider the pointwise singular value decomposition since Λ^{opt} is low-rank for all $\theta \in [0, 2\pi]$

$$\Lambda^{opt}_{m,m}(\theta) = \Gamma_{m,l}(\theta)\Omega_{l,l}(\theta)\Gamma^*_{m,l}(\theta). \tag{9}$$

The i-th column of $\Gamma_{m,l}(\theta)$ contains the strength of the conditional dependence relation between the i-th unobserved latent variable and each of the m manifest variables. Following component analysis nomenclature [6] we call the i-th column of $\Gamma_{m,l}(\theta)$ the i-th *latent component* and denote it $\gamma_{i,.}(\theta)$. In other words, $\gamma_{i,j}(\theta)$ with $i \in [1\dots l]$ and $j \in [1\dots m]$ represents the weight of the conditional dependence between the latent variable $x_{m+i}(t)$ and the manifest variable $x_j(t)$ in the expression of $x_j(t)$. In the static case, $\gamma_{i,.}(\theta)$ reduces to a constant vector because $\Delta = I$ in (4) and Λ does not depend on θ. On the contrary, in the dynamic case each latent component is a function of $\theta \in [0, 2\pi]$.

III. Alternating direction method of multipliers

We use the alternating direction method of multipliers (ADMM) of [12] to solve (8). This choice is motivated by the fact that this algorithm both inherits the strong convergence properties of the method of multipliers and exploits decomposability of the dual problem formulation leading to efficient partial updates of the variables. We show how we rewrite (8) in the ADMM format by separating the sparse and low-rank constraints, then explain how we choose an adequate stopping criterion and recover the primal variables.

In order to decouple the constraints related to sparsity and low-rank we introduce the new variable $Y \in Q_{m(p+1)}$ and reformulate (8) as

$$\begin{array}{lll} \min\limits_{Z\in M_{m,p}} & \Psi(C+\mathcal{T}(Z)) & \\ \text{subject to} & A(Z)\le b, & (C_1) \\ & Y=\mathcal{T}(Z)+\lambda I, & \\ & Y\succeq 0 & (C_2), \end{array} \tag{10}$$

where the first constraint (C_1) gathers the first two constraints on Z of (8) and A and b are defined accordingly; and (C_2) is the last constraint of (8) imposing positive semidefiniteness of the new variable Y. Using the *augmented Lagrangian* formulation, we introduce L_ρ defined by

$$L_\rho(Z,Y,M) = \Psi(C+T(Z)) - \langle M, Y-\mathcal{T}(Z)-\lambda I\rangle + \frac{\rho}{2}\|Y-\mathcal{T}(Z)-\lambda I\|_F^2,$$

where $\|\cdot\|_F$ is the Frobenius norm. Subsequently, the ADMM updates are

$$\begin{cases} Z^{k+1} = \min\limits_{Z\,\in\,C_1} L_\rho(Z,Y^k,M^k), & \text{(S1)} \\ Y^{k+1} = \min\limits_{Y\,\in\,C_2} L_\rho(Z^{k+1},Y,M^k), & \text{(S2)} \\ M^{k+1} = M^k - \rho(Y^{k+1}-\mathcal{T}(Z^{k+1})-\lambda I). & \text{(S3)} \end{cases}$$

It should be noted that (S1) has no closed form solution and corresponds to the sparsity set of constraints of [10]. We approximate the solution by a projective gradient step as in [10]. Following this approach,

$$Z^{k+1} = \Pi_{C_1}(Z^k - t_k\nabla_Z L_\rho(Z_k,Y_k,M_k)),$$

where

- $\nabla_Z L_\rho(Z_k,Y_k,M_k) = \nabla\Psi(Z^k) + \mathcal{D}(M^k) + \rho\mathcal{D}(\mathcal{T}(Z^k)+\lambda I - Y^k)$,
- t_k is found from the Armijo conditions,
- Π_{C_1} is the projection onto C_1 which reduces to a projection onto the l_1-norm ball [10], [18].

The optimization problem (S2) has a closed form solution and is computed as

$$Y^{k+1} = \Pi_{C_2}(\frac{1}{\rho}M^k - \mathcal{T}(Z^{k+1}) - \lambda I),$$

where Π_{C_2} is the projection onto the cone of symmetric positive semidefinite matrices of size $m(p+1)\times m(p+1)$, which is done by selecting the eigenvectors corresponding to *positive* eigenvalues. This leads to the final updates of the ADMM algorithm.

ADMM for sparse plus low-rank inverse power spectral density estimation. Initialize Z_0, Y_0, M_0; set $\rho > 0$; and successively update variables as follows:

$$\begin{array}{ll} Z^{k+1} = & \Pi_{C_1}(Z^k - t_k\nabla_Z L_\rho(Z_k,Y_k,M_k)), \\ Y^{k+1} = & \Pi_{C_2}(\frac{1}{\rho}M^k - \mathcal{T}(Z^{k+1}) - \lambda I), \\ M^{k+1} = & M^k - \rho(Y^{k+1} - \mathcal{T}(Z^{k+1}) - \lambda I). \end{array} \tag{11}$$

Following [12], a stopping criterion for (11) is based on the primal and dual *residuals* r and s that respectively measure

satisfaction of the equality constraint of (10) and the distance between two successive iterates of the additional variable Y. r and s should satisfy $\|r\|_F \leq \epsilon^{pri}$ and $\|s\|_F \leq \epsilon^{dual}$ where ϵ^{pri} and ϵ^{dual} are defined as

$$\begin{aligned}\epsilon^{pri} &= m(p+1)\epsilon^{abs} \\ &\quad +\epsilon^{rel}\max\{\lambda\sqrt{m(p+1)}, \|T(Z^k)\|_F, \|Y^k\|_F\}, \\ \epsilon^{dual} &= m\sqrt{p+1}\epsilon^{abs} + \epsilon^{rel}\|\mathcal{D}(M^k)\|_F.\end{aligned}$$

Here ϵ^{abs} and ϵ^{rel} are the predefined absolute and relative tolerances for the problem.

A variation is obtained when ρ is multiplied by a factor of $\tau > 1$ at each iteration up to a maximum value ρ_{max} starting from a value ρ_0 depending on the application.

Convergence analysis of the ADMM algorithm follows from [12, Section 3.2][19]. The computational cost per iteration of the updates (11) depends on the projections onto C_1 and C_2, the gradient evaluation of L_ρ, and the linear mappings $\mathcal{T}$ and $\mathcal{D}$ leading to a final complexity $\mathcal{O}(m^3(p+1)^3)$.

IV. Numerical Examples

In this section we apply the proposed ADMM algorithm to solve the sparse plus low-rank decomposition on synthetic and real datasets and explore the type of information encoded in the identified latent *components*. The Matlab code for the algorithm is available from the webpage `http://www.montefiore.ulg.ac.be/~rliegeois/`

A. Application on linear synthetic data

This synthetic dataset consists of time series corresponding to a first order AR model (dynamic model, $p = 1$) with the interaction graph presented in Fig. 1. The interaction graphs of the manifest variables (support of Σ) identified for different values of λ and $\lambda\gamma$ are represented in Fig. 2 as well as l, the number of latent components (rank of Λ) that were identified. In order to discriminate between models we compute a score function f, defined in [10], taking into account fitting to the data and complexity of the model.

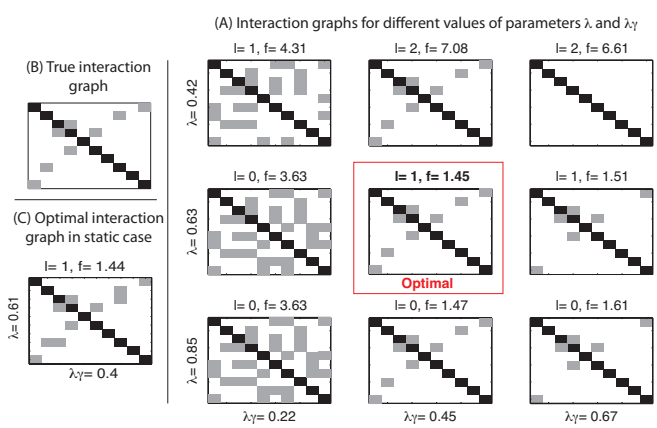

Fig. 2. (A) Interaction graphs of estimated models for different values of λ and $\lambda\gamma$ using a first order AR model ($p = 1$). (B) True interaction graph. (C) Optimal interaction graph of estimated model obtained in the static case ($p = 0$).

As expected, higher values of λ promote models with less latent components, and higher values of $\lambda\gamma$ favor models with few interactions between the manifest variables. The model with the best (lowest) score function recovers the true interaction graph with the correct number of latent components (Fig. 2A). The optimal interaction graph in the static case (Fig. 2C), on the contrary, does not recover exactly the true interaction graph.

The stopping criterion based on the primal and dual residuals is illustrated in Fig. 3 using this dataset. The algorithm stops when $\|r\|_2 \leq \epsilon^{pri}$ and $\|s\|_2 \leq \epsilon^{dual}$ are both satisfied.

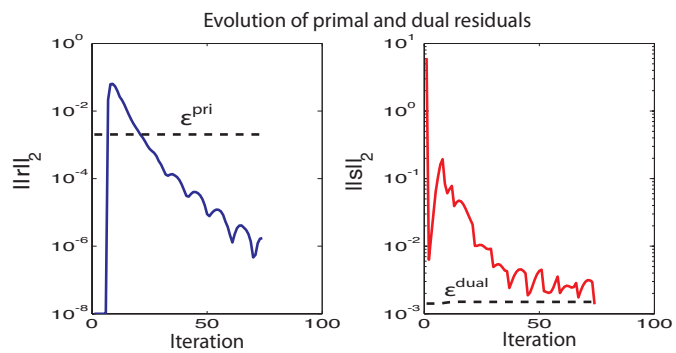

Fig. 3. Typical convergence scheme of the ADMM algorithm obtained with the linear synthetic dataset with the parameter values $\rho_0 = 1$, $\tau = 1.1$, $\rho_{max} = 1000$, $\epsilon^{abs} = 10^{-5}$, and $\epsilon^{del} = 10^{-4}$.

B. Application on non-linear synthetic data

The interest in considering a non-linear generative model is that it can produce endogenous sustained oscillations in networks similarly to what is observed in neuroimaging data such as fMRI time series. A popular non-linear model is the Hopfield model widely used in neural networks [20]:

$$\dot{x_i} = -x_i + \text{sat}(\sum_{j\neq i} G_{ij}x_j) + \epsilon,$$

where x_i denotes the level of activity of the variable i, G_{ij} is the strength of the connection from x_j to x_i, and sat is a sigmoïdal saturation function. Fig. 4 shows how we generate oscillations in two clusters. In the first case (Fig. 4B) the oscillations are coupled, leading to dephased oscillations of the same frequency whereas in the second case the clusters are decoupled (Fig. 4C), leading to oscillations of different frequencies in the two clusters. We finally generate three different datasets for each configuration by sampling these time series at different frequencies. The first dataset is produced by using the original time series (no sampling), the second and third datasets are obtained by sampling the time series with a period of T1 and T2 to generate synthetic data with higher frequency content.

Fig. 5 shows the static and dynamic latent components as defined in section II-C that are identified in the optimal models from the synchronous oscillations datasets of Fig. 4B. Since the results are very similar within the nodes of each cluster and for clarity purposes we plot only the value of the latent components in node 1 ($\gamma_{i,1}$) and in node 5 ($\gamma_{i,5}$).

Preprint submitted to 54th IEEE Conference on Decision and Control.
Received March 24, 2015.

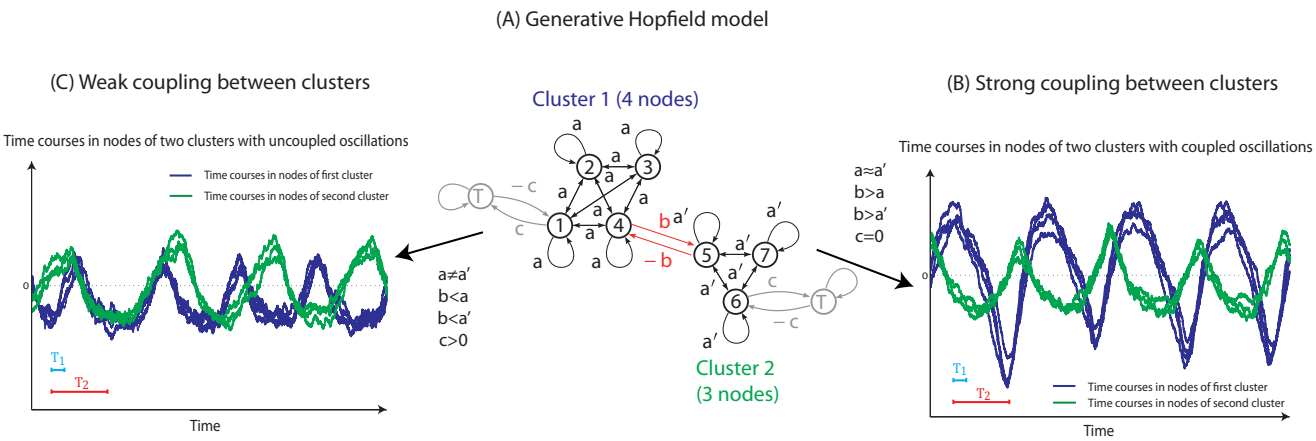

Fig. 4. (A) The generative model contains two clusters and the values of the directed connectivity between the nodes is indicated on the corresponding arrows. These values correspond to the matrix C in (12). (B) Set of parameters leading to coupled oscillations and (C) decoupled oscillations.

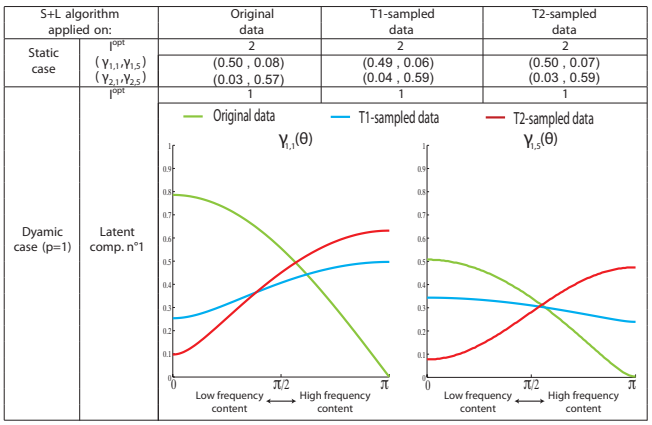

Fig. 5. Latent components identified in the static and dynamic cases from the three datasets generated from coupled oscillators (Fig. 4 B).

In the static case two latent components corresponding to the two clusters of the generative model are identified. There is no significant difference for the three input datasets suggesting that the frequency content of original data is not encoded in the static latent components. In the dynamic case ($p = 1$), the latent components are encoded in $\gamma_{i,j}(\theta)$ and a single latent component is identified in the optimal model. Interestingly, the T2-sampled synthetic data (high-frequency time series) leads to a latent component showing a strong high-frequency content whereas the original dataset leads to a latent component with dominant low-frequency content.

Having the same approach on asynchronous oscillations, we get the results shown in Fig. 6. In the static case we still obtain two latent components recovering the two clusters with no distinction between the three starting datasets. In the dynamic case ($p = 1$), however, the optimal model now has two latent components, each one capturing the oscillations in one of the two clusters. As in the previous case the frequency content of the synthetic data is encoded in the frequency content of the latent component. Identifying two latent components probably comes from the fact that the frequency of oscillation in the two clusters are different,

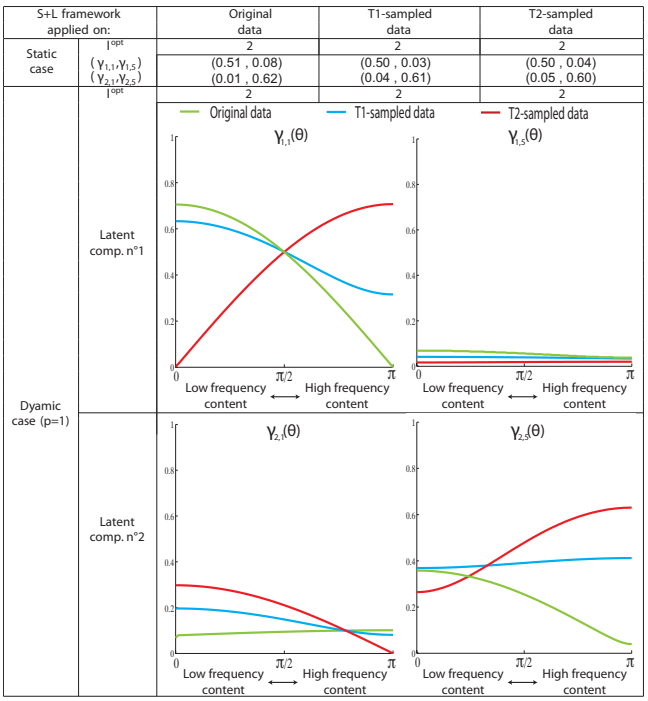

Fig. 6. Latent components identified in the static and dynamic cases from the three datasets generated from decoupled oscillators (Fig. 4C).

suggesting that the dynamic latent component identifies a *spatio-temporal* subspace of variation common to different manifest variables.

C. Application to neuroimaging data

We illustrate the application of our proposed algorithm on real neuroimaging data consisting of functional magnetic resonance imaging time series in 90 brain regions collected on 17 patients during rest [21]. It should be noted that this problem dimension is not tractable with a standard optimization tool such as the CVX toolbox of [13]. The classical approach is to use component analysis to extract neuronal networks. During rest, three networks are robustly

Preprint submitted to 54th IEEE Conference on Decision and Control.
Received March 24, 2015.

identified using component analysis: the visual network (VN), the default mode network (DMN) and the executive control network (ECN) that are represented in Fig 7.

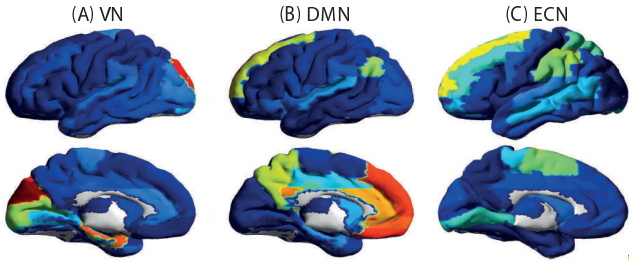

Fig. 7. Three resting state neuronal networks are commonly recovered using component analysis: (A) the visual network, (B) the default mode network, and (C) the executive control network.

By fitting a first order AR model ($p = 1$) to this dataset we identify latent components corresponding to these networks. In Fig. 8 we plot only these components for $\theta = 0$, a low-frequency contribution called $\gamma_0 = \gamma(\theta = 0)$ from the definition of Section II-C, and $\theta = \pi$, a high-frequency contribution called $\gamma_\pi = \gamma(\theta = \pi)$. Indeed, for a first order model the latent components are characterized by these two extreme values γ_0 and γ_π.

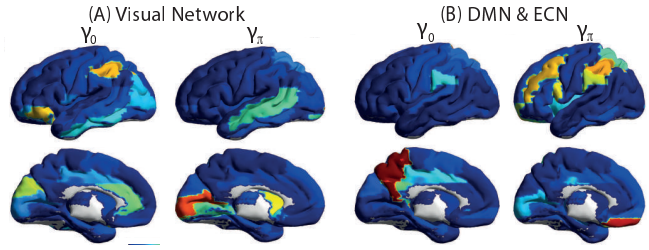

Fig. 8. (A) The visual network is recovered in 14 out of 17 subjects. (B) The DMN and ECN are coupled into a unique latent component in 12 out of 17 subjects.

In 14 subjects out of 17 we observe that VN is recovered in one latent component, with no significant differences between γ_0 and γ_π. On the other hand, in 12 subjects the DMN and ECN networks are gathered in a unique latent variable, the DMN corresponding to γ_0 and ECN to γ_π. This interestingly echoes some recent results suggesting that the DMN and the ECN are both consciousness related processes and are anti-correlated, oscillating at similar frequencies [21].

V. Conclusion

The contribution of our work is twofold. First, we reformulated the sparse plus low-rank autoregressive identification problem into the ADMM framework in order to scale it to larger datasets encountered in neuroimaging applications. Second, we presented a deeper exploration of the type of information encoded in the latent components identified in the low-rank part of the decomposition. Fig. 5 & 6 suggest that this information is richer in dynamic models than in static models in the sense that dynamic latent components recover *spatio-temporal* properties of the original time series such as common spectral content. Applied to a neuroimaging dataset, this interpretation led to a novel characterization of the dynamical interplay between two neuronal networks mediating consciousness, echoing recent experimental results.

As a future research direction we intend to explore whether additional information can be recovered in higher order dynamical latent components ($p > 1$). It would also be interesting to study the mathematical link between the latent components of the proposed framework and the widely used static principal components that also capture a 'common subspace' of variation of many variables from the covariance matrix instead of the precision matrix. Finally, in addition to using separability into sparse and low-rank constraints, it could be beneficial to exploit these structures in the algorithm updates in order to scale it to even larger problems.

References

[1] Bullmore E., Sporns O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* Vol. 10, No.3, pp. 186-98

[2] Koller, D., and Friedman, N. 2009. Probabilistic Graphical Models. *Massachusetts: MIT Press*.

[3] Dempster, A., 1972. Covariance selection. *Biometrics*, Vol. 28, pp. 157-75.

[4] Banerjee, O., El Ghaoui, L., and d'Aspremont, A., 2008. Model Selection Through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data. *J. Mach. Learn. Res.*, Vol. 9, pp. 485-516.

[5] Chandrasekaran, V., Parrilo, P. A., and Willsky, A. S., 2012. Latent Variable Graphical Model Selection via Convex Optimization, *Ann. Stat.*, Vol. 40, No. 4, pp. 1935-67.

[6] Jolliffe, I. T., 2002. Principal Component Analysis. *Sec. ed. (Springer)*.

[7] Hyvärinen, A., Karhunen, J., and Oja, E., 2001. Independent Component Analysis. *John Wiley and Son*.

[8] Stoica, P., and Moses, R. L., 1997. Introduction to spectral analysis. *Prentice Hall, Upper Saddle River, N.J.*

[9] Brillinger, D., 1981. Time Series: Data Analysis and Theory. *New York: Holden-Day.*

[10] Songsiri, J., and Vandenberghe, L., 2010. Topology selection in graphical models of autoregressive processes. *J. Mach. Learn. Res.*, Vol. 11, pp. 2671-2705.

[11] Zorzi, M., and Sepulchre, R., 2015. AR identification of Latent-variable Graphical Models. *Submitted* http://arxiv.org/abs/1405.0027.

[12] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J., 2010. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Found. Trend. Mach. Learn.*, Vol. 3, No. 1, pp. 1-122.

[13] Grant M., Boyd S., and Ye, Y., 2005. CVX: Matlab software for disciplined convex programming. Available from www.stanford.edu/ boyd/cvx/

[14] Dahlhaus, R., 2000. Graphical interaction models for multivariate time series. *Metrika*, Vol. 51, No. 2, pp. 157-72.

[15] Horn, R. A., and Johnson, C. R. (1990). Matrix analysis. *Cambridge University Press*.

[16] Byrnes, C., Gusev, S., and Lindquist, A. 1998. A convex optimization approach to the rational covariance extension problem *SIAM J. Optimiz.*, Vol. 37, No. 1, pp. 211-29, 1998

[17] Cover, T., and Thomas, J. 1991. Information Theory. *New York: Wiley*.

[18] Berg, E., Schmidt, M., Friedlander, M., and Murphy, K., 2008. Group sparsity via linear-time projection. Dept. of Computer Science, Univ. of British Columbia, Vancouver, BC, Canada, Technical Report.

[19] Ouyang, Y., Chen, Y., Lan, G., and Pasiliao, E., 2014. An accelerated linearised alternating direction method of multipliers. *Submitted* http://arxiv.org/pdf/1401.6607v3.pdf.

[20] Hopfield, J. J., 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.*, Vol. 79, No. 8, pp. 2554-58.

[21] Vanhaudenhuyse, A., Demertzi, A., Schabus, M., Noirhomme, Q., Bredart, S., Boly, M., Phillips, C., Soddu, A., Moonen, G., and Laureys, S., 2011. Two distinct neuronal networks mediate the awareness of environment and of self. *J. Cogn. Neurosci.*, Vol. 23, pp. 570-8.

Preprint submitted to 54th IEEE Conference on Decision and Control.
Received March 24, 2015.