Sparse plus low-rank graphical models of time series for functional connectivity in MEG

Nicholas J. Foti
Department of Statistics
University of Washington
nfoti@uw.edu

Rahul Nadkarni
Computer Science and Engineering
University of Washington
rahuln@cs.washington.edu

Adrian KC Lee
Institute for Learning and Brain Sciences
University of Washington
akclee@uw.edu

Emily B. Fox
Department of Statistics
University of Washington
ebfox@uw.edu

ABSTRACT

Inferring graphical models from high dimensional observations has become an important problem in machine learning and statistics because of its importance in a variety of application domains. One such application is inferring functional connectivity between brain regions from neuroimaging data such as magnetoencephalograpy (MEG) recordings that produce signals with good temporal and spatial resolution. Unfortunately, existing techniques to learn graphical models that have been applied to neuroimaging data have assumed the data to be i.i.d. over time, ignoring key temporal dynamics. Additionally, the signals that arise from neuroimaging data do not exist in isolation as the brain is performing many tasks simultaneously so that most existing methods can introduce spurious connections. We address these issues by introducing a method to learn Gaussian graphical models between multiple time series with latent processes. In addition, we allow for heterogeneity between different groups of MEG recordings by using a hierarchical penalty. The proposed methods are formulated as convex optimization problems that we efficiently solve by developing an alternating directions method of multipliers algorithm. We evaluate the proposed model on synthetic data as well as on global stock index returns and a real MEG data set.

Keywords

1. INTRODUCTION

Consider the challenge of inferring networks in the brain from noisy, high-dimensional neuroimaging recordings. In particular, our focus is on deciphering *functional connectivity* networks activated during various auditory attention tasks based on magnetoencephalography (MEG) data. The concept of *functional connectivity* in neuroscience is based on the correlations between signals in two different brain regions when the linear effects of all other regions have been removed [4]. This definition is precisely that of *conditional independence* under a Gaussian likelihood. Such conditional independencies are equivalently captured by zeros in the precision (i.e., inverse covariance) matrix, which encodes the structure in a Gaussian graphical model.

Learning Gaussian graphical models (i.e., sparse precision matrices) from high-dimensional data is a broadly popular tool in machine learning and statistics to discern the underlying conditional independence structure between random variables. Beyond providing a natural definition of functional connectivity in neuroscience [16], inferring such structure has far reaching applications, such as allowing us to quantify risk between financial instruments [2, 5]. However, most existing approaches to learning Gaussian graphical models assume that the observations are independent and identically distributed (i.i.d.) Gaussian random vectors. In many applications of interest, such as stock data or our MEG recordings, the data represent a multivariate time series. Simply treating the observations as i.i.d. across time (cf., [5]) ignores important information contained in the dynamics that we might want to account for when assessing conditional independencies.

Recently, there has been work to try to learn graphs of Gaussian stationary time series [1, 18, 13, 19]. Most of this work builds on an elegant result of Dahlhaus [6], which states that zeros at all frequencies in the *inverse spectral density* matrices characterize conditional independence between time series. This represents a direct analog of the standard sparse precision result for Gaussian i.i.d. random variables. Focusing on likelihood-based approaches to structure learning, recent work has considered transforming the time series into the frequency domain and specifying Bayesian priors on the inverse spectral density matrices [19], or a penalized likelihood approach using a group-lasso penalty [18]. The latter builds on the common ℓ_1-based *graphical lasso (glasso)* approach for Gaussian graphical model structure learning [9], but with a group structure to capture shared zero patterns across frequencies in the spectral domain.

In this paper, motivated by our application, we build on this *graphs of time series* framework to account for latent unobserved processes that, when omitted, may introduce

KDD '16

DOI:

spurious edges between the observed series. Such spurious connections are particularly problematic in our neuroscientific application since neuroimaging data is notoriously noisy making existing methods result in graphs that are too dense to interpret. The rationale behind the presence of such latent processes is manifold. First, the brains of subjects involved in a cognitive experiment are most likely also performing a variety of other tasks not related to the experiment. Such signals will be measured in the MEG data and, if unaccounted for, can introduce spurious connections unrelated to the task. Second, the MEG recording mechanism is subject to what is referred to as *point spread* where due to an ill-posed mapping from sensor space (raw recordings) to source space (projecting to a tessellated grid of the cortex), the effect of a true signal in the brain gets diffused amongst many regions. The result is that disconnected regions can artificially appear to be strongly connected. We believe that both of these effects can be ameliorated through the incorporation of latent processes.

Accounting for latent processes has an elegant solution in the i.i.d. setting through a *sparse plus low-rank* decomposition of the precision matrix for the observed variables [5]. Interesting challenges arise when applying this approach to graphs of time series since the zeros have to be consistent across frequencies. We devise a penalty that encourages the inverse spectral density matrices of the observed process to decompose into sparse and low rank components per frequency, while having shared sparsity structure across frequencies. Unfortunately, existing algorithms to solve sparse plus low-rank models like `logdetPPA` [21] do not scale to the problems we consider due to the size of the problem rapidly increasing with the number of frequencies analyzed. To address this computational issue we develop efficient alternating direction method of multipliers (ADMM) algorithms that can be parallelized over frequencies and allowing us to perform efficient inference for the model. We apply the methods to synthetic observations and demonstrate the utility on two real-world data sets including stock index data where interpreting the learned graphs is simpler and on MEG data collected during an experiment on auditory attention.

2. BACKGROUND

Gaussian Graphical Models.

Let $X \in \mathbb{R}^p$ be a random variable distributed according to a multivariate Gaussian distribution, $X \sim \mathbb{N}(0, \Sigma)$, and let $G = (V, E)$ be a graph with vertices $V = \{1, \ldots, p\}$ and edge set $E \subset \{(i, j) \in V \times V : i \neq j\}$. When $\Sigma^{-1}_{ij} \neq 0$ for all pairs $(i, j) \in E$, we say that X respects the graph G and that X follows the *Gaussian graphical model* specified by the precision matrix Σ^{-1}.

Given a sample covariance estimate $\hat{\Sigma}$ based on a set of observations $X_1, \ldots, X_N$, a common approach to inferring a Gaussian graphical model is the *graphical lasso* given by the convex program:

$$\underset{\Omega \in \mathbb{S}^p_{++}}{\arg\min} - \log\det\Omega + \mathrm{tr}\{\Omega\hat{\Sigma}\} + \lambda \left|\left|\Omega\right|\right|_1, \tag{1}$$

where $\mathbb{S}^p_{++}$ is the cone of $p \times p$ symmetric positive-definite matrices and $\left|\left|\Omega\right|\right|_1 = \sum_{i<j} |\Omega_{ij}|$ is the 1-norm of the matrix and can encourage a sparse solution for Ω to be learned.

Graphical Models of Stationary Time Series.

Let $X_t = (X^1_t, X^2_t, \ldots, X^p_t)^T \in \mathbb{R}^p$ for $t \in \mathbb{Z}$ be a stationary multivariate Gaussian time series so that

$$\mathbb{E}\left[X_t\right] = \mu, \ \forall t \in \mathbb{Z} \tag{2}$$

$$\mathbb{E}\left[X_t, X_{t+h}\right] = \Gamma(h), \ \forall t \in \mathbb{Z}, \tag{3}$$

where $\Gamma(h)$ is the *autocovariance function* of the time series so that for all $h \in \mathbb{Z}$, $\Gamma(h)$ is a $p \times p$ symmetric positive definite matrix. Under the stationarity assumptions we have that $\sum_{-\infty}^{\infty} ||\Gamma(h)||_2 < \infty$ where $||\cdot||_2$ is the spectral norm. The *spectral density matrix* of the time series is given by the discrete Fourier transform of the autocovariance function

$$S(\omega) = \frac{1}{2\pi} \sum_{\infty}^{\infty} \Gamma(h) e^{-ih\omega} \tag{4}$$

for $\omega \in [0, 2\pi]$ and $S(\omega) \in \mathbb{C}^{p \times p}$ is a $p \times p$ Hermitian positive definite matrix.

The idea of a Gaussian graphical model is naturally extended to Gaussian stationary time series: Instead of conditional independencies being encoded as zeros in the inverse covariance (precision), such statements are encoded as zeros in the inverse spectral density matrix, $S(\omega)^{-1}$ [6]. Specifically, if $S(\omega)^{-1}_{ij} = 0$ for all $\omega \in [0, 2\pi]$ then the components X^i and X^j are independent conditioned on the entire trajectories of the other series.

One approach to learning such structure is through a penalized likelihood approach. In the frequency domain, the likelihood of the time series can be efficiently computed using the *Whittle approximation* to the likelihood:

$$\begin{aligned} \log\, p(X_1, \ldots, X_T) \approx \\ - \frac{1}{2} \sum_{k=0}^{T-1} \left[\log\det S(\omega_k) + \mathrm{tr}[S(\omega_k)^{-1} \hat{S}(\omega_k)]\right] \\ - \frac{Tp}{2} \log 2\pi. \end{aligned} \tag{5}$$

Here, $\hat{S}(\omega_k)$ is the spectral density estimate at frequency $\omega_k = 2\pi k/T$. A sample-based estimate of the spectral density matrix is given by the *periodogram*

$$I(\omega_k) = \frac{1}{2\pi} d(k) d(k)^*, \tag{6}$$

where $d(k)$, $k = 0, ..., T-1$, is the discrete Fourier transform of x_t, a realization of the Gaussian time series X_t, $t \in 0, ..., T$. That is,

$$d(k) = \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} x_t e^{-i\omega_k t}. \tag{7}$$

It is well known, however, that the periodogram does not provide a consistent estimator of the spectral density. For consistency, it is common to instead smooth the periodogram as follows:

$$\hat{S}(\omega_k) = \sum_{j=-\infty}^{\infty} W(j) I(\omega_{k+j}) \tag{8}$$

for $W(j)$ a smoothing window that is symmetric and sums to one. To determine the width of the smoothing window we use a technique that maximizes the AIC [1].

In the context of the Whittle approximation, it is natural to apply a group lasso penalty, extending the vanilla graphical lasso approach to capture shared zero patterns across

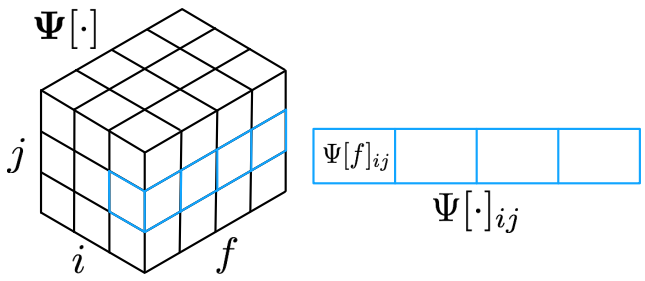

Figure 1: Representing the cross-spectrum of a multivariate time series as a tensor in $\mathbb{C}^{p\times p\times F}$. A group-lasso penalty is applied to the slices $\Psi[\cdot]_{ij}$ to encourage sharing zeros over all frequencies to learn a graphical model between the component series.

frequencies. This idea was recently explored by Jung, et. al. [13]. Let $\mathcal{C} = \{\mathbf{\Psi}[\cdot] = (\Psi[1], \ldots, \Psi[F]) : \Psi[f] \succ \mathbf{0}, \forall f \in [F]\}$ be a sequence of $p \times p$ Hermitian positive definite matrices. Also, let $\hat{S}[f] = \hat{S}(\omega_f) \in \mathbb{C}^{p\times p}$ for $f = 1, \ldots, F$. Intuitively, we can think of $\mathbf{\Psi}[\cdot]$ as a tensor in $\mathbb{C}^{p\times p\times F}$ where each $p \times p$ slice $\Psi[f]$ is a Hermitian positive-definite matrix containing the cross-spectrum at Fourier frequency ω_f, and each slice $\Psi[\cdot]_{ij}$ contains the cross-spectral density between the ith and jth signals over all frequencies. See Fig. 1 for a depiction. The goal is then to solve the following convex optimization problem:

$$\underset{\mathbf{\Psi}[\cdot]\in\mathcal{C}}{\arg\min} \sum_{f=1}^{F} -\log\det\Psi[f] + \left\langle \Psi[f], \hat{S}[f] \right\rangle + \lambda \left|\left|\mathbf{\Psi}[\cdot]\right|\right|_1 \quad (9)$$

where $\langle A, B\rangle = \mathrm{tr}\{AB\}$ for A and B Hermitian positive-definite, and $||\mathbf{\Psi}[\cdot]||_1 = \sum_{i<j}\sqrt{\sum_{f=1}^{F}\Psi[f]_{ij}}$ is the element-wise *group-lasso* penalty. The first term in Eq. (9) is the Whittle likelihood of the data represented by the smoothed periodogram at each frequency, $\hat{S}[f]$, and the group-lasso term enforces zeros to be shared across frequencies. Convex problems of this form have been used previously to simultaneously learn Gaussian graphical models for different but related groups of observations [7].

We note that moving to the frequency domain is doubly beneficial. First, the Whittle likelihood requires computing T inverses of $p \times p$ matrices in contrast to the naive method of inverting a $Tp \times Tp$ matrix in the time domain likelihood computation. Additionally, via [6], the structure learning problem is straightforward to define in the frequency domain.

3. RELATED WORK

Gaussian graphical models and spectral analysis have become popular tools for inferring brain connectivity from neuroimaging data. A similar convex penalty to Eq. (9) was developed to account for inter-subject variability in fMRI data [20]. A convex formulation to learn Gaussian graphical models specifically for vector autoregressive (VAR) processes has been developed and was applied to fMRI data [18]. This approach was then extended by incorporating a low-rank component to account for unobserved effects in neuroimaging data, however, the method remains restricted to VAR processes [14]. Additionally, notions of connectivity derived from the inverse spectral density (thus corresponding to a graphical model) have been shown to reproduce existing knowledge about functional connectivity and thus can be useful tools for neuroscientists [17].

4. GRAPHS OF TIME SERIES WITH LATENT STRUCTURE

The formulation of Eq. (9) assumes that all series of interest have been observed. Often in practice this is not the case and interactions between pairs of unrelated observed series can be inferred due to an unobserved, or latent, series. An example is when analyzing stock prices, many stocks may appear correlated, however, when the price of oil is observed the prices of those stocks become uncorrelated with each other. For i.i.d. observations a sparse plus low-rank decomposition of the underlying precision matrix has proven useful to discern the connections between observed variables from those arising from an unobserved variable [5]. These ideas have been used to learn sparse autoregressive processes in the presence of latent variables in the time domain [11]. The sparse plus low-rank framework has also been incorporated into a method to learn graphical models of VARs in the frequency domain [14], building on the work in [18]. In contrast, we consider the general family of Gaussian stationary time series rather than restricting ourselves to VAR processes.

Let $x_t = [y_t, u_t]^T \in \mathbb{R}^{p+r}$ be a completely observed stationary time series at times $t \in \{1, \ldots, T\}$ where $y_t \in \mathbb{R}^p$ are the observed components and $u_t \in \mathbb{R}^r$ are the latent components. Throughout, we assume that $r \ll p$. Using this decomposition of x_t, we can write the inverse spectral density matrix of x_t at frequency ω as

$$S(\omega)^{-1} := K(\omega) = \begin{bmatrix} K_{YY}(\omega) & K_{YU}(\omega) \\ K_{UY}(\omega) & K_{UU}(\omega) \end{bmatrix}. \quad (10)$$

The marginal inverse spectral density of the observed time series y_t is then

$$K_Y(\omega) = K_{YY}(\omega) - K_{YU}(\omega)K_{UU}(\omega)^{-1}K_{UY}(\omega), \quad (11)$$

where we will make the assumption that $K_{YY}(\omega)$ is *sparse* and $K_{YU}(\omega)K_{UU}(\omega)^{-1}K_{UY}(\omega)$ is *low-rank* [5, 8]. We then optimize the following adaptation of Eq. (9):

$$\begin{aligned}\underset{\mathbf{\Psi}[\cdot],\mathbf{L}[\cdot]}{\arg\min} &\sum_{f=1}^{F} -\log\det\{\Psi[f] - L[f]\} + \left\langle \hat{S}[f], \Psi[f] - L[f] \right\rangle \\ &+ \lambda_\Psi \left|\left|\mathbf{\Psi}[\cdot]\right|\right|_1 + \lambda_L \sum_{f=1}^{F} \mathrm{tr}\{L[f]\} \\ &\text{s.t. } \Psi[f] - L[f] \succ \mathbf{0}_{p\times p},\ L[f] \succcurlyeq \mathbf{0}_{p\times p},\ \forall f \in [F]\end{aligned} \quad (12)$$

where $\mathrm{tr}\{\mathbf{L}[\cdot]\}$ is a surrogate for the rank function and is equivalent to the sum of the eigenvalues for symmetric positive-definite matrices.

4.1 Optimizing LVSglasso

We solve Eq. (19) using a consensus ADMM algorithm [15]. First, we introduce an auxiliary variable $\mathbf{R}[\cdot] := \mathbf{S}[\cdot] - \mathbf{L}[\cdot]$

which allows us to rewrite Eq. (19) as

$$\begin{aligned}\underset{\boldsymbol{\Psi}[\cdot],\mathbf{L}[\cdot],\mathbf{R}[\cdot]}{\arg\min} & \sum_{f=1}^{F} -\log\det R[f] + \left\langle \hat{S}[f], R[f] \right\rangle \\ & + \lambda_{\Psi} \left|\left|\boldsymbol{\Psi}[\cdot]\right|\right| + \lambda_L \sum_{f=1}^{F} \left(\mathrm{tr}\{L[f]\} + \mathbb{I}[L[f] \succ \mathbf{0}] \right) \\ & \text{s.t. } \mathbf{R}[\cdot] - \mathbf{S}[\cdot] + \mathbf{L}[\cdot] = \mathbf{0}[\cdot].\end{aligned} \tag{13}$$

Applying ADMM naively to Eq. (20) is not guaranteed to converge since ADMM for three blocks of variables has been shown to converge only for a small class of functions [15]. Instead, we rewrite the problem in terms of two blocks of variables $Z = [\mathbf{R}[\cdot], \boldsymbol{\Psi}[\cdot], \mathbf{L}[\cdot]]$ and $\tilde{Z} = [\tilde{\mathbf{R}}[\cdot], \tilde{\boldsymbol{\Psi}}[\cdot], \tilde{\mathbf{L}}[\cdot]]$

$$\begin{aligned}\underset{\mathbf{Z}[\cdot],\tilde{\mathbf{Z}}[\cdot]}{\arg\min}\ & g(Z) + h(\tilde{Z}) \\ & \text{s.t. } Z - \tilde{Z} = [\mathbf{0}[\cdot], \mathbf{0}[\cdot], \mathbf{0}[\cdot]] \\ & g(Z) = \sum_{f=1}^{F} -\log\det R[f] + \left\langle \hat{S}, R[f] \right\rangle \\ & \quad + \lambda_{\Psi} \left|\left|\boldsymbol{\Psi}[\cdot]\right|\right|_1 + \lambda_L \sum_{f=1}^{F} \mathrm{tr}\{L[f]\} \\ & h(\tilde{Z}) = \mathbb{I}[\tilde{\mathbf{R}}[\cdot] - \tilde{\boldsymbol{\Psi}}[\cdot] + \tilde{\mathbf{L}}[\cdot] = \mathbf{0}[\cdot]].\end{aligned} \tag{14}$$

The updates for both Z and $\tilde{Z}$ can be found in Appendix A. We note that updating $\mathbf{R}[\cdot]$, $\mathbf{L}[\cdot]$, and $\tilde{Z}$ all parallelize over frequencies and that the update for $\boldsymbol{\Psi}[\cdot]$ parallelizes over the off-diagonal entries implying that the proposed algorithm can efficiently scale to both high-dimensional data and to many frequencies allowing for high-resolution analysis of spatio-temporal signals.

Over the course of the algorithm we diminish μ by a constant factor in order to aid convergence [15]. We use the standard ADMM stopping scheme where we keep track of the primal and dual residuals, $r_k = \sum_f R[f] - \Psi[f] + L[f]$ and $s_k = \tilde{Z}^{(k+1)} - \tilde{Z}^{(k)}$, respectively, as well as

$$\begin{aligned}\epsilon_{\text{pri}} &= \sqrt{p^2 F}\epsilon_{\text{abs}} + \epsilon_{\text{rel}} \sum_f \max\left(\left|\left|Z\right|\right|_F, \left|\left|\tilde{Z}\right|\right|_F \right) \\ \epsilon_{\text{dual}} &= \sqrt{p^2 F}\epsilon_{\text{abs}} + \epsilon_{\text{rel}} \sum_f \left|\left|\Lambda\right|\right|_F\end{aligned} \tag{15}$$

and terminate the algorithm when $\left|\left|r_k\right|\right|_2 < \epsilon_{\text{pri}}$ and $\left|\left|s_k\right|\right|_2 < \epsilon_{\text{dual}}$ [3].

5. CAPTURING GROUP HETEROGENEITY WITH A HIERARCHICAL PENALTY

Sometimes we observe multiple populations or experimental conditions and want to estimate separate graphs for each while sharing common structure between the groups. For instance, when analyzing brain imaging data from different conditions (e.g., "switch" versus "maintain" focus between auditory stimuli) we expect that similar structures might exist under different conditions (e.g., both are auditory tasks), with some variation between the conditions due to noise and other factors. The methods presented in the preceding sections can be used for this by analyzing groups separately. However, we can directly encourage shared structure within our sparse plus low-rank graphs of time series from Sec. 4 in the following manner, where the shared structure is actually the *non-edges* rather than the edges. Unlike most existing work that shares structure between populations (e.g. [7]), propagating the non-edges is useful when we expect there to be weak idiosyncratic signals within groups that existing methods will falsely detect.

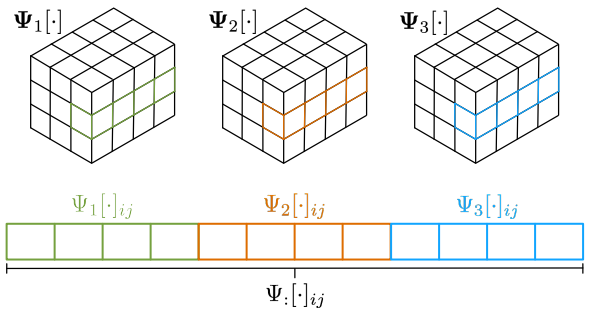

Figure 2: Representing the cross-spectra of a group of multivariate time series, $\boldsymbol{\Psi}_1[\cdot], \ldots, \boldsymbol{\Psi}_G[\cdot]$ as a group of tensors. To encourage zeros within a group we use a group-lasso penalty on each $\boldsymbol{\Psi}_g[\cdot]_{ij}$. To share zeros between groups, we concatenate the $\boldsymbol{\Psi}_g[\cdot]_{ij}$s into $\Psi_{:}[\cdot]_{ij} \in \mathbb{C}^{GF}$ and apply a group-lasso penalty.

Assume that we observe estimates of the spectral density matrices for G groups denoted $\hat{\mathbf{S}}_1[\cdot], \ldots, \hat{\mathbf{S}}_G[\cdot]$. We then consider the optimization problem

$$\begin{aligned}\underset{\substack{\boldsymbol{\Psi}_1[\cdot],\ldots,\boldsymbol{\Psi}_G[\cdot] \\ \mathbf{L}_1[\cdot],\ldots,\mathbf{L}_G[\cdot]}}{\arg\min} & \sum_{g=1}^{G} -\log\det \boldsymbol{\Psi}_g[\cdot] + \left\langle \hat{\mathbf{S}}_g[\cdot], \boldsymbol{\Psi}_g[\cdot] \right\rangle \\ & + P\{\boldsymbol{\Psi}[\cdot], \mathbf{L}[\cdot], \lambda_{\Psi}, \lambda_L\},\end{aligned} \tag{16}$$

where $\left\langle \hat{\mathbf{S}}_g[\cdot], \boldsymbol{\Psi}_g[\cdot] \right\rangle = \sum_f \mathrm{tr}\{\hat{S}_g[f]\Psi_g[f]\}$. The penalty $P\{\boldsymbol{\Psi}[\cdot], \mathbf{L}[\cdot], \lambda_{\Psi}, \lambda_L\}$ should encourage sparsity over all frequencies both within each group as well as between groups.

We use a hierarchical penalty that propagates non-edges (zeros) between the groups while simultaneously [12]. Specifically, we define

$$P\{\boldsymbol{\Psi}[\cdot], \lambda_1, \lambda_2\} = \lambda_1 \sum_{g=1}^{G} \left|\left|\boldsymbol{\Psi}_g[\cdot]\right|\right|_1 + \lambda_2 \sum_{i>j} \left|\left|\Psi_{:}[\cdot]_{ij}\right|\right|_2 \tag{17}$$

where $\Psi_{:}[\cdot]_{ij} = [\boldsymbol{\Psi}_1[\cdot]_{ij}, \ldots, \boldsymbol{\Psi}_G[\cdot]_{ij}] \in \mathbb{C}^{GF}$ and is depicted in Fig. 2. For a fixed pair of entries (i, j) the hierarchical penalty consists of two terms: The first is the 2-norm over frequencies within a group which encourages zeros to occur across all frequencies within a group. The second term is the 2-norm treating all frequencies over all groups as a vector of length GF and encourages zeros to occur across groups. This is a hierarchical penalty since the group-lasso for all frequencies within a group are nested within the group-lasso that encourages zeros between groups. As such, the proximal operator can be computed efficiently in a bottom-up fashion by first considering the within-group penalty and then feeding the result to the global group-lasso proximal update [12]. Fig. 2 depicts the hierarchical structure of this penalty. The ADMM updates can be found in Appendix B that has the same computational complexity as the algorithm presented in Appendix A.

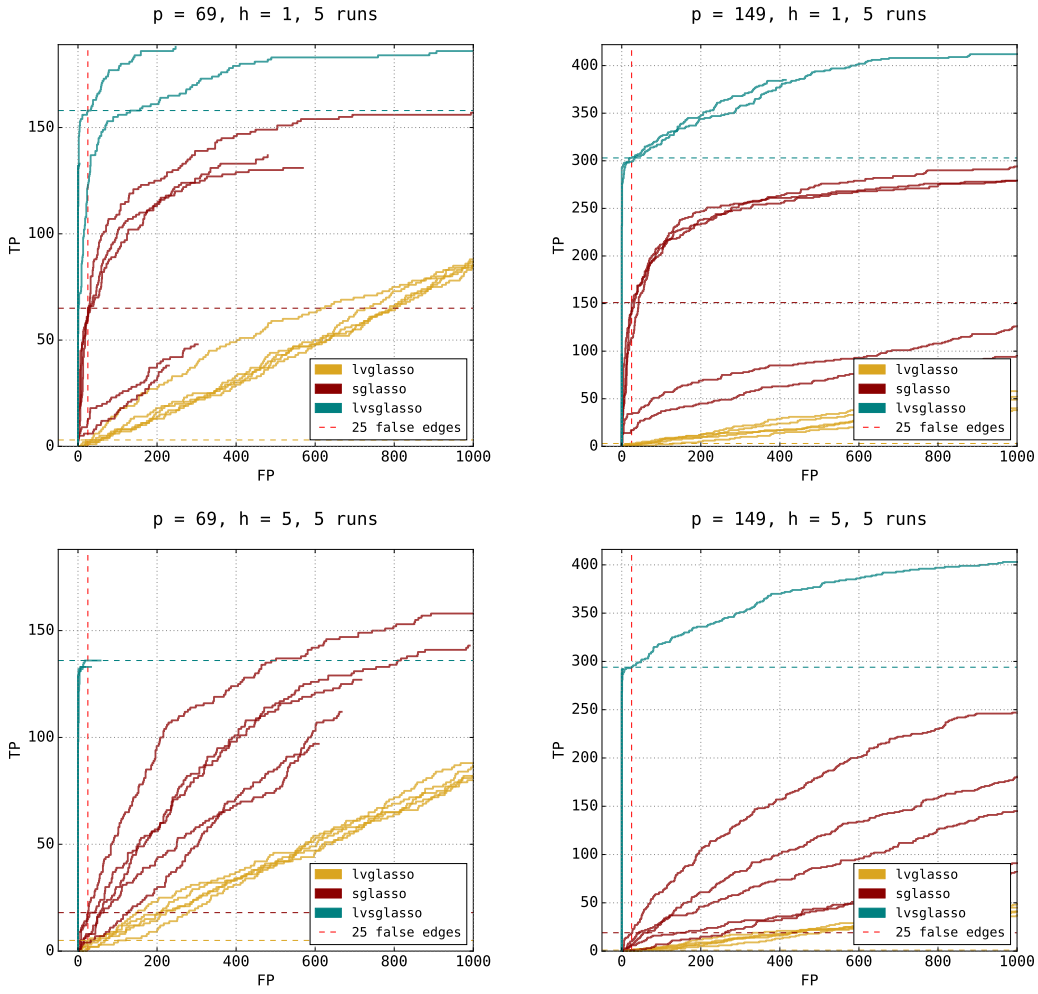

Figure 3: True positives (TP) versus false positives (FP) for the synthetic VAR(1) data for $p = 69$ and $p = 149$ and a single latent variable ($h = 1$) (top) and 5 latent variables (bottom) using `LVglasso`, `Sglasso`, and `LVSglasso`. Incorporating time-varying dynamics results in large gains over the i.i.d method and incorporating latent variables into that provides larger gains.

6. EXPERIMENTS

We first demonstrate the proposed models on synthetic data with known structure. Since interpreting inferred connectivity from MEG data is challenging, we first apply the proposed latent variable model to analyze closing prices of global stock indices which have been studied previously [18, 19]. We then apply the methods to real MEG data collected from an auditory attention task to demonstrate the structures inferred, the validation of which is left to future work. Source code for the experiments will by released on GitHub.

6.1 Synthetic Data

Single Group Comparisons.

We first evaluate the ability of our proposed sparse and low-rank decomposition of the inverse spectral density matrix on a synthetic data set with known graphical model structure for which spurious connections have been introduced by the introduction of latent variables. Specifically, we construct a VAR(1) process observed variables, $x_t \in \mathbb{R}^p$ and latent variables, $u_t \in \mathbb{R}^r$, as

$$\begin{bmatrix} x_t \\ u_t \end{bmatrix} = \underbrace{\begin{bmatrix} A & B \\ C & D \end{bmatrix}}_{A^*} + \epsilon_t \tag{18}$$

where $\epsilon_t \sim \mathbb{N}(0, \sigma I_{p+r})$ and $A \in \mathbb{R}^{p \times p}$, $B \in \mathbb{R}^{p \times r}$, $C \in \mathbb{R}^{r \times p}$, and $D \in \mathbb{R}^{r \times r}$ are the observed to observed, hidden to observed, observed to hidden, and hidden to hidden transition matrices, respectively. We set C to the matrix of all zeros, and D to a diagonal matrix with entries drawn from a standard Gaussian. The entries of B are drawn as $b_{ij} \sim N(0, 2)$ and then a random 20% of the entries in each column of B are set to zero. To construct A we first set the diagonal entries to be 0.2, then, in each row of A we randomly select two entries at random and set them to one of $\{-0.5, 0.5\}$ with probability $\frac{1}{2}$. We then construct $A^* = [A, B; C, D]$ and divide the entries by the maximum eigenvalue in order to make the process stable. Finally, we set σ so that the

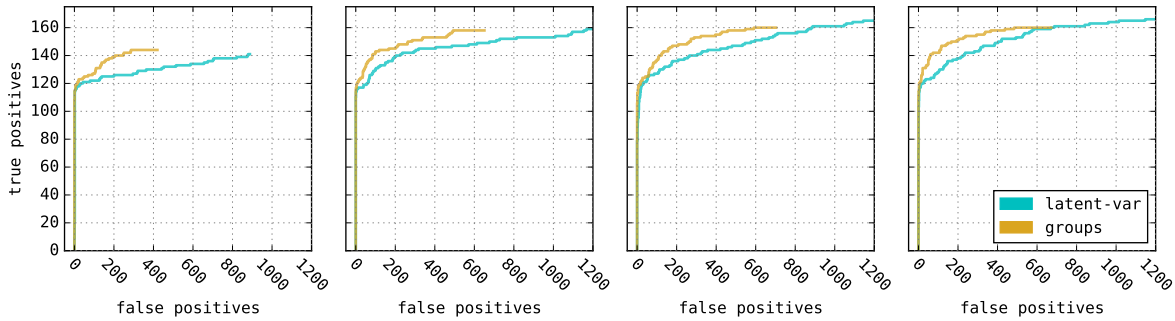

Figure 4: True-positives vs. false-positives for group synthetic example with $p = 69$ dimensions and $h = 5$ latent variables. Both `LVSglasso` and `group-LVSglasso` find the "strong" connections, however, `group-LVSglasso` introduces fewer false positives among the "weaker" connections.

signal-to-noise ratio of the resulting process is 2.[1] We then simulate realizations $[x_t, u_t]^T, t \in \{1, \ldots, 512\}$ according to Eq. (18) and only retain the x_t values. For each simulated time series we compute an estimate of the spectral density estimate by smoothing the average periodogram computed over all replicates of the series and choosing the bandwidth adaptively to minimize the AIC [1].

We compare the ability of three convex formulations to infer graphical models on the task of recovering the true underlying edges in synthetic data. Specifically, we compare the spectral graphical lasso from Sec. 2 (`Sglasso`), the sparse plus low-rank graphical lasso [5] (`LVglasso`), and our proposed sparse plus low-rank time series graphical model from Sec. 4 (`LVSglasso`). In all cases we select the regularization parameters using grid search and minimizing the Bayesian information criteria (BIC).

We report the number of true positives vs. the number of false positives learned by both models. In Fig. 3 we see that `LVSglasso` introduces many fewer false positives for the number of true positives detected as compared to `Sglasso`, where each line in the plot corresponds to a different randomly sampled VAR(1) process. From a neuroscience perspective this is a desirable property as returning many false-positives is more detrimental than missing some true positives. We additionally see that as the number of latent variables increases the gap between `LVSglasso` and the other methods widens.

Multiple Group Comparisons.

We evaluate the efficacy of the hierarchical sparsity inducing penalty when presented with multiple groups of time series. We create a synthetic data set consisting of four VAR(1) processes as discussed previously. Each group shares the same connectivity structure, except that 25% of the connections in each group are designated as "weak" connections with weights set to one of $\{-0.1, 0.1\}$. These connections represent idiosyncratic connections within a single group that are not present in the entire population. Each group also has its own latent components that interact as in Eq. (18).

We apply our `group-LVSglasso` model (Eq. (16)) with the hierarchical penalty to the generated data and `LVSglasso` independently to each group. For both models we choose the regularization parameters using grid search minimizing the BIC. The true-postive vs. false-positive curves are depicted in Fig. 4. First, note that both models are able to find the strong signals in each group. However, we see that `group-LVSglasso` introduces false-positive edges at a slower rate than `LVSglasso`.

[1]The signal-to-noise ratio of a VAR(1) process is defined as $\frac{\rho(A^*)}{\sigma}$ where $\rho(A^*)$ is the spectral norm of A^*.

6.2 Global Stock Indices

Before turning to our MEG data of interest, we first examine learning conditional independencies between the financial systems of various countries [18]. Following the experiment in Tank, et. al. [19], we acquired the daily closing prices in US dollars of 17 stock indices from various countries (see Appendix C for the full list) from June 3, 1997 until June 30, 1999 from the website `www.globalfinancialdata.com`. Missing prices were backfilled and we only considered dates where all exchanges traded in our analysis resulting in a 17-dimensional time series of length 542. As is standard when analyzing stock prices, we transformed the closing price on day t, p_t, to the log-return as $r_t = 100 \log(p_t/p_{t-1})$. Using the log-returns for each series we compute the periodogram of the 17-dimensional series, which we smooth using the method described in Sec. 2 in order to obtain a consistent estimate. We apply the three models `LVglasso`, `Sglasso`, and `LVSglasso` to the resulting estimate of the spectral density.

The graphs learned by the three models are shown in Fig. 5. `LVSglasso` learned a graph with 19 edges and used one latent variable while `Sglasso` learned a graph with 33 edges. Though the two graphs share many of the same edges, the inferred strength of the edges (thickness of the line in the figure) is different between the two graphs: `Sglasso` has many edges with comparable small weights, while `LVSglasso` depicts two strongly connected groups of nodes, i) the US, Canada, and Australia, and ii) Italy, France, and Spain. Both of these groups then have weaker connections to other Asian and Eurozone countries, respectively. This analysis demonstrates that incorporating latent variables into learning graphical models of time series can help in interpreting the inferred edges.

6.3 Auditory Attention MEG Recordings

Next, we turn our attention to our motivating application of analyzing magnetoencephalography (MEG) data. We

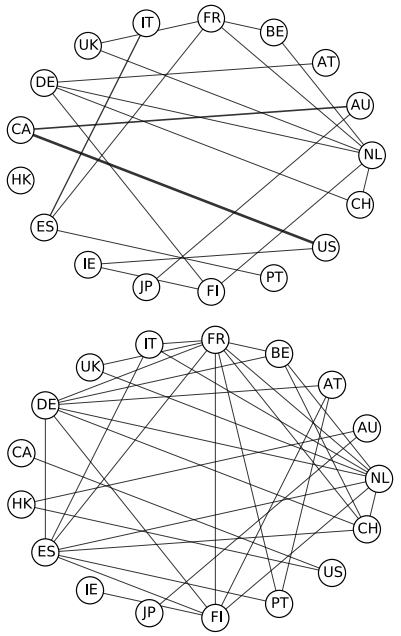

Figure 5: Graphs learned from global stock index data by `LVSglasso` (left) and `Sglasso` (right). The edge thickness indicates the strength of connection.

demonstrate that our proposed `group-LVSglasso` algorithm learns sparser graphical models than the `LVSglasso` method. Sparser models are desirable to neuroscientists in order to aid in interpreting the inferred models to then drive future experimental designs that target potential networks inferred by the model, where each experiment is costly and time consuming.

The data consists of MEG recordings from 16 individuals performing an auditory attention task. All subjects gave informed consent to participate in the study as overseen by the University of Washington Institutional Review Board. Each subject was presented with an auditory stream and was instructed to either maintain attention to that stream or switch to a different stream after presented with a stimulus. For each subject, the data consists of time series of length 1000 at each of 10,000 locations on the cortical surface. We downsample the number of series to 149 dimensions by averaging over series within predefined regions according to an existing parcellation. The resulting data set consists of a 149-dimensional time series for each subject and four conditions (maintain and switch attention over space and pitch). We split each series into small segments and consider one segment, 'Gap1', that occurs directly after stimulus is presented in order to allow transient brain activity to subside. For all four conditions we average the series over all of the subjects and compute an estimate of the spectral density matrix using the smoothed periodogram with automatic bandwidth selection [1].

We consider learning a graphical model for all four conditions for the 'Gap1' segment using the `LVSglasso` and `group-LVSglasso` models. The graphs inferred by both methods are shown in Fig. 6, where we see that `group-LVSglasso` infers a much sparser graph than `LVSglasso` which infers a denser graph with many of the edges being spatially localized and potentially spurious due to the point-spread artifact mentioned previously that arises from the MEG inverse imaging problem and affects all similar MEG analyses [10].

We are currently exploring the neuroscientific significance of the inferred connections and are applying the method to the remaining segments of the collected data where additional connections can be learned. This preliminary analysis has demonstrated that `group-LVSglasso` enables statistical information to be shared between groups of time series, resulting in learning sparser representations as compared to applying `LVSglasso` independently to all groups. Recall, when analyzing neuroimaging data that inferring false positives is more costly than missing some true positives so that the extra sparsity that `group-LVSglasso` achieves is desirable.

7. DISCUSSION

We proposed a method to learn graphical models of multiple time series that accounts for latent processes that – if omitted – could cause existing methods to learn spurious connections. We also proposed using a hierarchical sparsity inducing penalty to account for heterogeneity between groups of time series. The resulting models were formulated as convex optimization problems and efficient ADMM algorithms were developed. We evaluated the proposed framework on synthetic data as well as applied it to financial data and MEG recordings where we showed that incorporating both latent- and group-structure allows us to learn sparse models. We plan to further investigate the interpretation of the inferred edges from the MEG data. Another future direction is to extend the framework to time-frequency analysis, relaxing the stationarity assumption. Finally, we plan to explore avenues to make the framework easier to use, especially incorporating more accurate optimization algorithms to solve the sub-problems at each frequency, and also developing automatic methods to address the difficult problem of determining the regularization parameters.

8. REFERENCES

[1] F. R. Bach and M. I. Jordan. Learning graphical models for stationary time series. *IEEE Trans. Signal Process.*, 52(8):2189–2199, 2004.

[2] D. A. Bessler and J. Yang. The structure of interdependence in international stock markets. *Journal of International Money and Finance*, 22(2):261–287, 2003.

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, 2011.

[4] D. R. Brillinger. *Time Series: Data Analysis an Theory*. Holden-Day, 1981.

[5] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky. Latent variable graphical model selection via convex optimization. *The Annals of Statistics*, 40(4):1935–1967, 2012.

[6] R. Dahlhaus. Graphical interaction models for multivariate time series. *Metrika*, 51(2):157–172, 2000.

[7] P. Danaher, P. Wang, and D. M. Witten. The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical*

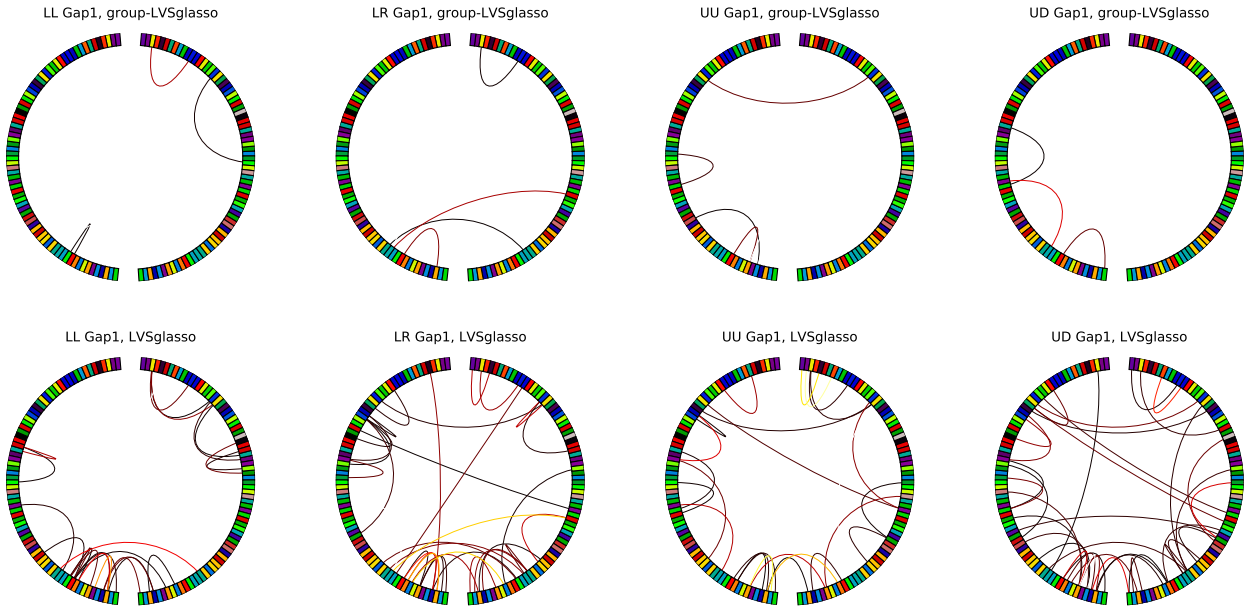

Figure 6: Inferred graphs using `group-LVSglasso` and `LVSglasso` from MEG recordings from 4 conditions of an auditory attention task. `group-LVSglasso` results in extremely sparse inferred graphs while `LVSglasso` contains many more edges making interpretation more difficult and the chances of false positives higher.

Society: Series B (Statistical Methodology), 76(2):373–397, 2014.

[8] J. Dauwels, H. Yu, X. Wang, F. Vialatte, C. Latchoumane, J. Jeong, and A. Cichocki. *Machine Learning and Interpretation in Neuroimaging: International Workshop, MLINI 2011, Held at NIPS 2011, Sierra Nevada, Spain, December 16-17, 2011, Revised Selected and Invited Contributions*, chapter Inferring Brain Networks through Graphical Models with Hidden Variables, pages 194–201. 2012.

[9] J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[10] M. S. HÃd'mÃd'lÃd'inen and R. J. Ilmoniemi. Interpreting magnetic fields of the brain: minimum norm estimates. 32:35–42, 1994.

[11] A. Jalali and S. Sanghavi. Learning the dependence graph of time series with latent factors. In *International Conference on Machine Learning*, 2012.

[12] R. Jenatton, J. Mairal, G. Obozinski, and F. R. Bach. Proximal methods for hierarchical sparse coding. *Journal of Machine Learning Research*, 12:2297–2334, 2011.

[13] A. Jung, G. Hannak, and N. Goertz. Graphical LASSO based model selection for time series. *IEEE Signal Processing Letters*, 22(10):1781–1785, 2015.

[14] R. Liégeois, B. Mishra, M. Zorzi, and R. Sepulchre. Sparse plus low-rank autoregressive identification in neuroimaging time series. In *54th IEEE Conference on Decision and Control, CDC 2015, Osaka, Japan, December 15-18, 2015*, pages 3965–3970, 2015.

[15] S. Ma, L. Xue, and H. Zou. Alternating direction methods for latent variable Gaussian graphical model. *Neural Computation*, 25(8):2172–2198, 2013.

[16] T. Medkour, A. T. Walden, and A. Burgess. Graphical modelling for brain connectivity via partial coherence. *Journal of Neuroscience Methods*, 180(2):374–383, 2009.

[17] R. Salvador, J. Suckling, C. Schwarzbauer, and E. Bullmore. Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 360(1457):937–946, 2005.

[18] J. Songsiri and L. Vandenberghe. Topology selection in graphical models of autoregressive processes. *Journal of Machine Learning Research*, 11:2671–2705, 2010.

[19] A. Tank, N. J. Foti, and E. B. Fox. Bayesian structure learning of stationary time series. In *Proc. Conference on Uncertainty in Artificial Intelligence*, July 2015.

[20] G. Varoquaux, A. Gramfort, J.-b. Poline, and B. Thirion. Brain covariance selection: Better individual functional connectivity models using population prior. In *Advances in Neural Information Processing Systems*, pages 2334–2342. 2010.

[21] C. Wang, D. Sun, and K.-C. Toh. Solving log-determinant optimization problems by a newton-cg primal proximal point algorithm. *SIAM J. Optimization*, 20:2994–3013, 2010.

APPENDIX

A. ADMM FOR LATENT VARIABLE SPECTRAL GRAPHICAL LASSO

Recall, we wish to solve the following optimization prob-

lem:

$$\begin{aligned}\underset{\boldsymbol{\Psi}[\cdot],\mathbf{L}[\cdot]}{\arg\min} \sum_{f=1}^{F} -\log\det\{\Psi[f]-L[f]\} + \left\langle \hat{S}[f], \Psi[f]-L[f]\right\rangle \\ + \lambda_\Psi \left|\left|\boldsymbol{\Psi}[\cdot]\right|\right|_1 + \lambda_L \sum_{f=1}^{F} \mathrm{tr}\{L[f]\} \\ \text{s.t. } \Psi[f]-L[f] \succ \mathbf{0}_{p\times p}, L[f] \succcurlyeq \mathbf{0}_{p\times p},\ \forall f \in [F]\end{aligned} \tag{19}$$

We will solve Eq. (19) using a consensus ADMM algorithm [3, 15]. First, we introduce an auxiliary variable $\mathbf{R}[\cdot] := \mathbf{S}[\cdot] - \mathbf{L}[\cdot]$ which allows us to rewrite Eq. (19) as

$$\begin{aligned}\underset{\boldsymbol{\Psi}[\cdot],\mathbf{L}[\cdot],\mathbf{R}[\cdot]}{\arg\min} \sum_{f=1}^{F} -\log\det R[f] + \left\langle \hat{S}[f], R[f]\right\rangle \\ + \lambda_\Psi \left|\left|\boldsymbol{\Psi}[\cdot]\right|\right| + \lambda_L \sum_{f=1}^{F} \left(\mathrm{tr}\{L[f]\} + \mathbb{I}[L[f] \succ \mathbf{0}]\right) \\ \text{s.t. } \mathbf{R}[\cdot] - \mathbf{S}[\cdot] + \mathbf{L}[\cdot] = \mathbf{0}[\cdot].\end{aligned} \tag{20}$$

We rewrite the problem in terms of two blocks of variables $Z = [\mathbf{R}[\cdot], \boldsymbol{\Psi}[\cdot], \mathbf{L}[\cdot]]$ and $\tilde{Z} = [\tilde{\mathbf{R}}[\cdot], \tilde{\boldsymbol{\Psi}}[\cdot], \tilde{\mathbf{L}}[\cdot]]$

$$\begin{aligned}\underset{\mathbf{Z}[\cdot],\tilde{\mathbf{Z}}[\cdot]}{\arg\min}\ & g(Z) + h(\tilde{Z}) \\ \text{s.t. } & Z - \tilde{Z} = [\mathbf{0}[\cdot], \mathbf{0}[\cdot], \mathbf{0}[\cdot]] \\ & g(Z) = -\sum_{f=1}^{F} \log\det R[f] + \left\langle \hat{S}, R[f]\right\rangle \\ & \quad + \lambda_\Psi \left|\left|\boldsymbol{\Psi}[\cdot]\right|\right| + \lambda_L \sum_{f=1}^{F} \mathrm{tr}\{L[f]\} \\ & h(\tilde{Z}) = \mathbb{I}[\tilde{\mathbf{R}}[\cdot] - \tilde{\boldsymbol{\Psi}}[\cdot] + \tilde{\mathbf{L}}[\cdot] = \mathbf{0}[\cdot]].\end{aligned} \tag{21}$$

The ADMM algorithm to solve Eq.(21) consists of repeating the following steps until convergence [3]:

1. $Z^{(k+1)} = \arg\min_Z g(Z) + \frac{1}{2\mu}\left|\left|Z - \tilde{Z}^{(k)} + \Lambda^{(k)}\right|\right|_F^2$
2. $\tilde{Z}^{(k+1)} = \arg\min_{\tilde{Z}} h(\tilde{Z}) + \frac{1}{2\mu}\left|\left|Z^{(k+1)} - \tilde{Z} + \Lambda^{(k)}\right|\right|_F^2$
3. $\Lambda^{(k+1)} = \Lambda^{(k)} - (Z^{(k+1)} - \tilde{Z}^{(k+1)})/\mu$

where $\Lambda = [\Lambda_R, \Lambda_\Psi, \Lambda_L]$ are dual variables.

In Step 1, the primal variables $\mathbf{R}[\cdot]$, $\boldsymbol{\Psi}[\cdot]$, and $\mathbf{L}[\cdot]$ are updated as follows. Note that the updates for $R[f]$ and $L[f]$ factorize over frequency and the updates for $\boldsymbol{\Psi}[\cdot]$ factorize of the entries. In particular, let UVU^T be the eigendecomposition of $\mu S[f] - \tilde{R}[f] - \mu\Lambda_R^{(k)}[f]$ and $v = \mathrm{diag}(V)$, then we have

$$R^{(k+1)}[f] = U\bar{V}U^T \tag{22}$$

where $\bar{V}$ is a diagonal matrix with ith diagonal entry given by $\bar{V}_{ii} = \frac{1}{2}[-v_i + \sqrt{v_i^2 + 4/\mu}]$ [7, 13]. The update for $L[f]$ is given by soft-thresholding the eigenvalues so that if $L[f] = UVU^T$, then

$$L^{(k+1)}[f] = U\bar{V}U^T \tag{23}$$

where $\bar{V}$ is a diagonal matrix with ith diagonal entry equal to $\bar{V}_{ii} = \max(V_{ii} - \lambda_L\mu, 0)$ [15]. Finally, the update for $\Psi[f]$ can be done by block-thresholding the off-diagonal entries [7]. Let $\mathcal{S}_\alpha(\mathbf{A}[\cdot]) = (B[1]_{ij}, \ldots, B[F]_{ij})$ where $B[f]_{ij} = (1 - \alpha/\left|\left|\mathbf{A}[\cdot]_{ij}\right|\right|)_+ A[f]$ and $\left|\left|\mathbf{A}[\cdot]_{ij}\right|\right| = \sqrt{\sum_{k=1}^{F} A[k]_{ij}}$. Then, the ADMM update for $\Psi[f]$ is

$$\Psi^{(k+1)} = \mathcal{S}_{\lambda_S\mu}(\Psi^{(k)} + \mu\Lambda_\Psi). \tag{24}$$

In Step 2 of the ADMM iteration we update the auxiliary variables $\tilde{Z} = [\tilde{R}, \tilde{\Psi}, \tilde{L}]$ by first defining

$$\bar{R}^{(k)} = R^{(k+1)} - \mu\Lambda_R^{(k)} \tag{25}$$

$$\bar{\Psi}^{(k)} = \Psi^{(k+1)} - \mu\Lambda_\Psi^{(k)} \tag{26}$$

$$\bar{L}^{(k)} = L^{(k+1)} - \mu\Lambda_L^{(k)} \tag{27}$$

so that we can write the updates as [15]

$$\tilde{R}^{(k+1)} = \bar{R}^{(k)} - (\bar{R}^{(k)} - \bar{\Psi}^{(k)} + \bar{L}^{(k)})/3 \tag{28}$$

$$\tilde{\Psi}^{(k+1)} = \bar{\Psi}^{(k)} - (\bar{R}^{(k)} + \bar{\Psi}^{(k)} + \bar{L}^{(k)})/3 \tag{29}$$

$$\tilde{L}^{(k+1)} = \bar{L}^{(k)} - (\bar{R}^{(k)} - \bar{\Psi}^{(k)} + \bar{L}^{(k)})/3. \tag{30}$$

The full algorithm is then given by iterating the following steps until a convergence criteria is met [3]:

1. Update R, L, and S using Eqs. (22), (23), and (24), respectively.
2. Update $\tilde{R}$, $\tilde{S}$, and $\tilde{L}$ using Eqs. (28), (29), and (30), respectively.
3. $\Lambda^{(k+1)} = \Lambda^{(k)} - (Z^{(k+1)} - \tilde{Z}^{(k+1)})/\mu$.

B. ADMM FOR GROUP-LVSGLASSO

Recall the `group-LVSglasso` problem:

$$\begin{aligned}\underset{\substack{\boldsymbol{\Psi}_1[\cdot],\ldots,\boldsymbol{\Psi}_G[\cdot] \\ \mathbf{L}_1[\cdot],\ldots,\mathbf{L}_G[\cdot]}}{\arg\min} \sum_{g=1}^{G} -\log\det\boldsymbol{\Psi}_g[\cdot] + \left\langle \hat{\mathbf{S}}_g[\cdot], \boldsymbol{\Psi}_g[\cdot]\right\rangle \\ + P\{\boldsymbol{\Psi}[\cdot], \mathbf{L}[\cdot], \lambda_\Psi, \lambda_L\}\end{aligned} \tag{31}$$

with the convex penalty penalty

$$P\{\boldsymbol{\Psi}[\cdot], \lambda_1, \lambda_2\} = \lambda_1 \sum_{g=1}^{G} \left|\left|\boldsymbol{\Psi}_g[\cdot]\right|\right|_1 + \lambda_2 \sum_{i>j} \left|\left|\Psi_{:}[\cdot]_{ij}\right|\right|_2. \tag{32}$$

We can derive an efficient ADMM algorithm by modifying the ADMM algorithm from Appendix A. Specifically, the updates for $R_g[f]$ and $L_g[f]$ are the same as in Appendix A but restricted to only the data within group g. The update for the $\boldsymbol{\Psi}_{:}[\cdot]$ variables are more complicated but can be computed efficiently due to the nesting of the proximal operators [12]. In particular, we compute the update for $\boldsymbol{\Psi}_{:}[\cdot]$ in two steps, first we update each $\boldsymbol{\Psi}_g[\cdot]$ according to

$$\boldsymbol{\Psi}_g[\cdot]^{\mathrm{tmp}} = \mathcal{S}_{\lambda_\Psi\mu}(\boldsymbol{\Psi}_s[\cdot]^{(k)} + \mu\Lambda_\Psi), \forall 1 \le i < j \le p, \tag{33}$$

where $\mathcal{S}_\alpha(\cdot)$ is again the block-thresholding operator over all frequencies. We then use these block-thresholded values as input to block-threshold the vector $\boldsymbol{\Psi}_{:}[\cdot]_{ij}$, i.e.

$$\boldsymbol{\Psi}_{:}[\cdot]_{ij}^{(k+1)} = \mathcal{S}_{\lambda_\Psi\mu}(\boldsymbol{\Psi}_{:}[\cdot]^{\mathrm{tmp}}). \tag{34}$$

We see that updating each off-diagonal entry of $\boldsymbol{\Psi}_{:}[\cdot]$ can be performed in time linear in the number of groups and frequencies. This update works because the penalty that enforces sparsity across frequencies within a group nests within the penalty that enforces zeros between groups.

C. GLOBAL STOCK INDEX COUNTRIES

Table 1 lists supplemental information about the stock indices that were analyzed in the main paper. The data was downloaded from `globalfinancialdata.com` for the dates June 3, 1997 to June 30, 1999.

Table 1: Stock index information.

Index Name	Ticker	Country	ISO 3166-1 Country Code
Amsterdam Exchange Index	AEX	Netherlands	NL
All Ordinary Composite	AORD	Australia	AU
Austrian Traded Index	ATX	Austria	AT
BEL 20	BFX	Belgium	BE
CAC 40	FCHI	France	FR
FTSEMIB	FTMIB	Italy	IT
FTSE 100	FTSE	United Kingdom	UK
DAX 30	GDAX	Germany	DE
Toronto Stock Exchange 300	GSPTSE	Canada	CA
Hang Seng Composite	HSI	Hong Kong	HK
IBEX 35	IBEX	Spain	ES
Irish Stock Exchange Index	ISEQ	Ireland	IE
Nikkei 225	N225	Japan	JP
OMX Helsinki 25	OMXH25	Finland	FI
Portugal Stock Index	PSI20	Portugal	PT
S&P 500	SPX	United States	US
Swiss Market Index	SSMI	Switzerland	CH