

Estimating Multi-Resolution Dependency Graphs within the Locally Stationary Wavelet Framework

Alex J. Gibberd
 Department of Statistical Science
 University College London
 Email: alexander.gibberd.12@ucl.ac.uk

James D. B. Nelson
 Department of Statistical Science
 University College London
 Email: j.nelson@ucl.ac.uk

Abstract—Very recently, the locally stationary wavelet framework has provided a means to describe the dependencies of co-varying time-series over a range of multiple scale levels. However, describing the many interactions between data-streams at different scale levels with only finite data poses some serious statistical estimation challenges. We illustrate that existing approaches suffer from large variance and are sometimes difficult to interpret. We here propose a sparsity-aware estimator which furnishes a set of multiresolution, dynamic graphs that describe how the dependency structure of the variables evolves through time and over multiple levels of scale. We show that the regulariser mitigates the variance and that, since the inference is performed using convex optimisation, it converges quickly to a global optima and scales well with respect to samples and nodes. Basic properties of the new method are established on simulated data. The method is applied to inferring dependency structure in multivariate EEG data-sets during epileptic seizures where it reveals evidence of band-limited dependency structure.

I. INTRODUCTION

The introduction of ever more complex sensor networks has led to an increase not only in the volume of data collected, but also its dimensionality. Traditionally, much of the literature in non-stationary time-series analysis has focused on analysing signals in the univariate setting (see [1], [2], [3], [4], [5]). When attempting to extend such methods to multivariate scenarios we are often faced with statistical estimation difficulties. Typically the number of parameters under consideration might grow quadratically with the number of variables P , whereas the number of data-points T remains the same. In the convergent disciplines of signal processing, machine learning and computational statistics, researchers look towards feature/variable selection methodologies [3], [6], [7], [8] to reduce the *degrees of freedom* in such multivariate (sometimes high-dimensional [9], [10]) settings.

We examine recent proposals [11], [12] to extend wavelet analysis to the multivariate setting, and in particular to look at dependency (or coherence) between data-streams. When considering examples with finite data sets, we demonstrate that previously proposed estimators for such models may not be consistent with the model specification. We show that some of these issues are due to an ill-conditioned estimation procedure and that the introduction of prior knowledge through regularisation can be of benefit. Introducing graphical model estimation [13], [10], [6], [14], [15] to the wavelet setting our proposed *Multiresolution exploratory graphical model*

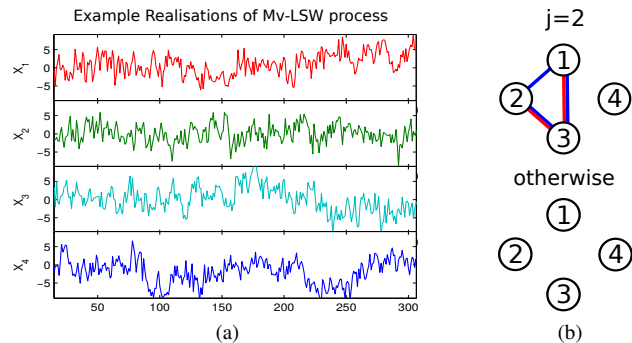


Figure 1: Band limited coherence structure. (a) Example draw from a $P = 4$ dimensional MV-LSW process. (b) Graphical representation of noise structure limited in this case to band $j = 2$, blue and red lines respectively represent non-zero coherence and partial coherence structure.

(*MR-EGM*) estimator makes a sparsity assumption about the conditional dependency structure between data-streams. This not only acts to stabilise the estimator (improving predictive performance), but also provides an interpretable representation of data-stream dependencies at different frequency/scale levels (see Fig. 1). The paper concludes with an example of our method applied to real EEG data where we study how dependency structures in brain activity may depend on frequency.

II. PRELIMINARIES

In this section we introduce the Locally Stationary Wavelet and offer some preliminaries on the interpretation of Gaussian graphical models for time-series.

Locally Stationary Wavelet Framework

The traditional family of complex exponential functions $\{e^{i\omega t}\}$ used in spectral analysis is not expressive enough to deal with non-stationarity [2]. To deal with this issue, a theory of evolutionary spectra was proposed by Priestly [2] (also see [5], [16]) whereby the exponential family is replaced by the more general class $\{\phi_t(\omega)\}$, in particular he considers families of the form $\phi_t(\omega) = C_t(\omega)e^{i\omega t}$, where $C_t(\omega)$ allows one to localise the basis about point t .

Such processes are constructed in order to generalise Fourier representations. However, an alternative basis for representing signals can be found through the application of *wavelet* methods [4], [3], [1], [11], [12]. Nason et. al. [1] construct the

locally stationary wavelet (LSW) framework by replacing the exponential basis $e^{i\omega t}$ with a set of non-decimated wavelets. Whilst the *non-decimated wavelet transform (NDWT)* is over-complete, it provides invariance with respect to translation and maintains a constant number of parameters at each scale level, i.e. for a time-series of length T we have up to $T \log(T)$ coefficients. In this paper we work with the P -variate LSW process, introduced by Park et. al. [11] and defined as

$$\mathbf{X}_{t;T} = \sum_{j=1}^{\infty} \sum_k \mathbf{V}_j(k/T) \psi_{j,t-k} \epsilon_{j,k}, \quad (1)$$

where $\mathbf{V}_j(k/T) \in \mathbb{R}^{P \times P}$ is a transfer matrix with lower-diagonal form (this encodes all dependency and contributions to the variance by wavelet at given scale/position). The discrete wavelets $\psi_{j,k}$ are formed according to the usual low-high pass iterative construction (see [1], [11]) based on the quadrature mirror filters $\{h_k\}, \{g_k\}$, such that $\psi_{1,l} = g_l$ and $\psi_{j+1,l} = \sum_k h_{l-2k} \psi_{j,l}$. Our setup follows that of Park and places all the dependency structure within the matrix $\mathbf{V}_j(u)$. The stochastic elements in the above are thus uncorrelated random vectors, whereby $E[\epsilon_{j,k}] = \mathbf{0}$ and $\text{cov}(\epsilon_{j,k}^i, \epsilon_{j',k'}^{i'}) = \delta_{j,j'} \delta_{i,i'} \delta_{k,k'}$ (this contrasts with similar work [12] where structure is split between the transfer matrix and the noise terms).

The principle object for estimation in this work is the *local wavelet spectral (LWS)* matrix:

$$\mathbf{S}_j(u) = \mathbf{V}_j(u) \mathbf{V}_j(u)^\top,$$

and inverse LWS matrix (iLWS) $\Theta_j(u) = (\mathbf{S}_j(u))^{-1}$. Loosely speaking, the LWS matrix is a covariance matrix at a specific scale level j . This is demonstrated by Park et al [11] who show the local cross-covariance $c^{(p,q)}(u, \tau)$ can be represented in terms of the LWS such that $c^{(p,q)}(u, \tau) = \sum_{j=1}^{\infty} \mathbf{S}_j^{(p,q)}(u) \Psi(\tau)$, where $\Psi(\tau) = \sum_j \psi_{j,k} \psi_{j,k-\tau}$ is the discrete auto-correlation wavelet for lag τ (see [17] for details). Asymptotically one can show $|c^{(p,q)}(u, \tau) - \text{cov}(X_{uT}^{(p)}, X_{[uT]+\tau}^{(q)})| = \mathcal{O}(T^{-1})$ [11].

Gaussian graphical models

Recent approaches to multivariate estimation using Gaussian graphical models (GGM) [13], [18], [6], [14] have utilised sparsity inducing priors to introduce bias but control variance in estimates. Our novel, proposed MV-LWS model extends these methods to estimation of the iLWS and LWS matrices.

In order to specify a valid probability density for the MV-Gaussian $\mathcal{N}(\mathbf{0}, \Sigma)$ one requires the existence of a unique inverse covariance matrix (often referred to as the precision matrix) $\Theta = (\Sigma)^{-1}$. The sparsity structure of the precision matrix plays an important role in that it specifies the conditional dependency structure of the resulting Gaussian density. If the P -variate Gaussian density (with variable index set $V = \{1, \dots, P\}$) can be decomposed as $f_{\mathcal{N}}(x_p, x_q | x_{V \setminus \{p,q\}}) = f_{\mathcal{N}}(x_p | x_{V \setminus \{p,q\}}) f_{\mathcal{N}}(x_q | x_{V \setminus \{p,q\}})$ then the corresponding precision matrix has entries $\Theta^{(p,q)} = 0$ (see [19] for a review of GGM). If $(p, q) \in E$ denote the variables for which the Gaussian can be decomposed in the aforementioned way, then one can interpret the conditional dependency structure of

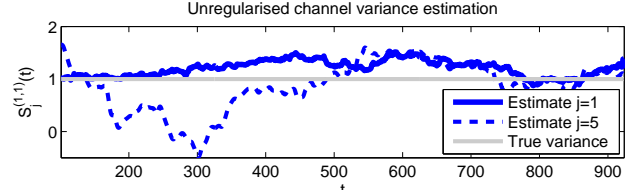


Figure 2: Estimation of the channel variance, i.e. $\hat{S}_j^{(p,q)}$ using Eq. (3), for scales $j = 1$ and $j = 5$. Whilst the true variance is equal to one in both cases, we see that the estimation error at the large scale lengths can cause negative variance estimates.

$\mathcal{N}(\mathbf{0}, \Sigma)$ as a GGM $G(V, E)$. Estimating a sparse precision matrix thus allows us to estimate the adjacency matrix of a representative graphical model.

III. ESTIMATING SPARSE PARTIAL COHERENCE STRUCTURES

Previous proposals [11] for estimating the LWS matrix are based on the estimated wavelet periodogram $\mathbf{d}_{j,k} = \sum_{t=0}^{T-1} \mathbf{X}_t \psi_{j,k}(t)$, where the smoothed periodogram matrix is

$$\mathbf{I}_{j,k} = \frac{1}{2M+1} \sum_{m=-M}^M \mathbf{d}_{j,k} \mathbf{d}_{j,k}^\top. \quad (2)$$

Whilst this estimator is consistent, it is unfortunately also biased due to leakage in the periodogram between scale levels. To correct for this Park et. al. propose the estimator:

$$\hat{\mathbf{S}}_{j,k} = \sum_{l=1}^J B_{jl}^{-1} \mathbf{I}_{j,k}, \quad (3)$$

where $B_{j,l} = \langle \Psi_j, \Psi_l \rangle = \sum_{\tau} \Psi_j(\tau) \Psi_l(\tau)$ reflects the (deterministic) leakage between wavelet scales.

In the limit $M, T \rightarrow \infty$ Park et al [11] demonstrate that asymptotically the estimator in Eq (3) is both bias free and consistent. However, in a finite sample setting entries in B_{jl}^{-1} , and therefore $\hat{\mathbf{S}}_{j,k}^{(p,p)}$, may become negative (see Fig. 2). This not only gives rise to an invalid covariance matrix, it also contradicts our process construction in Eq. (1) because, for example, $\hat{\mathbf{S}}_{j,k}$ is no longer guaranteed to be positive-semi-definite. In this paper, we introduce a means to stabilise the estimation of the covariance structure, thereby ensuring that the model assumptions are met and that an interpretable estimate of the LWS and iLWS matrices is realised.¹

Statistical estimation

If one assumes that the the wavelet periodogram is drawn from a multivariate Gaussian, $\mathbf{d}_{j,k} \sim \mathcal{N}(\mathbf{0}, \Sigma_{j,k})$ then maximum likelihood estimation of the covariance matrix is given by the empirical covariance estimator $\hat{\Sigma}_{j,k} = \mathbf{d}_{j,k} \mathbf{d}_{j,k}^\top$. If we assume that there are $k = 1, \dots, 2M+1$ observations drawn i.i.d from a Gaussian (such that $\Sigma_{j,k} = \Sigma_{j,l}$ for all $k, l \in \{0, \dots, 2M+1\}$) and take into account the deterministic bias from \mathbf{B} , then the estimator in Eq (3) $\hat{\Sigma} := \hat{\mathbf{S}}_{j,k}$ is the

¹Note: The *coherence* and *partial-coherence* referred to in Park/Sanderson [11], [12] are analogous to the usual definitions of *correlation* and *partial-correlation* whereby one normalises the LWS/iLWS matrices by the square of the diagonal component.

ML estimator. Unfortunately, as demonstrated the properties of such an estimate ($\hat{S}_{j,k}^{(i,i)} < 0$, $\hat{S}_{j,k} \not\equiv 0$) are not compatible with the parameterisation of a P -variate Gaussian [19].

To enable estimation of valid covariance matrices, we propose to estimate a sparse precision matrix (and thus GGM) at each scale level $j = 1, \dots, J$ and discrete time-step $k = 1, \dots, T$ according to the following MR-EGM estimator:

$$\hat{\Theta}_{j,k} := \arg \min_{\mathbf{Z} \succeq 0} [-\log \det(\mathbf{Z}) + \text{tr}(\hat{S}_{j,k} \mathbf{Z}) + \lambda \|\mathbf{Z}\|_1], \quad (4)$$

where $\|\mathbf{Z}\|_1 = \sum_{p,q} |Z_{p,q}|$. In the context of estimating the LWS and iLWS matrices, the above estimator has several advantages:

- 1) The bias induced by $\lambda \|\mathbf{Z}\|_1$ imposes a sparsity structure on the resultant precision matrix estimate. If the assumption of a sparse ground-truth Θ is valid, this can improve estimation performance and reduce estimator variance.
- 2) The estimated graphical models enable one to easily visualise key dependencies within a data-stream.
- 3) The constrained problem forms a convex optimisation problem, enabling fast convergence to a global optima.

There are many algorithms developed for solving problems of the form Eq (4), for simplicity we adopt the approach of Hastie et al [13] whereby the dual of (4) is solved such that rather than actively updating $\hat{\Theta}_j$ we update the constrained covariance estimator $\hat{\mathbf{W}}_j = (\hat{\Theta}_j)^{-1}$. The *graphical lasso (glasso)* algorithm [13] iterates through updating columns/rows of this matrix to arrive at a global optima. One benefit of this method, is that it can take advantage of the sparsity structure within the target matrix, generally the sparser we assume the matrix (i.e. the larger we set λ) the faster a solution will be found. The process for computing the multi-scale graphical estimators we propose is outlined in Algorithm 1.

Input: $M, \lambda_j, \psi_{j,k}, \mathbf{X}_{t:T}$
Result: $\hat{\mathbf{W}}_{j,k}, \hat{\Theta}_{j,k}$
for $j, l = 1, \dots, J$, and $k = M, \dots, T - M$ **do**
 $\mathbf{d}_{j,k} = \sum_{t=0}^{T-1} \mathbf{X}_t \psi_{j,k}(t)$
 $\mathbf{I}_{j,k} = \frac{1}{2M+1} \sum_{m=-M}^M \mathbf{d}_{j,k+m} \mathbf{d}_{j,k+m}^\top$
 $B_{j,l} = \sum_{\tau} \Psi_j(\tau) \Psi_l(\tau)$
end
 $\hat{S}_{j,k} = \sum_{l=1}^J B_{jl}^{-1} \mathbf{I}_{j,k}$
for $j = 1, \dots, J$, $k = M, \dots, T - M$ **do**
 $(\hat{\Theta}_j, \hat{\mathbf{W}}_{j,k}) = \text{glasso}(\hat{S}_{j,k} + \lambda_j \mathbf{I}, \lambda_j)$
end

Algorithm 1: MR-EGM estimation procedure

IV. SYNTHETIC EXPERIMENTS

In this section we consider synthetic experiments where we have full knowledge of the ground-truth structure. We consider the performance of the basic smoothed estimator and the graphical estimator to recover this structure when the ground-truth exhibits a degree of sparsity.

Experimental setup

In the LSW framework, data is generated according to Eq (1). The target of our estimation procedure is the inverse LWS matrix $\Theta_j^* := \mathbf{S}_j^{*-1} = (\mathbf{V}_j^* \mathbf{V}_j^{*\top})^{-1}$ (we use * to

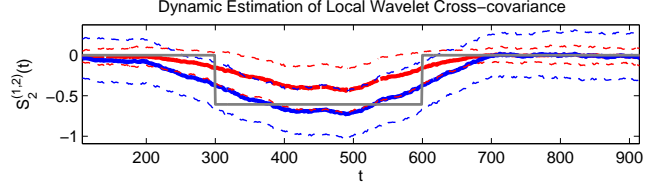


Figure 3: Example of estimation for off-diagonal cross-covariance $S_2^{(1,2)}(t)$ structure in band $j = 2$: Blue - standard estimator $\hat{S}_2^{(1,2)}$, Red - MR-EGM estimator $\hat{W}_2^{(1,2)}$, Grey - $S_2^{*(1,2)}$ ground-truth cross-covariance. The solid lines represent the mean from $N = 100$ realisations of an LSW process, dashed lines represent empirical uncertainty estimates to one standard-deviation.

denote ground-truth structures). We simulate a ground-truth iLWS matrix, which encodes a GGM with adjacency matrix $\mathbf{A} \sim \text{ErdosRenyi}(P, n)$, and then construct:

$$\Theta_j^{*(i,l)} = \begin{cases} \sim \text{Uniform}[-s, s] & A_j^{(i,l)} = 1 \\ 0 & A_j^{(i,l)} = 0 \end{cases}, \text{ for } i \neq l \quad (5)$$

where s acts to scale the iLWS off-diagonal elements (i.e. increasing the partial correlation between variables at a given scale). The transfer function $\mathbf{V}_j^*(u)$ can now be derived through LU decomposition. Realisations of the LWS process are given by Eq (1). For an example realisation see Fig (1).

Tracking dependency dynamics

To examine the ability to track band-limited structure we simulate precision matrices $\{\Theta_j^*\}$ for $P = 4$ and $T = 1024$ $j = 1, \dots, \log_2(T) = 10$. However, we restrict off-diagonal structure (as generated by Eq. 5) to a specific band $j = 2$ as depicted in Fig (1). In the interests of space, we only present a single example of estimation in this setting and contrast this to the unregularised estimator.

The example in Fig (3) has piecewise constant ground-truth $\mathbf{S}_j^*(t)$ structure, however, we observe both estimators are somewhat able to track changes. As discussed the Park estimator is unbiased, however, we observe that it is relatively sensitive to the data, i.e. has large variance (see dashed lines). Our proposed MR-EGM estimator has biased estimation in the active period ($t = 300 - 600$), but reduced estimator variance. We note, that this is especially useful, when one considers that if the sparsity assumption is true then the imposed bias actually helps (for example the region $t = 600 - 900$).

Estimator performance

In addition to the smoothing parameter M , the regularised estimator introduces a sparsity tuning parameter λ . In the synthetic setting, where the ground-truth parameterisation is known, one can easily generate test and training sets by simply simulating more realisations from the LSW process in Eq. (1). If one desired, both smoothing and regularisation parameters (M, λ) could be selected according to cross-validation based on an appropriate risk function.

For model consistency (i.e. selecting the correct sparsity pattern) one may consider maximising a balance of *precision* = $|\hat{E}_j \cap E_j^*|/|\hat{E}_j|$ and *recall* = $|\hat{E}_j \cap E_j^*|/|E_j^*|$, where

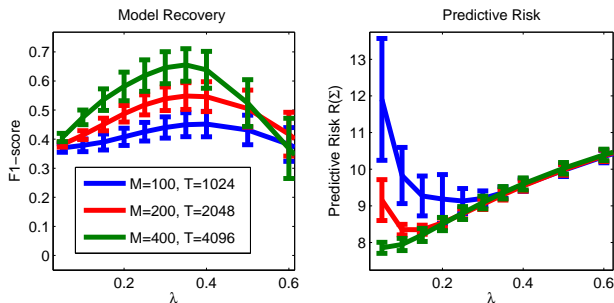


Figure 4: For each setting of λ we simulate $N = 30$ realisations and assess the mean recovery performance across the simulated data. In this case, we don't consider dynamics within the generating distribution, i.e. $\Theta^*(t) = \Theta^*(l)$, for all $t, l \in \{1, \dots, T\}$.

\hat{E}_j , E_j^* , denote the estimated and ground truth edge sets at the j th level. One popular choice for assessing model selection performance [14] is the F_1 -score defined as: $F_1 = 2(\text{precision} \times \text{recall})/(\text{precision} + \text{recall})$. Alternatively, for predictive risk (approximating the true distribution) one could minimize $R(\hat{\Theta}_j) = \text{tr}(\hat{\Theta}_j \mathbf{S}_j^*) + \log \det(\hat{\Theta}_j)$. See also Zhou et al. [15] for analysis of such risks in the non-wavelet setting.

Figure (4) gives a demonstration of how model recovery and predictive performance behave as we change the regularisation parameter and scale of the smoothing parameter (here we use $P = 10$ with $n = 10$ true edges simulated from Equation 5). With sparse models (high λ) we observe that estimator variance is reduced, however bias is increased as one would expect due to the reduced degrees of freedom. The high variance inherent within the weakly regularised estimator is apparent through the large error bars in the predictive risk for small values of λ . The large variance in estimating the covariance directly translates into variation within the predictive risk measure. We observe that regularising has clear benefits in terms of predictive performance when data is at a premium (i.e. for small M), but as M is increased the need for regularisation is reduced. Additionally, one observes that estimation of graph structure improves with increasing M .

V. APPLICATION TO HUMAN BRAIN CONNECTOMICS

In this section, we apply the proposed MR-EGM estimator in a real life setting with the study of Electroencephalography (EEG) data from the PhysioNet project (<http://www.physionet.org/>) [20], [21]. We consider studying the partial coherence structure of EEG data before and during an epileptic seizure. In particular we are interested in how the dependency structure is distributed across scale levels. If such scale localised features exist they may be useful in further analysing/diagnosing epileptic activity; for previous work in this direction see [22]. This study is not meant to form a complete analysis of EEG patterns or seizure behaviour, but rather illustrate our proposed method and its ability to analyse dependencies at different scale levels.

Given the cost of extracting the graphical model scales with $\mathcal{O}(T)$ we restrict ourselves a short segment of the data ($P = 23$ electrodes) around 100 seconds either side of a seizure and,

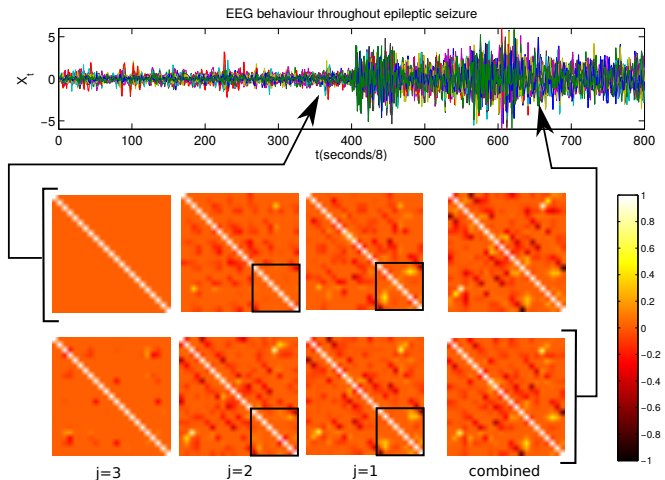


Figure 5: Estimated partial coherence structure at multiple scale levels ($j = 1, 2, 3$) via MR-EGM. In contrast with the proposed multiresolution approach, the 'combined' estimate is formed by applying graphical lasso directly to the raw data.

for the purposes of this example, down-sample the data from 256Hz to 8Hz and consider $T \approx 1000$ data points.

Results

In practice a labelled training set may not available and it can be difficult to 'learn' optimal parameters automatically. An alternative is to use an in-sample measure of estimation error by employing criteria such as AIC/BIC that actively penalise model complexity. However, when used with regularised feature selection methods such as MR-EGM they often select too many edges [18], [23]. In this work we take a more pragmatic approach and simply choose a regularisation parameter $\lambda = 0.3$ that produces an interpretable level of sparsity. The results of our analysis, illustrated in Fig. (5), demonstrate the estimated graphical dependency structure in the partial coherence matrix, i.e. $P_j^{(p,q)}(t) = \hat{\Theta}_j^{(p,q)}(t)/(\hat{\Theta}_j^{(p,p)}(t))^{\frac{1}{2}}(\hat{\Theta}_j^{(q,q)}(t))^{\frac{1}{2}}$. We observe that for the coarser scale levels, the shrinkage imposed by the ℓ_1 term effectively sets all the off-diagonal elements to zero. The combined matrices, on the right-hand side represent the result of applying the graphical lasso without performing any wavelet decomposition. We notice that the combined matrix appears to be a fix of edges estimated at the different scale levels. Interestingly, extracted structure at the start of the seizure appears to show some evidence of band limited dependency. Activity in the highlighted regions appear to only persist at higher frequencies corresponding to $j = 1$.

VI. CONCLUSION

We have demonstrated a method for extracting conditional dependency graphs from data-streams assuming a Gaussian MV-LSW process construction. Our proposed MR-EGM estimator enhances previous proposals by reducing estimator variance and increasing predictive performance in the finite sample setting. Demonstrations on EEG data suggest that the graphs extracted by our method may be useful for studying brain dynamics or other multivariate signals where we expect that some of the dependency structure may be band-limited.

REFERENCES

- [1] G. Nason, "Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pp. 1–28, 2000.
- [2] M. B. Priestly, *Spectral Analysis and Time Series*. 1981.
- [3] D. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Society*, pp. 1–28, 1995.
- [4] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, 1990.
- [5] R. Dahlhaus, "Fitting time series models to nonstationary processes," *The Annals of Statistics*, pp. 1–37, 1997.
- [6] A. J. Gibberd and J. D. B. Nelson, "High dimensional changepoint detection with a dynamic graphical lasso," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014.
- [7] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 1996.
- [8] H. Zou, T. Hastie, and R. Tibshirani, "On the degrees of freedom of the lasso," *The Annals of Statistics*, vol. 35, pp. 2173–2192, Oct. 2007.
- [9] P. Zhao and B. Yu, "On model selection consistency of Lasso," *The Journal of Machine Learning Research*, vol. 7, pp. 2541–2563, 2006.
- [10] N. Meinshausen and P. Bühlmann, "High-dimensional graphs and variable selection with the lasso," *The Annals of Statistics*, 2006.
- [11] T. Park, I. A. Eckley, and H. C. Ombao, "Estimating Time-Evolving Partial Coherence Between Signals via Multivariate Locally Stationary Wavelet Processes," *IEEE Transactions on Signal Processing*, vol. 62, pp. 5240–5250, Oct. 2014.
- [12] J. Sanderson, P. Fryzlewicz, and M. W. Jones, "Estimating linear dependence between nonstationary time series using the locally stationary wavelet model," *Biometrika*, vol. 97, pp. 435–446, Apr. 2010.
- [13] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, pp. 432–41, July 2008.
- [14] R. P. Monti, P. Hellyer, D. Sharp, R. Leech, C. Anagnostopoulos, and G. Montana, "Estimating time-varying brain connectivity networks from functional MRI time series," *NeuroImage*, Aug. 2014.
- [15] S. Zhou, J. Lafferty, and L. Wasserman, "Time varying undirected graphs," *Machine Learning*, vol. 80, pp. 295–319, Apr. 2010.
- [16] A. Walden and E. Cohen, "Statistical Properties for Coherence Estimators From Evolutionary Spectra," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4586–4597, 2012.
- [17] I. A. Eckley and G. P. Nason, "Efficient computation of the discrete autocorrelation wavelet inner product matrix," *Statistics and Computing*, vol. 15, pp. 83–92, Apr. 2005.
- [18] P. Danaher, P. Wang, and D. M. Witten, "The joint graphical lasso for inverse covariance estimation across multiple classes," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2013.
- [19] S. L. Lauritzen, *Graphical Models*. 1996.
- [20] A. Shoeb and J. Guttag, "Application of machine learning to epileptic seizure detection," *Proceedings of the 27th International Conference on Machine Learning*, 2010.
- [21] A. Goldberger, L. Amaral, and L. Glass, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," *Circulation*, 2000.
- [22] B. Hunyadi, *Learning from structured EEG and fMRI data supporting the diagnosis of epilepsy*. PhD thesis, 2014.
- [23] J. Chen and Z. Chen, "Extended Bayesian information criteria for model selection with large model spaces," *Biometrika*, vol. 95, pp. 759–771, Sept. 2008.