SPARSE GRAPHICAL MODELING OF PIECEWISE-STATIONARY TIME SERIES

Daniele Angelosante and Georgios B. Giannakis
University of Minnesota, Dept. of ECE, 200 Union St. SE, Minneapolis, MN 55455, USA

ABSTRACT

Graphical models are useful for capturing interdependencies of statistical variables in various fields. Estimating parameters describing sparse graphical models of stationary multivariate data is a major task in areas as diverse as biostatistics, econometrics, social networks, and climate data analysis. Even though time series in these applications are often non-stationary, revealing interdependencies through sparse graphs has not advanced as rapidly, because estimating such time-varying models is challenged by the curse of dimensionality and the associated complexity which is prohibitive. The goal of this paper is to introduce novel algorithms for joint segmentation and estimation of sparse, piecewise stationary, graphical models. The crux of the proposed approach is application of dynamic programming in conjunction with cost functions regularized with terms promoting the right form of sparsity in the right application domain. As a result, complexity of the novel schemes scales gracefully with the problem dimension.

Index Terms— Graphical models, sparsity, segmentation, dynamic programming, statistical learning.

1. INTRODUCTION

Graphical models offer a pictorial means of describing dependencies among random variables – a critical issue in multivariate time series analysis and statistical learning [7]. Unveiling such dependencies is of paramount importance in various applications: discovery of active genes in protein data [4], identification of controlling companies for portfolio selection in financial data [10], and data compression [9], to name a few. To facilitate interpretation and model-based inference tasks, the principle of parsimony dictates that the simplest graphical model adequately explaining the given time series data is most desirable.

This has motivated a number of recent approaches to identifying sparse graphical models of stationary time series; see e.g. [2, 4] and references therein. For stationary Gaussian series, the so-called graphical least-absolute selection and shrinkage operator (Lasso) yields sparse graphical models by promoting zeros in the inverse covariance matrix [2]. For stationary non-Gaussian multivariate series of protein data, the so-called paired group Lasso provides sparse regression models describing each variable [4].

The present paper introduces algorithms for segmenting and estimating parameters of sparse graphical models for non-stationary, and in particular piecewise stationary, time series. The intended application domains include speech, social, and financial data, where isolated events (such as company acquisitions) effect changes in one or more edges of the graph. A related approach for financial data is reported in [10], where dynamic programming is employed first to isolate segments of stationarity, and the graphical Lasso is invoked subsequently to identify sparse models per segment. Unfortunately, the heuristic initialization of [10] affects critically the convergence of this two-step iterative approach. Instead, the joint segmentation-estimation approach developed here relies on iteratively minimizing well-motivated cost functions with guaranteed convergence in polynomial time.

The rest of this paper is structured as follows. Section 2 contains preliminaries and the problem statement; the novel segmentation-identification methods are the subject of Section 3. Simulations are presented in Section 4, and conclusions are drawn in Section 5.

Notation: Column vectors (matrices) are denoted with lower- (upper-) case boldface letters; (·)H stands for transposition; 0P×Q is the P × Q matrix with all zero entries; N(μ, σ2) denotes the Gaussian probability density function with mean μ and variance σ2; for N-dimensional column vectors x1, ..., xP, X1:P := [x1, ..., xP] ∈ RN×P. The (i, j) entry of matrix X is denoted as X[i,j]; Given the square matrix X ∈ RN×N, tr(X) and det(X) denote its trace and determinant, respectively, and diag(X) := [X[1,1], ..., X[N,N]]T. The ℓ1- and ℓ2-norms of x := [x1, ..., xP]H ∈ RP are defined, respectively, as ||x||1 := ∑Pp=1|xp| and ||x||2 := √Pp=1xp2. The ℓ1- and ℓ2-norms of the square matrix X ∈ RP×P are defined, respectively, as ||X||1 := ∑Pp=1∑Pp=1|X[p,q]| and ||X||2 := √Pp=1Pp=1|X[p,q]|2.

2. PRELIMINARIES AND PROBLEM STATEMENT

Let x[t] := [x1,t, ..., xN,t]H ∈ RN denote the realization of an N-variate random process at time t = 1, ..., T, which is assumed to be piecewise-stationary and Gaussian over time segments [tk, tk+1 − 1], k = 1, ..., K; that is, with μk and
Σ_k denoting the mean vector and covariance matrix of x_t over the kth segment, it holds that
\[ x_t \sim N(\mu_k, \Sigma_k), \quad \text{for} \ t \in [\tau_k, \tau_{k+1} - 1]. \] (1)

It is further assumed that most entries of x_t per segment are pairwise uncorrelated when conditioned on the remaining entries. For a pair (x_{i,t}, x_{j,t}) on segment k, this means that in the undirected graph describing the dependencies among variables, nodes i and j are not connected; and implies that the entry (i, j) of the inverse covariance matrix \( \Theta_k := \Sigma_k^{-1} \) is zero [2]; i.e., \( \Theta_k_{i,j} = [\Theta_k]_{j,i} = 0 \) for most \( i \neq j \).

Given \( \{x_t\}_{t=1}^T \) and the number of segments K, the goal is to estimate \( \{\tau_k\}_{k=2}^K \) and the sparse matrices \( \{\Theta_k\}_{k=1}^K \).

### 3. Joint Segmentation and Identification

Suppose temporarily that \( \{\tau_k\}_{k=2}^K \) are known, and consider estimating \( \Theta_k \) using the snapshots \( \{x_t\}_{t=\tau_k}^{\tau_{k+1}-1} \), assumed to be independent over the kth segment \([\tau_k, \tau_{k+1} - 1]\). After maximizing over the mean \( \mu_k \) and omitting irrelevant terms, the log-likelihood can be expressed in terms of the inverse covariance matrix as \( L_k(\Theta_k) = \log \det(\Theta_k) - \text{tr}(S_k \Theta_k) \), where \( S_k := (\tau_{k+1} - \tau_k)^{-1} \sum_{t=\tau_k}^{\tau_{k+1}-1} (x_t - m_k)(x_t - m_k)^H \) and \( m_k := (\tau_{k+1} - \tau_k)^{-1} \sum_{t=\tau_k}^{\tau_{k+1}-1} x_t \) denote the sample covariance and sample mean, respectively.

With no sparsity constraint, the maximum likelihood (ML) estimate of \( \Theta_k \) is given by \( S_k \) [6]. One approach to account for sparsity, is to penalize the log-likelihood with the \( \ell_1 \)-norm of \( \Theta_k \). The resultant (so-called graphical Lasso) estimator is
\[
\hat{\Theta}_k = \arg \min_{\Theta_k \succeq 0, \in N^{n \times N}} -L_k(\Theta_k) + \lambda \| \Theta_k \|_1 \tag{2}
\]
and can be computed efficiently since the associated cost is convex [2, 4]. The scale \( \lambda > 0 \) can be selected via cross-validation, and controls sparsity: the larger \( \lambda \), the more zeros emerge in \( \hat{\Theta}_k \), and thus a sparser graphical model is obtained.

A natural approach to jointly estimate all \( \{\tau_k\} \) and \( \{\Theta_k\} \) is to employ as in (2) the \( \ell_1 \)-norm regularization of the negative log-likelihood of all independent snapshots \( \{x_t\}_{t=1}^T \). This leads to the following novel estimator of piecewise stationary sparse graphical model parameters:
\[
\{\tilde{\tau}_k\}_{k=2}^K, \{\tilde{\Theta}_k\}_{k=1}^K := \arg \min_{\tau_k \in \mathbb{Z}_{\geq 0}, \Theta_k \succeq 0, \in N^{n \times N}} \frac{1}{K} \sum_{k=1}^K \left( \tau_{k+1} - \tau_k \right)
\]
\[
\times \left[ \text{tr}(S_k \Theta_k) - \log \det(\Theta_k) + \lambda \| \Theta_k \|_1 \right] \tag{3}
\]

Similar to (2), the sparsity-tuning parameter \( \lambda \) can be selected using a modified form of cross-validation.

The cost in (3) is non-convex. Existence of at least one minimum is ensured since, in principle, one can try exhaustively all possible combinations of \( \{\tau_k\}_{k=2}^K \) over the interval \((1, T)\), and for each solve \( K \) graphical Lasso problems as in (2). Fortunately, a closer look reveals that this prohibitive complexity can be avoided and the cost in (3) can be minimized in polynomial time via dynamic programming. The latter yields estimates of \( \{\tilde{\tau}_k\}_{k=2}^K \), while \( \hat{\Theta}_k \)'s can be obtained as in (2). Indeed, it is possible to re-write (3) as
\[
\tilde{\tau}_k := \arg \min_{\tau_k \in \mathbb{Z}_{\geq 0}, \in N^{n \times N}} \sum_{k=1}^K \tau_{k+1} - \tau_k \sum_{t=\tau_k}^{\tau_{k+1}-1} \left( \tau_{k+1} - \tau_k \right) \sum_{t=\tau_k}^{\tau_{k+1}-1} x_t \]
\[
\text{subject to } \Theta_k \succeq 0, \in N^{n \times N} \tag{4}
\]
where
\[
F_1(\tau_k, \tau_{k+1} - 1) := (\tau_{k+1} - \tau_k) \Theta_k \min \left\{ \Theta_k \geq 0 \right\} - L_k(\Theta_k) + \lambda \| \Theta_k \|_1
\]
which involves only the variables \( \{\tilde{\tau}_k\}_{k=2}^K \).

Given \( \{\tilde{\tau}_k\}_{k=2}^K \), the function \( F_1(\tau_k, \tau_{k+1} - 1) \) can be obtained in polynomial time via graphical Lasso at complexity order \( O(N^{4.5}) \) [2]. Furthermore, the function in (4) is precisely in the form that can be minimized via dynamic programming. The latter incurs computational burden that scales as \( O(T^2 \tau N^{4.5}) \); see e.g., [5, 6].

**Remark.** As acknowledged also by [5], assuming that the number of segments is known is not formidable, since \( K \) can be estimated using order determination schemes, such as the minimum description length, with complexity order \( O(T^3 \tau N^{4.5}) \).

### 3.1. Sparse graphical models with unknown distributions

When the distribution of \( x_t \) is unknown, it is possible to estimate sparse graphical model parameters using the paired Lasso approach [4]. In the present context of piecewise stationary models, the idea is to express each variable per segment using a sparse linear regression model with the remaining variables acting as regressors.

Specifically, let the \( i \)th entry \( x_{i,t} \) of \( x_t \) over the \( k \)th segment be expressible as \( x_{i,t} = g_k^T x_t + \epsilon_{i,t} \), where \( \epsilon_{i,t} \) accounts for modeling errors, and \( g_k \) comprises the unknown vector of regression coefficients that are to be estimated, except for the \( i \)th one that is equal to zero. Concatenating these models for \( i = 1, \ldots, N \), yields a model for \( x_t \); and using the latter for \( t \in [\tau_k, \tau_{k+1} - 1] \) as columns to form \( X_k \), one arrives at the corresponding matrix model for the \( k \)th segment: \( X_k = G_k X_k + E_k \), where now \( G_k \) has all diagonal entries equal to zero. Without prior information on the distribution of \( E \), the least-squares (LS) criterion can be adopted to estimate \( G_k \), and thus reveal the linear interdependencies among variables per segment \( k \).

To account for sparsity in the links (meaning a few linear interdependencies) among variables, it suffices to regularize the LS cost in estimating the regression coefficients with a term promoting zeros in the off-diagonal entries of \( G_k \). However, since (presence) absence of a link from node variable \( i \) to variable \( j \) implies also (presence) absence of the link from \( j \) to \( i \), it follows readily that the \((i,j)\) and \((j,i)\) entries of \( G_k \)
should be (non) zero as a pair. This motivates the paired group regularization term \( \|G_k\|_2^2 + \|G_{k+1}\|_2^2 \) per entry \((i, j)\).

Putting things together, the proposed approach for sparse, piecewise stationary, graphical modeling with unknown distributions, amounts to solving

\[
\{\hat{\tau}_k\}_{k=2}^K, \{G_k\}_{k=1}^K = \arg \min_{\{\tau_k\}_{k=2}^K} \left\{ \sum_{k=1}^K (\tau_{k+1} - \tau_k) \right. \\
\left. \times \left[ \|X_k - G_k X_k\|_2^2 + \lambda \sum_{j>i} \sqrt{[G_k]_{i,j}^2 + [G_{k+1}]_{j,i}^2} \right] \right\}
\]

s.t. \( [G_k]_{i,i} = 0, \) for \( i = 1, \ldots, N, \) \( k = 1, \ldots, K \). \hspace{1cm} (5)

Noting that the cost is additive, it is possible to recast (5) as

\[
\{\hat{\tau}_k\}_{k=2}^K := \arg \min_{\{\tau_k\}_{k=2}^K} \left\{ \sum_{k=1}^K F_2(\tau_k, \tau_{k+1} - 1) \right\}
\]

where

\[
F_2(\tau_k, \tau_{k+1} - 1) := (\tau_{k+1} - \tau_k) \min_{G_k} \left[ \|X_k - G_k X_k\|_2^2 \right] \triangleq \min_{G_k} \left[ \|X_k - G_k X_k\|_2^2 \right]
\]

and

\[
+ \lambda \sum_{j>i} \sqrt{[G_k]_{i,j}^2 + [G_{k+1}]_{j,i}^2}.
\]

Given \( \{\tau_k\}_{k=2}^K \), function \( F_2(\tau_k, \tau_{k+1} - 1) \) for each segment \( k \) can be obtained by solving a paired group Lasso, while the cost in (6) is again in a form that can be efficiently solved to estimate the segmentation parameters \( \{\tau_k\}_{k=2}^K \) via dynamic programming.

### 3.2. Slowly-varying sparse graphical models

One limitation of the approaches in (3) and (5) is that the graphical models corresponding to adjacent segments are assumed to be statistically independent. In many applications, however, link changes in the graphical model occur slowly; that is, only a few connections vary while most remain invariant. In other words, graphical models in neighboring segments can be strongly dependent. For such cases, slowly-varying, sparse graphical model parameters can be estimated as follows:

\[
\{G_t\}_{t=1}^T = \arg \min_{\{G_t\}_{t=1}^T} \left\{ \sum_{t=1}^T \|X_t - G_t X_t\|_2^2 \right. \\
\left. + \lambda \sum_{t=1}^T \sum_{j>i} \sqrt{[G_t]_{i,j}^2 + [G_t]_{j,i}^2} \right\}
\]

\[
+ \gamma \sum_{t=1}^{T-1} \sum_{j>i} \sqrt{[G_{t+1}]_{i,j} - [G_t]_{i,j}}^2 + ([G_{t+1}]_{j,i} - [G_t]_{j,i})^2 \right\}
\]

s.t. \( [G_t]_{i,i} = 0, \) for \( i = 1, \ldots, N, \) \( t = 1, \ldots, T \). \hspace{1cm} (8)

The cost in (8) resembles the fused Lasso, which is convex and can thus be solved efficiently; see e.g., [1, 3]. The first regularization term promotes sparsity in \( G_t \), while the second one encourages most entries of \( G_t \) to approach those of \( G_{t+1} \). Clearly, the larger the \( \gamma \), the less the number of changes between \( G_t \) and \( G_{t+1} \).

### 4. SIMULATIONS

Due to space limitations, joint segmentation-estimation of sparse graphical models is performed using only the approach in (3). Simulated data following the model in (1) were generated with the following parameters: \( N = 9, T = 240, K = 3, \tau_2 = 81, \tau_3 = 161, \mu_1 = \mu_2 = \mu_3 = \Theta_0 \times 1 \).

The change-points and the graphical model parameters estimated using (3) with \( \lambda = 0.2, 0.3, 0.4, \) and 0.45, are depicted in Table 1 and Figures 1-4, respectively. In Figs. 1-4, a connection between nodes \( i \) and \( j \) is missing if \( [\Theta_0]_{i,j} = 0 \). Since, the number of segments is assumed known, each figure depicts the graphs in the three segments. Thus, the graph on the left of each figure corresponds to the graphical model in the first segment, the graph in the center corresponds to the graphical model in the second segment, and the graph on the right corresponds to the graphical model.
Table 1. Estimated change-points using (3).

<table>
<thead>
<tr>
<th>λ = 0.2</th>
<th>λ = 0.3</th>
<th>λ = 0.4</th>
<th>λ = 0.45</th>
</tr>
</thead>
<tbody>
<tr>
<td>τ₂</td>
<td>79</td>
<td>81</td>
<td>81</td>
</tr>
<tr>
<td>τ₃</td>
<td>161</td>
<td>161</td>
<td>161</td>
</tr>
</tbody>
</table>

in the third segment. Observe that both connectivity as well as the change-points are estimated correctly for λ = 0.4.

5. CONCLUSIONS

Computationally tractable algorithms were developed in this paper for joint segmentation-estimation of sparse graphical models to reveal dependencies among variables of piecewise-stationary multi-variate time series. The novel schemes rely on promoting the right form of sparsity in the pertinent domain, and casting the resultant optimization criteria in forms that can be solved using dynamic programming. The proposed algorithms considerably broaden the scope of existing graphical Lasso and paired group Lasso approaches to the piecewise stationary regime. The underlying multi-variate distribution can be either Gaussian, unknown, or even slowly varying. Complexity of the developed schemes scales gracefully with the problem dimension, which allows their application to large data sets. Preliminary tests on simulated data demonstrated the effectiveness of the algorithms in identifying the change-points as well as the parameters of sparse graphical models. Future research will focus on linear inter-dependencies with memory as well as non-linear interdependencies using sparse Volterra models along the lines of [8].

6. ACKNOWLEDGMENT

Work in this paper was supported by the NSF grants CCF-0830480, 1016605, and ECCS-0824007, 1002180.

7. REFERENCES


