Graph Learning for Spatiotemporal Signal with Long Short-Term Characterization

Yueliang Liu, Student Member, IEEE, Wenbin Guo, Member, IEEE, Kangyong You, Lei Zhao, Tao Peng, Wenbo Wang, Senior Member, IEEE

Abstract—Mining natural associations from high-dimensional spatiotemporal signals have received significant attention in various fields including biology, climatology and financial analysis, etcetera. Due to the widespread correlation in diverse applications, ideas that taking full advantage of correlated property to find meaningful insights of spatiotemporal signals have begun to emerge. In this paper, we study the problem of uncovering graphs that better reveal the relations behind data, with the help of long and short term correlated property in spatiotemporal signals. A spatiotemporal signal model considering both spatial and temporal relationship is firstly presented. Particularly, a low-rank representation together with a Gaussian Markov process is adopted to describe the signals’ time-correlated behavior. Next, we cast the graph learning problem as a joint low-rank component estimation and graph Laplacian inference problem. A Low-Rank and Spatiotemporal Smoothness-based graph learning method (GL-LRSS) is proposed, which novelty introduces spatiotemporal smooth prior to the field of time-vertex signal analysis. Through jointly exploiting the low-rank property of long-time observations and the smoothness of short-time observations, the overall performance is effectively improved. Experiments on both synthetic and real-world datasets demonstrate the significant improvement on learning accuracy of the proposed GL-LRSS over current state-of-the-art low-rank estimation and graph learning methods.

Index Terms—Graph learning, spatiotemporal signal, graph signal, low rank, spatiotemporal smoothness.

I. INTRODUCTION

In a variety of modern applications, from finance and sociology to transportation and sensor network, many problems in signal processing, machine learning and statistics involve the analysis of spatiotemporal signals. Much of these signals take the form of long time series measured over a certain spatial range. Examples include biomedical imaging data [1], video sequences [2], social interactions among individuals [3], and environmental sensing [4]. Due to the complex spatial and temporal correlation, together with the space-time interactions, analyzing spatiotemporal signals is a challenging problem.

Graphs are useful tools for data analysis, as it shows a flexible description of data living on an irregular domain. In recent years, graph signal processing (GSP) [5] offers an engineering paradigm for processing spatiotemporal signals on graphs, referred to as time-varying graph signals, based on spectral graph theory [6]. For analyzing and learning purposes, it is often meaningful to represent the data by means of graph, and utilize graph Laplacian matrix which is equivalent to the graph topology to deal with numerous problems including graph filtering [7], [8], graph signal compression [9] and graph signal reconstruction [10], [11], etcetera.

Nevertheless, though graph-based methods have been successful for many tasks, so far the graph structure is not always available and its natural choice (e.g., geographical K-nearest-neighbors) may not well capture the intrinsic relationship among data. The demands for graph learning that aims to spot trends or forecast future behavior from data analysis is raising. Therefore, how to extract the underlying relationship from observed spatiotemporal signals is important. In our previous work [12], we study the graph learning problem for time-varying graph signals where temporal dynamics are particularly described through a proposed space-time signal model. As such, we successfully propose an efficient graph learning method by regularizing spatiotemporal smoothness of the graph signal. However, in many cases, the collected spatiotemporal data is approximately low-rank over a long-term horizon and has short-term stability. It is essential to consider these properties in signal representation, while many studies ignore the time-correlation of signals, for example by treating the successive signal independently or processing in the entire dimensional space [13], [14]. Even though the temporal relationship in our previous work is modeled as a first-order Markov process, it lacks the long-term characterization of spatiotemporal signals. Thus, this paper focuses on an enhanced graph learning method by making full use of long short-term correlated properties in spatiotemporal signals.

A. Related works

In the literature, there has been a group of works addressing the problem of low-rank component recovery and graph learning. But these issues have been independently studied yet. For low-rank component recovery, many researches approximate spatiotemporal signals as low-rank matrices [15]–[17], which assumes the matrix collecting the time sequences to be approximately low-rank and achieves satisfying results. Recently, GSP-based approaches are proposed to recover approximated low-rank components by using spectral graph regularization.
This incorporates graph smoothness on the low-rank matrices and improves both clustering and recovery performance. It is worth mentioning that, in all these works, the graph structure is predefined based on the geometric distance, which may not be accurate enough for further analysis.

For graph learning, the early studies aim at learning graphical model, which estimates an inverse covariance matrix (i.e., precision matrix) for the Gaussian graphical model [21]–[23]. Nowadays, the fast-growing field of GSP provides a new way to solve graph learning problems. The basic idea of these methods is to identify Gaussian Markov Random Field (GMRF) models with precision matrix denoted by graph Laplacian matrix or its variants. By leveraging the smooth property of graph signals, smoothness-based methods have been adopted for graph inference. Dong et al. [13] firstly propose a valid combinatorial graph Laplacian (CGL) learning method under a smooth graph representation. Following this work, Kalofolias [14] reformulates the problem in terms of the adjacency matrix and proposes a computationally efficient algorithm. To generalize the restriction of the precision matrix being CGL, Egilmez et al. [24] identify a GMRF model whose precision matrix can be multiple types of graph Laplacian. Alternative smoothness-based approaches [12], [25], [26] also show effectiveness, with the methodological implementation provided in the former two based on temporal dynamics and edge selection, respectively. A theoretical analysis of reconstruction error is provided in [26]. The above methods learn a graph from smooth graph signals, while a few works make extra assumptions on graph dynamics for time-varying graph learning. For instance, the work in [27] proposes an algorithm to learn dynamic graphs under the hypothesis that graph changes smoothly over time, and the method of Koki et al. [28] postulate the sparseness of temporal variation on the series of learned graphs.

There is another family of approaches to tackle the graph learning problem by incorporating physical insights on graph signals. In this case, the observations are modeled as the results of a physical process on the graph, for example, diffusion-based [29]–[32] and causality-based [33]–[35] methods. To be specific, Segarra et al. [29] and Pasdeloup et al. [30] identify graphs from stationary observations that are assumed to be generated by a diffusion process. To generalize the work, Shaﬁpour et al. [31] explore the graph learning strategy that can be applied to non-stationary graph signals. Thanou et al. [32] propose a graph learning framework where the graph signals are the outcomes of heat diffusion processes. In addition, causality-based methods focus on estimating asymmetric adjacency matrix corresponding to a directed graph. In [33], Mei et al. consider a causal graph process to characterize the time series and apply it to temperature analysis. Under a structural equation model, authors in [34] propose a recursive least-square estimator to track both signal state and graph topology. Similarly, Shen et al. [35] describe the nonlinear dependency of signals via a structural vector autoregressive model and develop an efficient estimator to infer a sparse graph. Notice that graphs can be extracted from the aforementioned graph learning methods, but none of these works consider temporal correlation of observations, especially the low-rank property.

The works of [13] and [36] are most related to our work. Different from the method in [13], we focus on the spatiotemporal signals, and hence fully exploit both long and short term correlated property of spatiotemporal signals to facilitate the graph learning. Rui et al. [36] and this work jointly estimate low-rank component and graph structure, while authors in [36] propose a single integrated scheme from [13] and [37] for lack of representation between graph and observed signals. Besides, the formulation and algorithm of [36] are different from this work. As will be discussed, we present a graph-based signal representation and propose a graph learning method where the superiority and utility are verified in multiple applications.

### B. Contributions

In this paper, in order to learn a graph with high quality, a graph learning method is proposed, which takes low-rank property and local smoothness of spatiotemporal signals into consideration. Therein, leverage on the procedure of low-rank component estimation, the quality of the learned graph is well improved. In turn, the low-rank component is better estimated with the help of a refined graph. The main contributions of this paper are summarized as follows,

1) To consider both the long and short term correlation of spatiotemporal signals, we first present a graph-based model that integrates low-rank representation and temporal dynamics for comprehensive characterization. Specifically, by accommodating time and graph setting, spatiotemporal smoothness is introduced as a new prior to facilitate the graph learning procedure.

2) Under the graph-based model, the graph learning problem is formulated as a joint graph refinement and low-rank component estimation problem, which is then solved by the proposed low-rank and spatiotemporal smoothness-based graph learning method (GL-LRSS) as an application of ADMM and alternating minimization schemes. Benefit from the two optimization steps of low-rank component and graph topology, the overall learning performance can be effectively improved.

3) We perform numerous experiments on both synthetic and real-world datasets. Visual and quantitative comparison is provided in synthetic data. Besides, several classification tasks on real-world datasets demonstrate the superior performance of the proposed GL-LRSS over the state-of-the-art low-rank estimation and graph learning methods.

The remainder of this paper is organized as follows. In Section II, an overview of the notation and the preliminaries in graph signal processing are reviewed. In Section III, a low-rank graph-based model is proposed and the corresponding spatiotemporal smoothness prior is introduced. In Section IV, we formulate the graph learning problem as a joint low rank and graph topology estimation problem, and propose GL-LRSS to alternatively solve the optimization problem. The performance of GL-LRSS is presented and compared with baseline methods on both synthetic and real-world datasets in Section V. Section VI concludes the whole paper.
Provided that the eigendecomposition of CGL is a semidefinite matrix, so its eigenvalues are all non-negative. Where \( \Lambda \) is the matrix of eigenvalue, and \( \mathbf{U} \) is the matrix of eigenvector. The graph frequency spectrum is defined by the ascending array of eigenvalue \( \lambda \) and the orthogonal eigenvectors \( \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N \) are the harmonics associated with graph frequencies. In addition, the CGL of a connected graph always has a zero value of eigenvalue (i.e., \( \lambda_1 = 0 \) with multiplicity one) which corresponds to the eigenvector \( \mathbf{u}_1 = 1/\sqrt{N} \cdot \mathbf{1} \).

**C. Smooth Graph Signals**

For a graph signal \( \mathbf{x} = [x_1, x_2, \ldots, x_N]^T \), where \( x_i \) is attached to vertex \( v_i \), its frequency component is defined by the graph Fourier transform (GFT), denoted as \( \hat{x} = \mathbf{U}^T \mathbf{x} \). Here, the frequency components corresponding to a higher eigenvalue indicate the larger variations between the signals of vertices, while the ones corresponding to small eigenvalue are relatively smooth. Actually, a large number of application data show that the signals residing on graphs change smoothly between connected vertices. Such smoothness property suggests how frequently a graph signal varies with respect to the underlying graph. To quantify the smoothness of a graph signal \( \mathbf{x} \), a typical metric can be written by graph Laplacian quadratic form \( \mathbf{L} \) as

\[
S(\mathbf{x}) = \mathbf{x}^T \mathbf{L} \mathbf{x} = \sum_{(i,j) \in \mathcal{E}} (\mathbf{W})_{i,j} [x_j - x_i]^2, \tag{2}
\]

where \( \mathcal{I} = \{(i,j) \mid \{v_i, v_j\} \in \mathcal{E}\} \) is the set of index pairs of vertices. As shown in Eq. (2), it measures the total variation of connected vertices associated with the edge set \( \mathcal{E} \). From the view of vertex domain, the smaller value of \( \mathcal{I} \), the smoother of the signals on the graph structure.

**D. Correlated property of Spatiotemporal Signals**

Spatiotemporal signals can be viewed as time-varying graph signals attached to a graph of the observation sites with edges labeling their spatial relationships. As pointed out in [38], nearby values in both space and time directions tend to be more alike than those far apart. In other words, the observed signals are usually redundant as spatial and temporal correlation commonly involved. The correlated properties of spatiotemporal signals are stated as follows.

**Correlation in a long-term horizon.** Spatiotemporal signals are time series data stacked into a sequence. Due to the existence of long-term correlation, the high-dimensional time-sequence essentially lies in a lower-dimensional subspace. That is to say, the observed spatiotemporal signals are approximately low-rank [15].

**Correlation in a short-term horizon.** In addition to the long-term correlation, spatiotemporal signals are also locally correlated [12]. Concretely speaking, the observations of a certain site are correlated in neighboring time instants, and hence the temporal sequences vary smoothly over time. Similarly, at each time instant, observation sites nearby are observed spatially correlated with values being close to each other. These two types of short-term correlations are evaluated by temporal smoothness and spatial smoothness, respectively.

The past works in GSP are based on spatial and temporal smoothness. For instance, spatial smoothness is widely applied in GSP tasks including [17], [14] and [20], while quite a few works, such as [39] and [40], take advantage of temporal smoothness. Combining the two types of smoothness, our previous work [12] introduces spatiotemporal smoothness and proposes a graph learning method based on it. The spatiotemporal smoothness is recalled in Assumption 1.

**Assumption 1** (Spatiotemporal smoothness). The weighted time differences of spatiotemporal signals are smooth with respect to the graph structure.
As shown above, the spatiotemporal smoothness characterizes the short-term property of time-varying graph signals.

III. A GRAPH-BASED REPRESENTATION FOR SPATIOTEMPORAL SIGNALS

A. Graph signal representation

As illustrated in Section II-D, spatiotemporal signals are time-varying signals residing on a graph of observation sites. A spatiotemporal signal can be characterized by a matrix \( X = [x_1, x_2, \ldots, x_M] \in \mathbb{R}^{N \times M} \), where \( N \) and \( M \) are the number of observation sites and the number of time instants, respectively. With special consideration of the spatial and temporal correlation, the observed signal can be modeled as

\[
y_t = x_t + n_t, \quad (3)
\]

\[
x_t = R x_{t-1} + v_t, \quad (4)
\]

where \( y_t \in \mathbb{R}^N \) is the observation, \( x_t \in \mathbb{R}^N \) is the time-varying graph signal at time instant \( t \), and \( n_t \in \mathbb{R}^N \) encodes the perturbation part that is adopted as an isotropic noise model. We assume that the noise \( n_t \) follows a multivariate Gaussian distribution denoted as \( n_t \sim N(0, \sigma_n^2 I_N) \), with zero mean and covariance \( \sigma_n^2 I_N \).

To characterize the short-term correlation of spatiotemporal signals, we impose a first-order Gaussian Markov autoregressive process on variable \( x_t \) in Eq. (4). To be more specific, the state transition matrix is expressed by the correlation matrix \( R = \text{diag}(c_1, c_2, \ldots, c_N) \), where element \( c_i \) is the local correlation coefficient of the \( i \)-th observation site with value ranging from 0 to 1, and could be varying between different observation sites. In particular, parameter \( c \) can be obtained from the training phase in advance. Furthermore, considering the long-term correlated property, we utilize principal component analysis (PCA) for determining the approximation of the low-rank component. Meanwhile, to establish a connection between a low-rank representation of spatiotemporal signals and the graph topology, the process variable \( v_t \) is defined in the following way

\[
v_t = U^{(r)} z_t \quad \text{and} \quad z_t \sim N\left(0, \Lambda^{(r)} \right). \quad (5)
\]

As shown in Eq. (5), the low-rank matrix with rank \( r \) is a product \( v_t = U^{(r)} z_t \), where \( U^{(r)} \in \mathbb{R}^{N \times r} \) contains the basis vectors obtained from the first \( r \) eigenvectors of the graph Laplacian and \( z_t \in \mathbb{R}^r \) is the coefficient matrix. As mentioned in Section II-C, the graph Laplacian can be interpreted as graph Fourier basis for graph signal representation, which offers a natural choice for selecting basis vector. For coefficient matrix \( z_t \), it is assumed to follow \( z_t \sim N\left(0, \Lambda^{(r)} \right) \), where \( \Lambda^{(r)} \) is the Moore-Penrose pseudoinverse of eigenvalue matrix with the first \( r \) eigenvalues.

The motivation of the above definition is twofold. First, we seek a low-rank representation in terms of a small number of basis vectors where most of the variability of the data lies. The selected eigenvectors corresponding to \( r \) smallest eigenvalues can bring benefits to the smooth property. Second, the above assumption leads to a multivariate Gaussian distribution of \( v_t \), i.e., \( v_t \sim N\left(0, \hat{L}^T \right) \), where \( \hat{L}^T \) is the pseudoinverse of approximate graph Laplacian \( \hat{L} \) and it admits the following eigendecomposition \( \hat{L}^T = U^{(r)} \Lambda^{(r)} U^{(r)^T} \). Recall that \( L^T = U A^T U^T = \hat{L}^T + \Delta L^{-1}_{-N-r} \) with \( \Delta L^{-1}_{-N-r} = U_{(N-r)} \Lambda_{(N-r)}^{-1} U_{(N-r)^T} \). Since \( L^T \) is dominated by the first \( r \) eigenvalues and the corresponding eigenvectors, \( \hat{L} \) is able to approximate the graph structure. Therefore, the multivariate Gaussian distribution of \( z_t \), together with basis vectors \( U^{(r)} \), directly links the temporal evolution of graph signals with their graph structure. In other words, spatial and temporal processes closely interact with each other.

B. Spatiotemporal smoothness and interaction

In order to have a better understanding the space-time interaction, we firstly introduce weighted difference operator of graph signal \( X \) as \( D(X) = X - RXB \), where \( R \) is the local correlation matrix and \( B \) is the shift operator denoted as

\[
B = \begin{bmatrix} 0 & 1 & 0 & \cdots \cdots & 0 \\ 0 & 0 & 1 & \cdots \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots \cdots & 1 \\
\end{bmatrix}_{M \times M}.
\]

Then we have the weighted difference signal \( X \) equals to \( D(X) = [x_1, x_2 - Rx_1, x_3 - Rx_2, \ldots, x_M - Rx_{M-1}] \). As described in Assumption 1, the temporal variation of weighted difference signal under the graph structure, namely spatiotemporal smoothness, is defined as

\[
S(D(X)) = \sum_{t=1}^{M} S(x_t - Rx_{t-1}) = tr (D(X)^T L D(X)), \quad (7)
\]

where the smoothness metric \( S(\cdot) \) is shown in Eq. (2). Next, according to the proposed model in (3) and (4), we rewrite the observed signals through weighted difference operator. The \( t \)-th component is shown as below,

\[
d_t = y_t - Rx_{t-1} = v_t + n_t - Rn_{t-1}, \quad t = 1, 2, \ldots, M \quad (8)
\]

where we set \( d_1 = y_1 \). Based on the expression in (5), the conditional probability of \( d_t \) given \( v_t \) and the probability of \( d_t \) can be formulated as

\[
d_t | v_t \sim N\left(v_t, \sigma_v^2 (I_N + RR^T)\right), \quad (9)
\]

\[
d_t \sim N\left(0, \hat{L}^T + \sigma_v^2 (I_N + RR^T)\right). \quad (10)
\]

As shown in (10), spatiotemporal interactions are well illustrated in a noiseless case, that is, weighted difference signal \( d_t \) follows a degenerate multivariate Gaussian distribution with precision matrix being graph Laplacian \( \hat{L} \). It also verifies that, by defining the process variable as (5), the temporal dynamics of graph signal is efficiently related to the graph topology.

Given the observed signal \( d_t \) and the multivariate Gaussian prior distribution on \( v_t \) in (5), the maximum a posteriori (MAP) estimate of \( v_t \) by applying Bayes’ rule is expressed as

\[
v_t^{MAP} (d_t) := \arg \max_{v_t} p(v_t | d_t) = \arg \max_{v_t} \left( \log p(d_t | v_t) + p(v_t) \right).
\]

\[
= \arg \min_{v_t} \left( -\log p_E (d_t - v_t) - \log p_V (v_t) \right), \quad (11)
\]

\[
= \arg \min_{v_t} \| d_t - v_t \|^2 + \alpha v_t^T \hat{L} v_t, \quad (12)
\]

where \( p_V (v_t) \) and \( p_E (d_t - v_t) \) are the probability density function (p.d.f.) of \( v_t \) and noise \( n_t - Rn_{t-1} \), respectively. To
obtain the optimization (12), we relax the objective function in (11) and detailed procedure of the relaxation is derived in Appendix A. Notice that in (12), the regularization term $v_i^T L v_i$ is the same as $S(x_t - R x_{t-1})$, which leads to (12) with the whole time $M$ involved. As a result, it shows that our proposed model favors the spatiotemporal smoothness which can be applied to the field of time-vertex signal analysis.

IV. GRAPH LEARNING BASED ON LOW RANK AND SPATIOTEMPORAL SMOOTHNESS (GL-LRSS)

In many cases, the graph structure is typically unavailable, which makes the MAP estimate in (12) difficult to solve. Moreover, an accurate graph inference calls for a deep understanding on the property of spatiotemporal signals. Therefore, in the following, jointly exploiting the local smoothness and the global correlated property of spatiotemporal signals, we propose an efficient graph learning method. In Subsection A, we first formulate the graph learning problem. After which, an optimization algorithm to the proposed problem, GL-LRSS, is presented in Subsection B based on ADMM and alternating minimization schemes. The complexity analysis of the proposed algorithm is provided as well.

A. Problem Formulation

As mentioned in Section II-D, time-varying graph signals smoothly evolve with respect to underlying graph topology, and meanwhile exhibit low-rank property due to the long-term correlation. Hence, given the observations of $M$ time instants $Y = [y_1, y_2, \ldots, y_M] \in \mathbb{R}^{N \times M}$, we focus on two objects of interest: (i) learn the graph Laplacian matrix $L$ that is equivalent to the graph structure. (ii) achieve better low-rank component estimation. Mathematically, by imposing additional constraints on graph Laplacian and low-rank component $X$, we reformulate the graph learning problem (12) as a joint optimization problem with respect to $L$ and $X$:

\[
\text{(P1)} \quad \min_{L, X} \quad Q_1(L, X)
\]
\[
\text{s.t.} \quad Q_1(L, X) \leq \|D(X - Y)\|_F^2 + \alpha \text{tr} \left( D(X)^T LD(X) \right) + \beta \|L\|_F^2 + \gamma \|X\|_A,
\]
\[
L \in \mathcal{L}^N, \quad \text{tr} (L) = N,
\]

where $\alpha$, $\beta$ and $\gamma$ are three positive regularization parameters corresponding to the regularization terms. In addition, $\text{tr} (\cdot)$, $\| \cdot \|_F$ and $\| \cdot \|_A$ denote the trace, Frobenius norm and nuclear norm of a matrix, respectively. The first regularization $\|D(X - Y)\|_F^2$ encodes both the spatial and temporal structure of $X$ in Laplacian $L$ and weighted difference operator $D$, respectively. It enforces the weighted difference signal (i.e., $D(X)$) being smooth on graph. The smaller value of the regularizer, the more smooth evolution of graph signals over time. Moreover, as observed in (10), weighted difference signals, instead of signals themselves, exhibit better smooth property. Hence, it will improve the performance of graph learning, which is shown in Section V.

The use of $\|X\|_A$ induces a low-rank recovery. According to the inequality

\[
\text{rank}(D(X)) \leq \text{rank}(X) + \text{rank}(R XB) \leq 2\text{rank}(X),
\]

the regularizer $\|X\|_A$ promotes the weighted difference $D(X)$ to be low-rank. Although the short-term correlated property of spatiotemporal signal is described through a local correlation matrix $R$, it is not precise enough for lack of long-term characterization. Besides, spatiotemporal smoothness regularizer only favors $D(X)$ residing on the low-frequency range, it does not consider the spectral sparsity. These are where the nuclear norm compensates. The effectiveness of introducing the nuclear norm term is verified in Section V-A.

Having given the above analysis, we will give the solution to the problem. According to the fact that the better low-rank component estimation promotes the quality of the learned graph, while a good graph, in turn, facilitates an accurate low-rank component estimation. It motivates our alternating minimization framework, which iteratively refines the graph and estimates the low-rank component.

### Algorithm 1: Graph learning based on low rank and spatiotemporal smoothness (GL-LRSS)

**Input:** Observations $Y$, local correlation $R$, regularization parameters $\alpha$, $\beta$, $\gamma$, maximum iteration $K$, threshold $\varepsilon$.

1. **Initialization:** $X^0 = Y$, $k = 1$;
2. **repeat**
3. 1) Graph topology refinement:
   \[
   L^{k+1} = G(X^k, Y) \text{ by (16)}
   \]
4. 2) Low-rank component estimation:
   \[
   X^{k+1} = C(L^{k+1}, Y) \text{ by (20)-(22)}
   \]
5. 3) $(\hat{L}, \hat{X}) = (L^k, X^k)$, $k = k + 1$;
6. **until** $k = K$ or $|Q_1(L^k, X^k) - Q_1(L^{k+1}, X^{k+1})| < \varepsilon$

**Output:** Refined graph $\hat{L}$, low-rank component $\hat{X}$.
B. Optimization algorithm

As the optimization problem (P1) is not jointly convex in \( L \) and \( X \), GL-LRSS is therefore proposed to solve the above non-convex problem through an alternating optimization scheme. At each step, we optimize only one variable while holding all other variables constant. The iteration is shown as follows

1. \( G(X, Y) \triangleq \arg \min_L Q_1(L, X), \quad (S_L) \)  
   \[ \text{s.t. } L \in \mathcal{L}^N, \, \text{tr}(L) = N, \]

2. \( C(L, Y) \triangleq \arg \min_X Q_1(L, X), \quad (S_X) \)

It is interesting to find that \((S_L)\) and \((S_X)\) are two subproblems with respective to the graph Laplacian \( L \) and graph signal \( X \), respectively. By iteratively refining graph from low-rank representation and estimating the low-rank component with \( \rho > 1 \) product of matrices, while \( Z \Xi \) is the Lagrange multiplier and \( \alpha \) is a penalty parameter. Based on the augmented Lagrangian in (19), a final solution is obtained through the following iterative scheme

\[
\begin{align*}
X^{k+1} &= \arg \min_X \mathcal{L}_\rho(X, P^k, Q^k), \quad (20) \\
P^{k+1} &= \arg \min_P \mathcal{L}_\rho(X^{k+1}, P, Q^k), \\
Q^{k+1} &= Q^k + \rho (X^{k+1} - P^{k+1}).
\end{align*}
\]

According to (19), the subproblem (20) can be rewritten as

\[
\begin{align*}
X^{k+1} &= \arg \min_X \|D(X - Y)\|_F^2 + \alpha \text{tr}(D(X)^T LD(X)) + \gamma \|P\|_N^2, \\
&\quad + \frac{\rho}{2} \|X - P^{k+1}\|_F^2.
\end{align*}
\]

As we can see, the subproblem (23) is a differentiable convex optimization problem that admits a closed-form solution. For the convenience of expression, we utilize the property of the vectorization operator, that is, \( \text{vec}(AXB) = (B^T \otimes A) \text{vec}(X) \). Then, the optimal update of \( X^{k+1} \) is denoted as

\[
\text{vec}(X^{k+1}) = \left(2T_dT_d^T + 2\alpha L + \rho_{MN}^dM\right)^{-1}\left(\text{vec}(\rho P^k - Q^k) + \text{vec}(Y)\right). \quad (24)
\]

where \( \text{vec}(\cdot) \) is the vectorization operator that stacks the columns of a matrix into a vector, and the dimension of the transformed vector is \( MN \times 1 \). In addition, the parameters \( L \) and \( \bar{Y} \) are respectively represented by \( T_d(I_M \otimes L)T_d^T \) and \( 2T_dT_d^T \text{vec}(Y) \), with \( \otimes \) denoting the Kronecker product operator and \( T_d \) denoting

\[
T_d = \begin{bmatrix} I_N & -R \\
I_N & -R \\
\vdots & \ddots & \ddots & \ddots \\
I_N & -R \end{bmatrix}_{NM \times NM}.
\]

The detailed derivation of (24) is described in Appendix B.

To be noted, the solution in (24) consists of calculating the inverse of an \( MN \times MN \) dimensional matrix. With the increasing number of vertices or time instants, this procedure can be expected to be time-consuming. The conjugate gradient method [42] can be used to deal with such a problem efficiently. For simplicity, we denote the objective function 2) Low-rank component Estimation in problem \((S_X)\): As shown in (P1), the first two terms of \( X \) are differentiable. But the nuclear norm term is undifferentiable, which is typically handled by the proximal operators. Due to the decomposability and converge property of ADMM, we also choose ADMM method to tackle the problem \((S_X)\). First of all, we provide an equivalent formulation of (P1) with respect to \( X \)

\[
\begin{align*}
\min_{X,P} & \|D(X - Y)\|_F^2 + \alpha \text{tr}(D(X)^T LD(X)) + \gamma \|P\|_N^2, \\
&\quad + \frac{\rho}{2} \|X - P\|_F^2, \\
\text{s.t. } & X = P.
\end{align*}
\]

Notice that the objective function is split into two part through introducing the linear equality constraint. Then the augmented Lagrangian of (18) is as follows

\[
\begin{align*}
\mathcal{L}_\rho(X, P, Q) &= \|D(X - Y)\|_F^2 + \alpha \text{tr}(D(X)^T LD(X)) \\
&\quad + \gamma \|P\|_N + (Q, X - P) + \frac{\rho}{2} \|X - P\|_F^2,
\end{align*}
\]

where \( Q \) is the Lagrange multiplier and \( \rho \) is a penalty parameter. Based on the augmented Lagrangian in (19), a final solution is obtained through the following iterative scheme

\[
\begin{align*}
X^{k+1} &= \arg \min_X \mathcal{L}_\rho(X, P^k, Q^k), \quad (20) \\
P^{k+1} &= \arg \min_P \mathcal{L}_\rho(X^{k+1}, P, Q^k), \\
Q^{k+1} &= Q^k + \rho (X^{k+1} - P^{k+1}).
\end{align*}
\]
Algorithm 2: Method for solving subproblem (23)

Input: Y, R, B, L^{k+1}, P^k, Q^k, α, ρ, K, error tolerance δ.
1: Initialization: X_0 = 0; ∆X_0 = −∇f_X(X_0);
2: repeat
3: 1) Dynamic stepsize selection:
4: μ = \frac{\text{tr} \left( \frac{(\Delta X_m)^T \nabla f_X(X_m)}{\nabla f_X(X_m)} \right)}{\text{tr} \left( \frac{(\Delta X_m)^T (\nabla f_X(X_m) + \psi)}{\nabla f_X(X_m) + \psi} \right)},
   with ψ = 2D(Y) − 2RD(Y)B^T + ρP^k − Q^k;
5: 2) Conjugate direction update:
6: X_{m+1} = X_m + μΔX_m;
7: ∆X_{m+1} = −∇f_X(X_{m+1}) + θΔX_m;
8: m = m + 1;
9: until Stopping criterion satisfied
Output: Recovered X.

in (23) as f_X(·). In each iteration, it updates the stepsize and the searching direction. Since the f_X is differentiable, the optimal stepsize at the mth step can be decided by exact line search [43], i.e., \min_{\mu} f_X(X^m + \muΔX^m), where \mu and ΔX^m denotes the stepsize and the search direction at the mth iteration, respectively. Taking derivative with respect to \mu and then setting to zero, we can have

\text{tr} \left[ (ΔX^m)^T \nabla f_X(X^m + \muΔX^m) \right] = 0,

with the gradient of f_X calculated as

\nabla f_X = 2D(X - Y) − 2RD(X - Y)B^T + ρ(X - P^k) + Q^k + 2α[L(DX) - RLXB^T + LXBB^T].

Therefore, we can determine the optimal stepsize \mu and update the searching direction by introducing the Fletcher-Reeves parameter given as \theta = \frac{\Vert \nabla f_X(X^{m+1}) \Vert_F^2}{\Vert \nabla f_X(X^m) \Vert_F^2}. The detailed procedure of iteration is described in Algorithm 2.

Similar to the subproblem (20), by adding a constant term \frac{1}{2}\text{tr} \left( \frac{(\Delta X^k)^T Q^k}{\rho} \right), the subproblem (21) is equivalent to the following optimization problem

P^{k+1} = \arg \min_{P} \frac{1}{2} \Vert P - X^{k+1} - Q^k \Vert_F^2 + \frac{\gamma}{\rho} \Vert P \Vert_*.

The above optimization has a closed-form solution

P^{k+1} = \Gamma_{\gamma/\rho} \left( X^{k+1} + Q^k \right),

where Γ is the singular value thresholding operator [44] that is the proximity operator associated with the nuclear norm. For each τ ≥ 0, the Γ is defined as follows

\Gamma_{\tau} (X) = U \Theta_{\tau} (Σ) V^T,

where U, V and Σ are obtained from the singular value decomposition (SVD) of X, that is, X = UΣV^T, with \sigma_i denoting the i-th singular value and \Theta_{\tau} (σ_i) = \text{sign} (σ_i) \max (|σ_i| - τ, 0).

The operator (30) applies a soft-thresholding rule to the singular values of X, effectively shrinking these towards zero. The stopping criterion for solving subproblem (S_L) and (S_X) could be either a maximum number of iterations, or the change of target variable less than a threshold. By alternately minimizing the two subproblems, we can get the final solution within a few iterations. Due to the nonconvexity of the joint optimization on L and X, we can only obtain a local minimum rather than a global minimum. The detailed algorithm for solving (P1) is summarized in Algorithm 1.

Complexity analysis: In the following, we briefly discuss the computational complexity of our graph learning algorithm. For problem (S_L), the computation is dominated by the update of L in (17). The update procedure is dominated by D(X)P(X)^T where the matrix-matrix product costs O(N^2M + M^2N + N^3) computational complexity. As for problem (S_X), there are two main steps that are computation consuming. When it comes to the first step updating X^k, we utilize the conjugate gradient method instead of the calculation of (24). As shown in Algorithm 2, the computation is dominated by the gradient calculation according to (26). The gradient procedure is mainly determined by the matrix-matrix product, i.e., RLXB^T, which consumes O(N^2M + M^2N + N^3) flops. When updating P^k in the second step (21), the computation of Γ dominates the computation consumption. It takes O(min(M^2N, N^3M)) for computing the SVD of matrix X [45]. The last step of Σ update and Q update involves the product of scalar and matrix, and cost O(MN). Overall, we learn that the computation of proposed GL-LRSS is dominated by the X update in (20) and the L update in (17).

V. Experiments

The suitability of the proposed method for graph learning problem is illustrated on a wide variety of datasets: (a) two synthetic datasets under different graph structures, (b) dancer meshes representing a dancing man [46], (c) the daily temperature dataset of China from National Oceanic and Atmospheric Administration (NOAA) [47] and (d) the daily evaporation data of California from the California Department of Water Resources [48].

The proposed GL-LRSS is compared with several state-of-the-art methods, including GL-Sigrep [13], GL-logdet [23], SpecTemp [29], LGE [36], PCAG [18] and RPCAG [19]. Notice that GL-logdet and SpecTemp are graph learning methods that only infer the graph topology from observations, while PCAG and RPCAG are methods for only estimating low-rank components under a KNN graph. However, GL-LRSS, GL-Sigrep and LGE simultaneously estimate the graph and low-rank component. Since the three real-world data show highly correlated, we heuristically take R = I_N in GL-LRSS.

In our experiments, we provide both visual and quantitative comparison between the edges of the learned graph and the ones of the groundtruth graph. Particularly, we perform Monte-Carlo simulations to test the average performance of the proposed and baseline methods. To measure the estimation performance, we use low-rank component estimation error (LCE): \Vert X - X_0 \Vert_F / \Vert X_0 \Vert_F and graph structure estimation error (GSE): \Vert L - L_0 \Vert_F / \Vert L_0 \Vert_F. In addition, to evaluate the performance in terms of the recovery of the edge position in the groundtruth graph, we use four evaluation criteria, namely, Precision, Recall, F-measure and Normalized Mutual Information (NMI) [49]. The above four criteria take a value...
between 0 to 1, where the value more close to 1 implies the better graph learning performance. Specifically, the F-measure is the overall criterion that takes both Precision and Recall into consideration, and it is defined as

$$F\text{-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (31)$$

where Precision measures the percentage of the correct edges in the learned graph and Recall evaluates the percentage of edges in the groundtruth graph that are presented in the learned graph. NMI is utilized to measure the mutual dependence between the learned edge set and the groundtruth one from an information theoretic perspective. To make a fair comparison on graph learning methods, we select the best combination of regularization parameters in each method that maximizes the F-measure score, and obtain the average performance over 20 independent Monte-Carlo experiments.

### Table II

| Random geometric graph $\mathcal{G}_{\text{RGG}}$ | Grid graph $\mathcal{G}_{\text{grid}}$ |
|----|----|
| GL-LRSS | 0.8201 | 0.7878 | 0.7196 | 0.6861 | - | - | 0.7832 | 0.6913 | 0.7029 | 0.6764 | - | - |
| GL-Sigrep | 0.8709 | 0.7834 | 0.6469 | 0.8565 | - | - | 0.7633 | 0.6547 | 0.6993 | 0.7517 | - | - |
| LGE | 0.7984 | 0.6561 | 0.8212 | 0.5793 | - | - | 0.8117 | 0.7554 | 0.7575 | 0.6456 | - | - |
| GL-logdet | 0.5096 | 0.2330 | 0.2761 | 0.2138 | - | - | 0.4198 | 0.3282 | 0.3339 | 0.3033 | - | - |
| PCAG | 0.3315 | 0.3814 | 0.3445 | 0.5375 | - | - | 0.7087 | 0.7196 | 0.6861 | - | - |
| RPCAG | 0.0545 | 0.2446 | 0.1424 | - | 0.4220 | 0.2432 | 0.0665 | 0.2465 | 0.1452 | - | 0.2223 | 0.1221 |

**A. Experiments on synthetic data**

In this subsection, we test the performance of the proposed method in synthetic datasets. We first create several synthetic datasets based on a 30-vertex undirected graph, chosen from two different graph connectivity models: the grid graph $\mathcal{G}_{\text{grid}}$ and the random geometric graph $\mathcal{G}_{\text{RGG}}$. For a grid graph, each vertex with random coordinate is connected to its five nearest neighbors and the edge weight between two vertices is inversely proportional to their distance. As for the perturbation, its five nearest neighbors and the edge weight between two vertices is inversely proportional to their distance. As for the random geometric graph, we generate the coordinates of vertices uniformly at random in the unit square, determine the edge weights by a threshold Gaussian function $W(i,j) = \exp\left(-\frac{d(i,j)^2}{2\sigma^2}\right)$, where $\sigma = 0.5$, and threshold weights that are less than 0.7. After the graph construction, we compute the graph Laplacian and normalize its trace to 30.

Given a specific groundtruth graph, we next generate $30 \times 100$ time-varying graph signals $\mathbf{Y}$ based on the proposed model in (4) and (5). Without loss of generality, the local correlation matrix $\mathbf{R}$ is set as an identity matrix. We select eigenvectors corresponding smallest $r = 3$ eigenvalues as the basis vectors, i.e., the columns of $\mathbf{U}$. As for the perturbation, the standard deviation of zero-mean Gaussian noise is set to 0.5. Notice that the initial signal $\mathbf{x}_0$ and the weighted difference signal $\mathbf{x}_t - \mathbf{Rx}_{t-1}$ are smooth graph signals residing on the subspace corresponding to the 3 smallest eigenvalues of graph Laplacian $\mathbf{L}$. Hence, the time-varying graph signal $\mathbf{Y}$ is approximately low-rank and satisfies the spatiotemporal smoothness. We then apply GL-LRSS, GL-Sigrep, LGE, GL-Logdet to learn the graph Laplacian matrices given only the observation $\mathbf{Y}$. Meanwhile, GL-LRSS, GL-Sigrep, LGE, together with PCAG and RPCAG are utilized to estimate the low-rank component. Finally, we average the performance of
the proposed and baseline methods over 20 random instances of two graphs with the associated graph signals.

1) Performance comparison: We first provide a visual comparison in Fig. 1, where from left to right denotes the groundtruth graph Laplacian, the Laplacian matrices learned by GL-LRSS, GL-Sigrep, LGE and GL-Logdet. The first and the second rows denote the results under the graph model $G_{RGG}$ and $G_{grid}$, respectively. As we can see in both cases, the graph Laplacian learned by GL-LRSS is visually more consistent with the groundtruth one than the other baseline methods. For further analyzing the performance, we next show the quantitative comparison in Table I. First, on the one hand, compared with four graph learning methods, the $F$-measure increases with the decreasing score of $LCE$. It indicates that the better low-rank component estimation leads to a more accurate graph estimation. On the other hand, when it comes to five low-rank estimation methods in $G_{grid}$, the $LCE$ decreases with the increasing score of $F$-measure. Specially, the performance of PCAG and RPCAG in $G_{grid}$ is better than that in $G_{RGG}$, since the predefined graph is more close to the groundtruth one in $G_{grid}$. These results suggest that a better graph inference improves the low-rank component estimation. Thus, as two estimation steps enhance each other, it is not surprising that the performance of GL-LRSS is better than that in GL-logdet, PCAG and RPCAG. Second, the proposed GL-LRSS shows superior performance compared to the others in both graph inference and low-rank component estimation. Especially, for $G_{RGG}$, GL-LRSS achieves highest $F$-measure at 0.8201, $NMI$ scores at 0.5096 and lowest $GSE$ at 0.3315, $LCE$ scores at 0.0545. It suggests that the graph learned by GL-LRSS is more similar to the groundtruth one and the low-rank component is better recovered. Similar results can be seen in another type of graph $G_{grid}$ as well. The improvement of GL-LRSS compared to GL-Sigrep is due to the exploitation of long-term correlation, i.e., low rank. The improvement of GL-LRSS over LGE comes from the proper modeling of short-term correlation in [4], which verifies the benefits of applying spatiotemporal smoothness in graph learning procedure.

2) Algorithm analysis: To better understand the behavior of GL-LRSS under different sets of regularization parameters, we choose different powers of 2 ranging from 0 to 5, with a stepsize 0.1 for $\gamma$, and different powers of 10 ranging from 0 to -2, with a stepsize 0.1 for $\alpha$ and 2 to 0, with a stepsize 0.4 for $\beta$. For the same $G_{RGG}$ as before, we firstly plot in Fig. 2(a) the learning performance given a selected $\gamma$ under different ratios of $\beta$ to $\alpha$. We see that in Fig. 2(a) as the learned graph approaches to the groundtruth one, the curve of Recall and Precision gradually interact, leading to a peak value of $F$-measure. This implies that an appropriate ratio of $\beta$ to $\alpha$ can maximize the graph learning performance of the proposed algorithm. A similar trend can be also observed in the curve of NMI. Secondly, to investigate the effect of parameter $\gamma$, we choose the best combination of $\alpha$ and $\beta$ as illustrated in Fig. 2(a) for each value of $\gamma$. The performance of GL-LRSS under different value of $\gamma$ is depicted in Fig. 2(b). It is interesting to find that $F$-measure initially increases as the value of $\gamma$ becomes larger. This can be attributed to the fact that the unclear norm in (P1) works better with larger $\gamma$. After $F$-measure reaches its peak at 0.93 and meanwhile $LCE$ reaching the minimum, the performance decreases as the influence of unclear norm is weakened. This implies that an appropriate $\gamma$ enhances low-rank component estimation and thus results in a better graph inference. Next, to test the effectiveness of the term $\|X\|_*$, we generate time-varying graph signals for a random instance of $G_{RGG}$ under the different value of $r$. Then we infer a graph by solving (P1) with $\gamma > 0$ and $\gamma = 0$, respectively. The performance comparison of the proposed GL-LRSS and the GL-LRSS ($\gamma = 0$) without nuclear norm under different rank index is shown in Fig. 2(c). In the case of $\gamma = 0$, the nuclear norm term does not work. As for metric $F$-measure, it can be observed that GL-LRSS with $\|X\|_*$ outperforms that without $\|X\|_*$ under low rank index and the advantage of GL-LRSS with $\|X\|_*$ is less obvious when the rank index increases. This possibly due to the introduction of the nuclear norm that efficiently works in the case of lower rank index and its influence is declining as the rank index is close to 30.
Similar results can be also obtained from the evaluation metric \(NMI\) and \(GSE\). The above test verifies the correctness of the optimization model in (P1).

Finally, for one random instance of random geometric graph, we investigate the influence of the number of signals varying from 20 to 200 in steps of 20. The performance of graph estimation is shown in Fig. 3(a), we plot the criteria of \(F\)-measure and \(GSE\) to evaluate the graph learning performance. We also present the performance of GSP-based methods to serve as a baseline for Laplacian recovery. As we can see, the performance of all methods initially increases as more signals are available to learn the graph, but remains stable when the number of signals is more than 80. Moreover, the proposed GL-LRSS attains highest \(F\)-measure around 0.82 and lowest edge recovery error \(GSE\) around 0.28, which shows better graph estimation. The error of low-rank components recovered by GL-LRSS, GL-Sigrep and LGE are depicted in Fig. 3(b). The tendency of \(LCE\) is similar to that of the \(F\)-measure metric. Looking at Fig. 3(a) and 3(b) together, GL-LRSS outperforms the other methods in both graph and low-rank component estimation, possibly due to the fact that our formulation utilizes long and short term correlation of spatiotemporal signals to facilitate the learning performance.

### B. Graph learning from dancer mesh dataset

We now test the proposed graph learning method on real-world data. We first consider the dancer mesh dataset describing a dance of man dancer. It collects 143 frames representing different phases of the dance. At each frame, we consider the distance of 300 mesh vertices from each coordinates to the centroid as our observed signals. This lead to 143 time-varying graph signals (i.e., one per frame), each of dimension 300. During the whole sequence, the graph between mesh vertices is unknown and assumed to be fixed. Our object is to uncover the intrinsic graph that captures the body connectivity between mesh vertices in terms of their distances in the dance.

As mentioned in Section V-A, low-rank component recovery and graph recovery benefit from each other, leading to a consistent optimal result. Even though the groundtruth graph of mesh data is unavailable, we can focus on low-rank recovery instead. As depicted in Fig. 4, according to the movement of different body parts, the frames can be labeled by three clusters indicating three phase of dance (i.e., moving arms, stretching legs and bending body). By performing k-means clustering on recovered low-rank component, the motion classification error can indirectly reflect the graph learning performance. The \(Purity\), \(NMI\) and \(RI\) scores are used to make a quantitative evaluation on the clustering results.

### C. Graph learning from temperature dataset

The daily average temperature data is collected from 60 observation sites in China [47] over 150 days starting from January 1, 2017, and the size of data is \(60 \times 150\). By applying our graph learning method, we would like to learn a graph structure to explore the inherent relationship between these observation sites in terms of the daily variations of temperature at their locations. In this example, we do not have an available groundtruth graph. Meanwhile, the natural choice of a graph based on the geometrical distance between observation sites does not seem appropriate, which will be shown in the following analysis. However, we have that the land of China can be divided into 4 zones (i.e., northern, southern, northwest and Qinghai-Tibet). This can be viewed as a groundtruth clustering of the 60 sites, which is shown by different colors in Fig. 5(a).

For performance evaluation, we apply spectral clustering [51] to the graphs learned by GL-LRSS, GL-Sigrep, GL-Logdet, SpecTemp and LGE, respectively, and partition the vertex set into four disjoint clusters. We then compare the resulting clusters with the groundtruth information.

In Fig. 5(b) and 5(c) we visually show the four-cluster partition and the graph topology learned by GL-LRSS. We can see that the four clusters are well distinguished, which is very close to the groundtruth one in Fig. 5(a). For comparison, we also show the natural choice of the graph constructed by 8 nearest neighbors in Fig. 5(d). It is interesting to find that such a graph does not seem accurate enough as it only considers physical distance, regardless of other influence, e.g., altitude. The observation sites that are geometrically close may be geographically separated. It can be also verified by the results shown in Table IV where the best \(RI\), \(Purity\) and \(NMI\) achieved by the graph learning algorithms are presented. Compare to the baseline methods, the GL-LRSS attains the highest score in terms of all three evaluation metrics. As expected, the performance of clustering by predefined KNN graph is worse than most of the graph learning methods, which shows the
Fig. 3. (a) Graph learning performance of the baseline and proposed methods under different number of signals, and (b) low-rank component estimation performance of the baseline and proposed methods under different number of signals, for a random instance of $\mathcal{G}_{ROG}$.

Fig. 4. Clustering of the dancer mesh: the plot (below) shows for each line the average distance from centroid, and the graph is visually very similar to the groundtruth clusters shown in Fig. 6(a).

Therefore, similar to the previous examples, we apply the spectral clustering to the learned graph and compare the resulting clusters to the groundtruth clusters.

Fig. 6(b) shows the clustering results of the proposed GL-LRSS. As depicted, the clusters obtained from the learned graph is visually very similar to the groundtruth clusters. Quantitative evaluation is further compared in Table V. The $RI$ scores of the GSP-based methods, GL-Sigrep, GL-Logdet, SpecTemp, LGE and the proposed GL-LRSS are 0.8065, 0.7653, 0.7612, 0.8153 and 0.8496, respectively. It indicates that the proposed GL-LRSS is superior to the comparison graph learning methods on this ETo dataset. A similar conclusion can be also drawn in the evaluation metrics $Purity$ and $NMI$.

VI. CONCLUSION

In this paper, we study the problem of learning graphs from spatiotemporal signals with long short-term correlated properties. By exploiting the low-rank property, as well as the spatiotemporal smoothness that accommodates both the time and graph structural information for graph learning procedure, we formulate the graph learning problem as a joint low-rank component and graph topology estimation problem. A correlation-aware graph learning method, GL-LRSS, is then proposed by applying alternating minimization and ADMM schemes to solve the proposed problem. These two optimization steps facilitate from each other, leading to a better graph learning performance. Experiments on synthetic datasets verify a significant performance improvement over the state-of-the-art graph learning and low rank estimation methods. Also, experiments on three real-world datasets demonstrate that the proposed GL-LRSS outperforms these compared methods.

APPENDIX A

DERIVATION OF THE OPTIMIZATION (12) VIA RELAXATION

We firstly deduce the the MAP estimate of $\mathbf{v}_t$ in (11)

$$v_{t,MAP}(d_t) := \arg\min_{\mathbf{v}_t} 2 \left( -\log p_E(d_t - \mathbf{v}_t, \log p_V(\mathbf{v}_t)) \right)$$

$$= \arg\min_{\mathbf{v}_t} \left( -2 \log e^{-(d_t - \mathbf{v}_t)^T W^{-1} (d_t - \mathbf{v}_t)} - \alpha \log e^{-\mathbf{v}_t^T L \mathbf{v}_t} \right)$$

$$= \arg\min_{\mathbf{v}_t} \left( (d_t - \mathbf{v}_t)^T W^{-1} (d_t - \mathbf{v}_t) + \alpha \mathbf{v}_t^T L \mathbf{v}_t \right)$$
Fig. 5. (a) The locations of 60 measuring stations in China. Different colors represent the groundtruth 4 clusters that correspond to 4 geographical regions. (b) The clustering results utilizing learned graph Laplacian obtained by the GL-LRSS. (c) Graph structure learned by the GL-LRSS, which achieves the best RI score in clustering performance. (d) Graph structure established by 8 nearest neighbors according to the physical location of measuring stations. The color code in (c) and (d) represents the realistic temperature in Celcius scale on the 20th day.

Fig. 6. (a) The groundtruth clusters of 62 observation sites in California. The colors from green, blue, cyan-blue to yellow represent ETo zone 14, zone 12, zone 6 and zone 9, respectively. (b) The resulting clusters obtained by proposed GL-LRSS method.

TABLE V
THE PERFORMANCE OF GRAPH LEARNING METHODS IN RECOVERING GROUNDTRUTH CLUSTERS OF ETO MEASURING STATIONS.

|           | GL-LRSS | GL-Sigrep | LGE | SpecTemp | GL-logdet |
|-----------|---------|-----------|-----|----------|-----------|
| RI        | 0.8496  | 0.8065    | 0.8153 | 0.7612   | 0.7653    |
| Purity    | 0.8225  | 0.7419    | 0.7903 | 0.6451   | 0.6290    |
| NMI       | 0.6544  | 0.5865    | 0.5945 | 0.4799   | 0.4613    |

where scalable value (i.e., 2 in this case) is introduced for the following relaxation, $W = I_N + RR^T$ and $\alpha$ is a constant parameter proportional to the variance of noise $\sigma^2$. As we can see, (32) is a difficult problem as the objective function involves the inverse of $W$. To solve it easier, we introduce the following inequality as

$$ (d_i - v_i)^T W^{-1} (d_i - v_i) \geq \lambda_{\text{min}} (W^{-1}) \|d_i - v_i\|_2^2, \quad (33) $$

where $\lambda_{\text{min}}$ indicates the minimum eigenvalue of matrix. As $0 \leq c \leq 1$, the inequality $0 \leq c^2 \leq 1$ is satisfied. Due to the diagonal matrix of $W$, the minimum eigenvalue of $W^{-1}$ is $\lambda_{\text{min}} = \frac{1}{2}$ when $c = 1$. Finally, taking advantage of the above property, we can derive the optimization (12).

**APPENDIX B**

**DERIVATION OF THE CLOSED-FORM SOLUTION IN (24)**

Being prepared for the following analysis, we first introduce the property of the vec-operator

$$ \text{tr} (A^T B) = \text{vec}(A)^T \text{vec}(B) \quad (34) $$

Then the second term in (23) can be transformed as

$$ \text{tr} (D(X)^T L D(X)) = \text{vec}(X - RXB)^T \text{vec}[L (X - RXB)] = \text{vec}(X)^T (I_M \otimes I_N) - (B^T \otimes R) \text{vec}(X) = \text{vec}(X)^T (I_M \otimes I_N) - (B^T \otimes R) \text{vec}(X) = \text{vec}(X)^T (I_M \otimes I_N) - (B^T \otimes R) \text{vec}(X). $$

Similarly, the first term in (23) can be denoted as

$$ \|D (X - Y)\|^2_F = \text{tr} (D (X - Y)^T D (X - Y)) = \text{vec}(X - Y)^T T_d T_d^T \text{vec}(X - Y), $$

and the objective function of problem (23) can be equivalently written as

$$ f_X (v) = (v^T - \text{vec}(Y)^T) T_d T_d^T (v - \text{vec}(Y)) + \alpha v^T G v + \frac{\rho}{2} [v^T - \text{vec}(P)^T + \text{vec}(Q)^T / \rho] [v - \text{vec}(P) + \text{vec}(Q) / \rho], $$

where $G = T_d (I_M \otimes I_N) T_d^T \in \mathbb{R}^{NM \times NM}$, and $v = \text{vec}(X)$. The gradient of $f_X (v)$ can be deduced as

$$ \nabla f_X (v) = 2 T_d T_d^T v - 2 T_d T_d^T \text{vec}(Y) + 2 \alpha G v + \text{vec}(Q) + \rho v - \rho \text{vec}(P). \quad (35) $$

By setting $\nabla f_X (v)$ to zero, the unique optimal solution $\text{vec}(X)$ can be obtained as (24).

**REFERENCES**

[1] A. R. McIntosh, W. K. Chau, A. B. Protzner, “Spatiotemporal analysis of event-related fMRI data using partial least squares,” *Neuroimage*, vol. 23, no. 2, pp. 764-775, 2004.

[2] I. Kompatsiaris, and M. Strintzis, “Spatiotemporal Segmentation and Tracking of Objects for Visualization of Videoconference Image Sequences,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 8, pp. 1388-1402, 2000.
