Spatio-Temporal Signal Recovery Based on Low Rank and Differential Smoothness

Xianghui Mao, *Student Member, IEEE*, Kai Qiu, Tiejian Li, Yuantao Gu, *Senior Member, IEEE*

***Abstract*—The analysis of spatio-temporal signals plays an important role in various fields including sociology, climatology, and environmental studies, etcetera. Due to the abrupt breakdown of the sensors in the sensor network, always there are missing entries in the observed spatio-temporal signals. In this paper, we study the problem of recovering spatio-temporal signals from partially known entries. Based on both the global and local correlated property of spatio-temporal signals, we propose a Low Rank and Differential Smoothness-based recovery method (LRDS), which novelly introduces the differential smooth prior of time-varying graph signals to the field of spatio-temporal signal analysis. The performance of the proposed method is analyzed theoretically. Considering the case where a priori information about the signal's global pattern is available, we propose Prior LRDS (PLRDS) to further improve the reconstruction accuracy. Such improvement is also verified by synthetic experiments. Besides, experiments on several real-world datasets demonstrate the improvement on recovery accuracy of the proposed LRDS over the state-of-the-art spatio-temporal signal recovery methods.**

***Index Terms*—spatio-temporal signal, graph signal, signal recovery, differential smoothness, low rank**

I. Introduction

A. Motivation

Spatio-temporal signals are time series collected over a certain spatial range. The analysis of spatio-temporal signals [1] plays an important role in many fields including sociology [2], [3], climatology [4], [5], environmental studies [6], [7], and medical imaging [8], [9], etcetera. For example, spatio-temporal signals are prevalent in climate researches. Typically a large number of sensors are deployed to collect hourly or daily data for a long time. Because of the complex spatial and temporal dependencies involved, analyzing spatio-temporal signals is a challenging problem.

In many cases, the collected spatio-temporal data is incomplete [10] due to sensor malfunctions [11] or conservation of resources [12], hence the recovery of spatio-temporal signals is a fundamentally important issue. For example, sea surface temperature provides important information of the earth's climate dynamics. However, the spatial distribution of the sea surface temperature data collected from ships has varied in history due to economic and political changes, such as the opening of new canals and world wars, which significantly raises demands for reconstructing the missing global sea surface temperature [5]. For another example, in low-cost commodity sensor network, such as air temperature sensor network, due to sensor malfunctions or communication failures, lost data are common [11], which also calls for the reconstruction of spatio-temporal signals. Thus this paper studies the recovery of spatio-temporal signals, i.e., to recover the missing values from partially known entries.

Received January 24, 2018; revised June 8, and August 20, 2018; accepted September 25, 2018. This work was partly funded by the National Natural Science Foundation of China (NSFC 61571263, 61531166005) and the National Key Research and Development Program of China (Project No. 2016YFE0201900, 2017YFC0403600). (Corresponding author: Yuantao Gu.)

X. Mao and Y. Gu are with Beijing National Research Center for Information Science and Technology (BNRist) and the Department of Electronic Engineering, Tsinghua University, Beijing, China (maoxh92@sina.com, gyt@tsinghua.edu.cn).

K. Qiu is with Academy of Military Sciences PLA China, Beijing, China (q1987k@163.com).

T. Li is with State Key Laboratory of Hydroscience and Engineering, Tsinghua University, Beijing, China (litiejian@tsinghua.edu.cn).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the author. The material includes the Matlab code and the datasets for the experiments in this paper. Contact gyt@tsinghua.edu.cn for further questions about this work.

B. Related Works

Nowadays many studies approximate spatio-temporal signals with low-rank matrices [12], [13], [14], [15], [16], and achieve satisfying results. A low-rank matrix estimation-based spatio-temporal image reconstruction method is investigated for dynamic photoacoustic computed tomography in [13], which assumes that the matrix collecting the sequence of vectorized images as its columns is approximately low-rank. The same assumption is also made in [14], which studies the recovery of arterial spin labeling MRI data from the noisy and corrupted observations. In [15] and [12], data collection in wireless sensor networks is studied. The former [15] assumes that the sensor network data is approximately low-rank and has short-term stability in the temporal dimension. The latter [12] further supposes that the sensor network data is correlated in the spatial dimension, that is, the closely located sensors usually have similar measurements. Recovery of traffic matrices is studied in [16], which makes the same assumptions as [12], including low rank, temporal smoothness, and spatial smoothness.

Besides the above mentioned recovery methods, the development of graph signal processing [17], [18], [19], [20], [21] provides a new way to solve spatio-temporal signal recovery problems. Many high-dimensional signals exhibit networked structure that can be expressed by a graph [22], [23], [24], and such signals could be naturally cast as graph signals. Many works have been delivered studying problems of reconstructing graph signals from partial entries [25], [26], [27]. By constructing a graph to connect the observation sites, spatio-temporal signals can be viewed as time-varying graph signals. Requiring the time-varying graph signal to

be strictly bandlimited at each time instant, a distributed reconstruction algorithm is proposed in [20]. In [28] and [29], a Laplacian regularizer is used to favor smooth solutions with respect to the graph topology. By minimizing the graph total variation of the signals at all time instants, a graph signal matrix completion algorithm is proposed in [30]. In [31] a differential-smoothness-based recovery method is proposed applying the property that the temporal differences of the time-varying graph signals exhibit better smoothness with respect to graph than the original signal. However, the aforementioned graph signal processing methods are not designed specifically for spatio-temporal signal recovery problem, and hence do not fully explore the properties of spatio-temporal signals to facilitate the recovery process.

C. Contributions

In this paper, in order to recover spatio-temporal signals with high quality, we take both the low-rank property and local smoothness into consideration. By constructing a graph to encode the pairwise relationship between the observation sites, we novelly introduce the differential smooth prior on time-varying graph signals to characterize the local smoothness of spatio-temporal signals. Then the spatio-temporal signal recovery is formulated as an optimization problem based on low-rank and differential smoothness with respect to the graph topology, which is then solved by the proposed Low Rank and Differential Smoothness-based recovery method (LRDS) as an application of the alternating direction method of multipliers (ADMM) [32]. Theoretical analysis on the recovery accuracy is provided and discussed in detail. When the prior knowledge on the global pattern of the target spatio-temporal signal is available, the proposed reconstruction method is modified to utilize such a priori information to further improve reconstruction performance. Experiments on both synthetic and real-world datasets demonstrate that the proposed method outperforms the state-of-the-art methods.

This paper is organized as follows. In section II, the sampling and recovery of spatio-temporal signals and graph signals are briefly reviewed. In section III, we formulate the spatio-temporal signal recovery problem as a convex optimization problem, and propose LRDS to solve the proposed optimization problem. Performance analysis of the proposed recovery method is included in section IV. In section V, the proposed recovery model is compared with the state-of-the-art models from the aspect of model prior. In section VI, a modified reconstruction method utilizing a priori information on signal's global pattern is provided. The recovery performance of the proposed algorithms is tested and compared with the state-of-the-art methods by experimenting on both synthetic and real-world datasets in section VI [1]. The paper is concluded in section VII.

[1]Simulation codes are provided to reproduce the results presented in this paper: http://ieeexplore.ieee.org and http://gu.ee.tsinghua.edu.cn/publications.

II. Preliminaries

A. Spatio-Temporal Signals Recovery

A spatio-temporal signal can be expressed by a matrix $\mathbf{X} = [\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_{M-1}] \in \mathbb{R}^{N\times M}$, where N is the number of observation sites and M is the number of time instants. The sampled signal can be modeled as

$$\mathbf{Y} = \mathbf{J} \circ \mathbf{X} + \mathbf{V}, \tag{1}$$

where $\mathbf{J} \in \{0,1\}^{N\times M}$ is the sampling operator, $\circ$ is the Hadamard (element-wise) product, and $\mathbf{V}$ denotes the additive Gaussian white noise. Each entry $J_{n,m}$ of matrix $\mathbf{J}$ takes the value of 1 if the nth observation site works at the mth time instant, otherwise it takes the value of 0.

Obviously, exploring some prior knowledge of $\mathbf{X}$ is necessary for signal recovery. It is pointed out in [1] that nearby values tend to be more alike than those far apart, where "nearby" refers to being proximal in both space and time. That is to say, spatio-temporal signals are usually redundant and have a strong correlation. Next, we review the correlated property prevalently existing in spatio-temporal signals.

1) Global Correlation: With the spatial structure of the set of observation sites fixed, the spatio-temporal signals are usually correlated in a global sense [28], [30]. Such global correlation can be demonstrated as the temporal sequence $\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_{M-1}$ generated from limited patterns, or in other words, the spatio-temporal signal $\mathbf{X}$ being approximately low-rank [14], [33].

2) Local Correlation: Besides global correlation, spatio-temporal signals are usually locally correlated [1], [12], [16] as well. To be more specific, the observation of a certain site at a certain time instant is usually strongly correlated with its neighbors in both spatial and temporal directions. These two types of local correlations are stated in detail in Assumptions 1 and 2 respectively.

Assumption 1 (spatial smoothness)**.** *Spatio-temporal signals are smooth in the spatial dimension: at each time instant, observations of geographically nearby observation sites are close to each other.*

Assumption 2 (temporal smoothness)**.** *Spatio-temporal signals are smooth in the temporal dimension: for each observation sites, its observation sequence varies smoothly over time.*

The spatial smooth prior given in Assumption 1 has been utilized in many spatio-temporal recovery works including [12], [20], [28], and [30]. There are also quite a few works, such as [12], [15], [16], and [34], recovering spatio-temporal signals based on the temporal smoothness in Assumption 2 .

B. Graph Signals Recovery

Graph signal sampling and recovery problem is widely studied in the field of graph signal processing, and can be roughly divided into two classes, one focusing on the sampling and recovery of static graph signals, and the other tackling time-varying graph signals. After recalling some basic definitions on graph, we review these two types of problem respectively.

Consider an undirected weighted graph $\mathcal{G}(\mathcal{V},\mathcal{E},\mathbf{W})$, where $\mathcal{V}$ is the set of vertices with $|\mathcal{V}|=N$, $\mathcal{E}$ is the edge set, and the symmetric matrix $\mathbf{W}$ is the weighted adjacency matrix. The (i,j)-th entry of $\mathbf{W}$, $W_{ij}\in\mathbb{R}^+$, denotes the edge weight between the ith and jth vertices is a quantitative expression of their underlying relation, such as similarity, dependency, or communication pattern [30]. The graph Laplacian is defined as $\mathbf{L}=\mathrm{diag}(d_1,\cdots,d_N)-\mathbf{W}$, where $d_i=\sum_{j=1}^{N}W_{ij}$ is the degree of the i-th vertex.

Since $\mathbf{L}$ is symmetric and semi-positive definite, its singular value decomposition can be denoted as $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\mathsf{T}$. With spectra of $\mathbf{L}$ indicating the frequency, the graph Fourier transform of graph signal $\mathbf{x}$ is defined as $\hat{\mathbf{x}}=\mathbf{U}^\mathsf{T}\mathbf{x}$. As a generalization of discrete bandlimited graph to the field of graph signal processing, the concept of bandlimited graph signals first appears in [35]. Specifically, a graph signal $\mathbf{x}$ is named as ω-bandlimited graph signal if $\mathbf{x}\in PW_\omega(\mathcal{G})$, where ω denotes the highest graph frequency component of $\mathbf{x}$, and $PW_\omega(\mathcal{G})$ is the ω-bandlimited subspace of graph $\mathcal{G}$, defined as $\mathrm{span}(\mathbf{u}_1,\cdots,\mathbf{u}_{m_\omega})$, with $\mathbf{u}_1,\cdots,\mathbf{u}_{m_\omega}$ denoting the first few eigenvectors of graph Laplacian $\mathbf{L}$, corresponding to the eigenvalues less or equal to ω.

1) Static Graph Signals: A static graph signal $\mathbf{x}\in\mathbb{R}^N$ is defined as a map on the graph that assigns signal value x_i to the i-th vertex. When the aim is to recover the graph signal with only a part of the vertices sampled, smoothness of graph signals is widely applied as a priori information.

The smoothness of graph signals is a qualitative characteristic that expresses how much signal samples vary with respect to the underlying graph [22], [30]. A typical metric measuring the variation of graph signal is as shown below,

$$S(\mathbf{x}):=\|\mathbf{L}^{\frac{1}{2}}\mathbf{x}\|_2^2=\mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x}=\frac{1}{2}\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{N}_i}W_{ij}\left(x_j-x_i\right)^2, \quad (2)$$

which is a summation of neighborhood variation over all the vertices. The smaller the value of $S(\mathbf{x})$ is, the smoother the graph signal $\mathbf{x}$ is.

Typically, bandlimited graph signals are always smooth, i.e., yielding smaller $S(\mathbf{x})$. The smoothness of ω-bandlimited graph signal $\mathbf{x}$ can be simply tested by $S(\mathbf{x})=\mathbf{x}\mathbf{L}^\mathsf{T}\mathbf{x}\leq\omega\|\mathbf{x}\|^2$.

2) Time-varying Graph Signals: A time-varying graph signal refers to a sequence of static graph signals $\{\mathbf{x}_0,\cdots,\mathbf{x}_T\}$ where $\mathbf{x}_t$ denotes the graph signal at the t-th time instant. Sampling of time-varying graph signals denotes that at each time instant only part of the vertices are sampled. According to when the recovery process is conducted, the recovery of time-varying graph signals can be divided into three types: tracking, batch recovery, and online recovery. Tracking of time-varying graph signals [20] refers to the problem that reconstructing $\mathbf{x}_t$ after the observation to $\mathbf{x}_{t-1}$ is completed, that is, only the observations before the t-th time instant is available for reconstructing $\mathbf{x}_t$. It is an online process and keeps forecasting the signal at the next time instant. Batch recovery of time-varying graph signals [31] refers to the case that reconstruction is conducted after the sampling at all time instants is finished and all the sampled data can be utilized for reconstruction. Online recovery [31] deals with the case that recovering $\mathbf{x}_t$ immediately after the observation to $\mathbf{x}_t$ is available.

In [20], a distributed tracking method is proposed based on both the temporal smooth prior and the smoothness of the signals $\{\mathbf{x}_0,\cdots,\mathbf{x}_T\}$ with respect to the graph topology. Combining the above two types of smoothness together, the work [31] introduces the differential smooth prior, and proposes both batch and online time-varying graph signal recovery algorithms based on it. The differential smoothness is recalled in Assumption 3.

Assumption 3 (differential smoothness)**.** *The temporal differences of spatio-temporal signals are smooth with respect to the graph topology.*

To be noted, the differential smoothness is a local property of time-varying graph signals.

C. Relationship of Spatio-temporal signal and graph signal

Spatio-temporal signals are time-varying graph signals residing on a graph of the observation sites with the edges labelling the geographical adjacency of the observation sites. A typical method to generate such geographical graph is k-nearest neighbor method, that is, an edge is placed between a pair of vertices if and only if one of the two vertices is among the k nearest neighbors of the other. The detailed generating procedure is postponed to Section III-A.

However, since graph topology is not necessarily based on the spatial positions of the vertices, time-varying graph signals are not all spatio-temporal signals.

III. Recovery Method Based on Low Rank and Differential Smoothness (LRDS)

In this section, we propose a spatio-temporal signal recovery method by introducing the differential smooth prior to the field of spatio-temporal signal analysis and jointly applying the global correlated property of spatio-temporal signals. In Subsection A, the recovery problem is formulated as a convex optimization problem. After which, in Subsection B, a solver to the proposed problem, LRDS, is provided by applying ADMM. We also provide complexity analysis to the proposed LRDS in Subsection C.

A. Problem Formulation

As mentioned in Section II-C, a spatio-temporal signal can be read as a time-varying graph signal attached to a graph over the observation sites. Based on such correspondence between spatio-temporal signals and time-varying graph signals, we are inspired to introduce the differential smooth prior of time-varying graph signals proposed in [31] to recover the spatio-temporal signals.

To apply such local prior to recover the missing entries of a spatio-temporal signal $\mathbf{X}=[\mathbf{x}_0,\cdots,\mathbf{x}_M]\in\mathbb{R}^{N\times M}$, we first generate an undirected weighted graph $\mathcal{G}(\mathcal{V},\mathcal{E},\mathbf{W})$ over the set of N observation sites, $\mathcal{V}$, by k-nearest neighbor method to model the geographical adjacency of the observation sites.

Concretely speaking, we first compute the distances between any pair of observation sites and achieve the distance matrix $\mathbf{G}$

with the (i,j)-th element $g_{i,j}$ indicating the distance between sites i and j. To avoid the connection between far away observation sites, a mask matrix $\mathbf{M}$ is generated based on $\mathbf{G}$. For each site i, we find the k smallest elements on the i-th row of $\mathbf{G}$, denoted as $\{g_{i,j}\}_{j\in K_i}$ and set $\{m_{i,j}\}_{j\in K_i}$ as 1 with all other elements in the i-th row of $\mathbf{M}$ as 0. Then the mask matrix is symmetrised by taking $\mathbf{M}$ as $\mathbf{M}^\mathsf{T}|\mathbf{M}$, with $|$ denotes the element-wise "or" operation. There exists an edge between sites i and j if and only if $m_{i,j}$ equals 1. Till now the underlying graph topology is achieved. To further suppress the influence of mal-connections and reflect the geographical adjacency quantitatively, the weight of each edge is set to be inversely proportional to the square of the distance between the connected two vertices, i.e., the (i,j)-th element of the graph adjacency matrix $\mathbf{W}$, $w_{i,j} = m_{i,j}/g_{i,j}^2$.

Let $\mathbf{L}$ denote the Laplacian matrix of this graph. Then we introduce the temporal differential operator $\mathbf{D}$ as follows,

$$\mathbf{D} = \begin{bmatrix} -1 & & \\ 1 & \ddots & \\ & \ddots & -1 \\ & & 1 \end{bmatrix}_{M\times(M-1).} \tag{3}$$

We have the temporal difference of $\mathbf{X}$ equal to $\mathbf{XD} = [\mathbf{x}_1 - \mathbf{x}_0, \mathbf{x}_2 - \mathbf{x}_1, \cdots, \mathbf{x}_M - \mathbf{x}_{M-1}]$. Then another way of saying that $\mathbf{X}$ satisfies Assumption 3 is that the mathematical expression (4) holds true for some positive parameter δ.

$$S(\mathbf{XD}) := \sum_{t=1}^{M} S(\mathbf{x}_t - \mathbf{x}_{t-1}) \overset{\text{(a)}}{=} \operatorname{tr}\left(\mathbf{D}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{L}\mathbf{X}\mathbf{D}\right) \le \delta, \tag{4}$$

where $\operatorname{tr}(\cdot)$ denotes taking the trace of a matrix, and (a) follows from the definition of the smoothness metric $S(\cdot)$ in (2) and (3). To be noted, $S(\mathbf{XD})$ measures the variation of the time difference of $\mathbf{X}$ with respect to the graph topology. Recalling the global correlation of spatio-temporal signals introduced in Section II-A, we have that $\mathbf{X}$ is approximately low-rank. Then jointly applying the global and local correlated prior, the spatio-temporal signal recovery problem can be formulated as an optimization problem

$$\begin{aligned} \min_{\mathbf{X}} \ & \operatorname{rank}(\mathbf{X}) + \lambda \operatorname{tr}\left(\mathbf{D}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{L}\mathbf{X}\mathbf{D}\right) \\ \text{s.t.} \ & \|\mathbf{J}\circ\mathbf{X} - \mathbf{Y}\|_F^2 \le \epsilon, \end{aligned} \tag{5}$$

where λ is the regularization parameter. Because matrix rank minimization is NP-hard, to solve it efficiently, we provide a convex relaxation to (5)

$$\min_{\mathbf{X}} \ \frac{1}{2}\|\mathbf{J}\circ\mathbf{X} - \mathbf{Y}\|_F^2 + \frac{\alpha}{2}\operatorname{tr}\left(\mathbf{D}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{L}\mathbf{X}\mathbf{D}\right) + \mu\|\mathbf{X}\|_*, \tag{6}$$

where α and μ are regularization parameters. In (6), $\operatorname{rank}(\mathbf{X})$ is replaced by nuclear norm $\|\mathbf{X}\|_*$, defined as the sum of the singular values of $\mathbf{X}$, which still promotes low rank [36], [37]. The convexity of (6) follows from the fact that $\|\mathbf{X}\|_*$ is the convex envelope of $\operatorname{rank}(\mathbf{X})$ over set $\{\mathbf{X}\in\mathbb{R}^{N\times M} : \|\mathbf{X}\|_2 \le 1\}$ [37].

Problem (6) is an unconstrained optimization problem with two regularization terms, with $\operatorname{tr}\left(\mathbf{D}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{L}\mathbf{X}\mathbf{D}\right)$ inducing the differential smoothness encoded in Assumption 3, and $\|\mathbf{X}\|_*$ encouraging the global correlated property of spatio-temporal signals. These two terms coincide with each other since they both favor a reconstruction result with approximate bandlimited temporal differences. Besides, they compensate each other and are indispensable. The detailed explanation is as below.

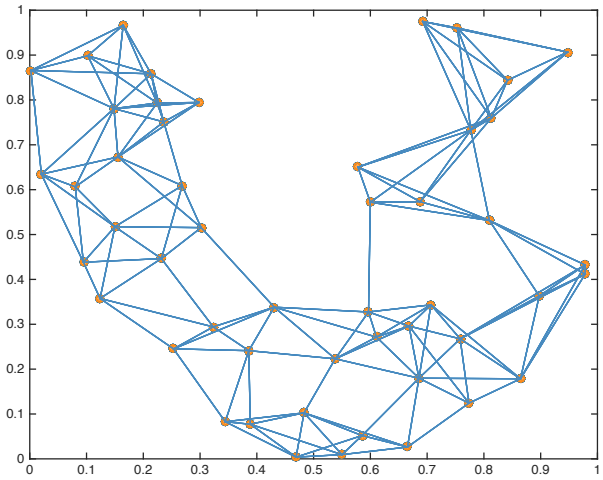


Fig. 1: Graph constructed by 5-nearest neighbor method. The weight of each edge is inversely proportional to the square of the distance between the connected two vertices.

- The regularizer $\operatorname{tr}\left(\mathbf{D}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{L}\mathbf{X}\mathbf{D}\right)$ encodes a priori information on the spatial and temporal structure of $\mathbf{X}$ in graph Laplacian $\mathbf{L}$ and $\mathbf{D}$ respectively. Notice that

$$\operatorname{tr}\left(\mathbf{D}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{L}\mathbf{X}\mathbf{D}\right) = \|\mathbf{L}^{1/2}\mathbf{X}\mathbf{D}\|_F^2 = \|\mathbf{\Lambda}^{1/2}\widehat{\mathbf{X}\mathbf{D}}\|, \tag{7}$$

  where the diagonal matrix $\mathbf{\Lambda}$ collects the graph frequencies as its diagonal elements, and each column of $\widehat{\mathbf{XD}}$ is the graph Fourier transform of the corresponding column of $\mathbf{XD}$. Equation (7) indicates that this regularizer penalizes the high frequency components of the temporal difference $\mathbf{XD}$. Hence it drags the temporal differences of the reconstructed signal toward the 0-frequency. An approximate bandlimited graph signal is favorable.
- With $\|\mathbf{X}\|_*$ favoring a low-rank reconstruction and

$$\operatorname{rank}(\mathbf{X}) - 1 \le \operatorname{rank}(\mathbf{XD}) \le \operatorname{rank}(\mathbf{X}),$$

  the regularizer $\|\mathbf{X}\|_*$ in (6) favors the temporal difference $\mathbf{XD}$ to be low-rank. Although the other regularizer $\operatorname{tr}\left(\mathbf{D}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{L}\mathbf{X}\mathbf{D}\right)$ encourages the energy of $\mathbf{XD}$ concentrating on the relatively low frequency range on graph, it does not emphasize the spectra sparsity of $\mathbf{XD}$. This is where the nuclear norm term compensates. For example, when the columns of $\mathbf{XD}$ resides in or close to the ω-bandlimited subspace $PW_\omega(\mathcal{G})$, low-rank inducing term $\|\mathbf{X}\|_*$ plays an important role.

To test the above intuitive understanding, we generate a spatio-temporal signal with both local and global correlated property and then reconstruct it from its noisy measurement by solving (6) with $\mu > 0$ and $\mu = 0$ respectively. Firstly, we randomly generate 50 points in a 1×1 square area as the observation sites of the spatio-temporal signal. The graph

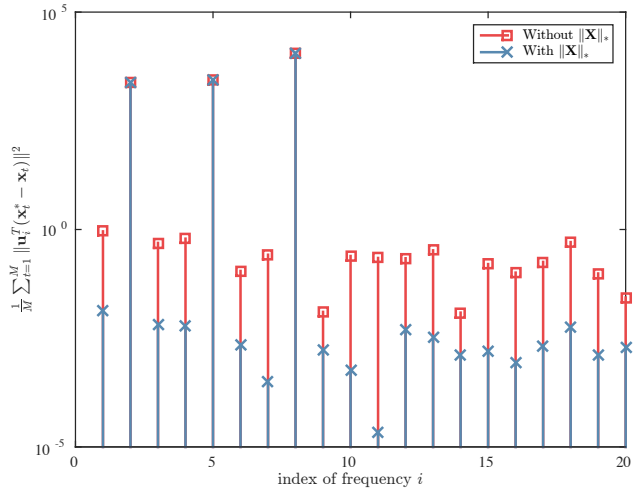


Fig. 2: The average energy at different graph frequencies of the reconstruction error obtained by solving problem (6).

constructed by 5-nearest neighbor is shown in Fig. 1. A 50×200 spatio-temporal signal $\mathbf{X}_0$ on this graph is generated. The generated signal satisfies $\mathbf{X}_0=\tilde{\mathbf{X}}+\mathbf{N}$, where the columns of $\tilde{\mathbf{X}}$ are bandlimited graph signals slowly evolving over time, and $\mathbf{N}$ encodes the perturbation part composed of the whole band signals with relatively low energy. Specifically, both the initial signal of $\tilde{\mathbf{X}}$, $\tilde{\mathbf{x}}_0$, and the difference signals $(\tilde{\mathbf{x}}_t-\tilde{\mathbf{x}}_{t-1})$'s are all smooth graph signal residing on $\mathrm{span}(\mathbf{u}_2,\mathbf{u}_5,\mathbf{u}_8)$, where $\mathbf{u}_2,\mathbf{u}_5$ and $\mathbf{u}_8$ are the eigenvectors of $\mathbf{L}$ corresponding to the 2nd, 5th and 8th smallest eigenvalues respectively. The signal strength satisfies $\|\tilde{\mathbf{x}}_0\|_2=100$, $\|\tilde{\mathbf{x}}_t-\tilde{\mathbf{x}}_{t-1}\|=10, \forall t$. As for the perturbation part, the elements of $\mathbf{N}$ are all identically independent distributed zero-mean Gaussian random variables with 0.01 variance. The sampling set is generated by randomly taking 20 sites for observation in each time instant.

We reconstruct $\mathbf{X}_0$ with $\mu>0$ and $\mu=0$ respectively. In the $\mu=0$ case, the nuclear norm regularizer is muted. Fig. 2 compares the average energy of the reconstruction error $\{\mathbf{x}_t^*-\mathbf{x}_t\}_t$ on the first 20 eigenvectors of Laplacian $\mathbf{L}$. It can be observed that the error components on the main directions $\mathbf{u}_2,\mathbf{u}_5$ and $\mathbf{u}_8$ of both reconstructed results are approximately the same, whereas the reconstruction error by solving optimization problem with $\|\mathbf{X}\|_*$ is smaller than that without as the regularizer $\|\mathbf{X}\|_*$ helps depress the energy on $\mathbf{u}_i$'s with $i\neq 2,5,8$. The above test verifies the correctness of the optimization model (6).

Remark 1. *In real applications, the graph construction process also influences the reconstruction performance of the proposed model. When constructing the base graph by k-nearest neighbors, k's value should be properly tuned. A larger k tends to generating a more densely connected graph, and hence is more likely to include the malconnections. A smaller k is likely to reduce the local correlation. To tackle these problems, a compromise proposal is to apply a moderately large k to emphasize the local correlation and distance dependent edge weights to depress the mutual influence between relatively far away but connected vertices.*

TABLE I: The procedure of solving subproblem (13)

Input: $\mathbf{Y}$, $\mathbf{J}$, $\mathbf{L}$, $\mathbf{Z}^k$, $\mathbf{P}^k$, α,ρ.
Output: $\mathbf{X}_i$: recovered signal
Initialization: $\mathbf{X}_0=0,\ \Delta\mathbf{X}_0=-\nabla f(\mathbf{X}_0),\ i=0$
Repeat:
1) **Stepsize decision:**
$\tau=-\dfrac{\langle\Delta\mathbf{X}_i,\nabla f(\mathbf{X}_i)\rangle}{\langle\Delta\mathbf{X}_i,\nabla f(\mathbf{X}_i)+\mathbf{Y}+\rho\mathbf{Z}^k-\mathbf{P}^k\rangle}$;
2) **Search direction updating:**
$\mathbf{X}_{i+1}=\mathbf{X}_i+\tau\Delta\mathbf{X}_i$;
$\gamma=\|\nabla f(\mathbf{X}_{i+1})\|_F^2/\|\nabla f(\mathbf{X}_i)\|_F^2$;
$\Delta\mathbf{X}_{i+1}=-\nabla f(\mathbf{X}_{i+1})+\gamma\Delta\mathbf{X}_i$;
$i=i+1$;
Until: i reaches maximum number of iterations.

TABLE II: The procedure of LRDS

Input: $\mathbf{Y}$, $\mathbf{J}$, $\mathbf{L}$, α, μ, ρ, stopping criterion.
Output: $\mathbf{X}^k$: recovered signal
Initialization: $\mathbf{X}^0=\mathbf{Y},\ \mathbf{Z}^0=\mathbf{Y},\ \mathbf{P}^0=\mathbf{0},\ k=0$
Repeat:
1) Update $\mathbf{X}^{k+1}$ by (15) or Table I;
2) Update $\mathbf{Z}^{k+1}$ by (17);
3) Update $\mathbf{P}^{k+1}$ by (12);
4) $k=k+1$;
Until: Stopping criterion satisfied.

B. Solver

Noticing that the first two terms of (6) are differentiable, and the third term is proximable, we choose to solve (6) by applying ADMM [32], [38], which takes advantage of both the decomposability and the superior convergence properties of the method of multipliers. Firstly, we introduce an equivalent splitting version of (6) as follows

$$\min_{\mathbf{X},\mathbf{Z}}\ \frac{1}{2}\|\mathbf{J}\circ\mathbf{X}-\mathbf{Y}\|_F^2+\frac{\alpha}{2}\mathrm{tr}\left(\mathbf{D}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{L}\mathbf{X}\mathbf{D}\right)+\mu\|\mathbf{Z}\|_*\quad \text{s.t.}\quad \mathbf{X}=\mathbf{Z}. \tag{8}$$

Such splitting step followed by an augmented Lagrangian method to handle the linear equality constraint is what constitutes ADMM. The augmented Lagrangian of (8) is

$$\mathcal{L}_\rho(\mathbf{X},\mathbf{Z},\mathbf{P})=\frac{1}{2}\|\mathbf{J}\circ\mathbf{X}-\mathbf{Y}\|_F^2+\frac{\alpha}{2}\mathrm{tr}\left(\mathbf{D}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{L}\mathbf{X}\mathbf{D}\right)+\mu\|\mathbf{Z}\|_*+\langle\mathbf{P},\mathbf{X}-\mathbf{Z}\rangle+\frac{\rho}{2}\|\mathbf{X}-\mathbf{Z}\|_F^2, \tag{9}$$

where $\mathbf{P}$ is the Lagrange multiplier, ρ is the penalty parameter, and $\langle\cdot,\cdot\rangle$ denotes the inner product of matrices. As proved in [32] and [39], ADMM finds a saddle point to (9) with the following iterative scheme

$$\mathbf{X}^{k+1}=\arg\min_{\mathbf{X}}\ \mathcal{L}_\rho\left(\mathbf{X},\mathbf{Z}^k,\mathbf{P}^k\right), \tag{10}$$

$$\mathbf{Z}^{k+1}=\arg\min_{\mathbf{Z}}\ \mathcal{L}_\rho\left(\mathbf{X}^{k+1},\mathbf{Z},\mathbf{P}^k\right), \tag{11}$$

$$\mathbf{P}^{k+1}=\mathbf{P}^k+\rho\left(\mathbf{X}^{k+1}-\mathbf{Z}^{k+1}\right). \tag{12}$$

Subproblem (10) is equivalent to

$$\mathbf{X}^{k+1} = \arg\min_{\mathbf{X}} \ \frac{1}{2}\|\mathbf{J}\circ\mathbf{X}-\mathbf{Y}\|_F^2 + \frac{\alpha}{2}\mathrm{tr}\left(\mathbf{D}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{L}\mathbf{X}\mathbf{D}\right) + \frac{\rho}{2}\|\mathbf{X}-\mathbf{Z}^k+\frac{\mathbf{P}^k}{\rho}\|_F^2. \quad (13)$$

Denoting the above objective function as $f(\mathbf{X})$, then its gradient is calculated as

$$\nabla f(\mathbf{X}) = \mathbf{J}\circ\mathbf{X}-\mathbf{Y}+\alpha\mathbf{L}\mathbf{X}\mathbf{D}\mathbf{D}^\mathrm{T}+\rho(\mathbf{X}-\mathbf{Z}^k)+\mathbf{P}^k. \quad (14)$$

By solving the linear equation $\nabla f(\mathbf{X}) = \mathbf{0}$, we achieve the closed-form solution to (13) as in (15).

$$\mathrm{vec}(\mathbf{X}^{k+1}) = \left(\tilde{\mathbf{J}}+\alpha\tilde{\mathbf{L}}\tilde{\mathbf{D}}+\rho\tilde{\mathbf{I}}\right)^{-1}\mathrm{vec}\left(\rho\mathbf{Z}^k+\mathbf{Y}-\mathbf{P}^k\right), \quad (15)$$

where $\mathrm{vec}(\cdot)$ denotes the vectorization operator stacking the columns of a matrix into a long vector, and the matrices with tilde are all square matrices of $MN\times MN$ dimensional. Specifically speaking, $\tilde{\mathbf{I}}$ denotes the MN dimensional identity matrix, $\tilde{\mathbf{J}} = \mathrm{diag}(\mathrm{vec}(\mathbf{J}))$, $\tilde{\mathbf{L}} = \mathbf{I}\otimes\mathbf{L}$, and $\tilde{\mathbf{D}} = \left(\mathbf{D}\mathbf{D}^\mathsf{T}\right)\otimes\mathbf{I}$, with $\mathbf{I}$ denoting the N dimensional identity matrix and $\otimes$ denoting the Kronecker product. To be noted, $\tilde{\mathbf{J}}$, $\tilde{\mathbf{L}}$ and $\tilde{\mathbf{D}}$ are all positive semi-definite matrices.

However, the calculation of (15) involves the inversion of an $MN\times MN$ dimensional matrix, which is computationally expensive. To reduce the computation complexity, conjugate gradient method can be applied to solve (13) iteratively. The detailed algorithm for solving problem (13) is listed in Table I. In each iteration, the algorithm mainly consists of two steps: the determination of stepsize and the update of the next search direction. Denoting the search direction of the ith step as $\Delta\mathbf{X}_i$, the optimal stepsize τ of the ith step is decided by the line minimization rule $\min_\tau \ f(\mathbf{X}_i+\tau\Delta\mathbf{X}_i)$. By taking derivative, we have

$$0 = \frac{\partial f(\mathbf{X}_i+\tau\Delta\mathbf{X}_i)}{\partial\tau} = \langle\Delta\mathbf{X}_i, \nabla f(\mathbf{X}_i+\tau\Delta\mathbf{X}_i)\rangle.$$

Thus we can get the optimal stepsize τ as in Table I.

Subproblem (11) is equivalent to

$$\mathbf{Z}^{k+1} = \arg\min_{\mathbf{Z}} \ \frac{1}{2}\|\mathbf{Z}-\mathbf{X}^{k+1}-\frac{\mathbf{P}^k}{\rho}\|_F^2+\frac{\mu}{\rho}\|\mathbf{Z}\|_*, \quad (16)$$

which has a closed-form solution

$$\mathbf{Z}^{k+1} = \mathrm{SVT}_{\frac{\mu}{\rho}}\left(\mathbf{X}^{k+1}+\frac{\mathbf{P}^k}{\rho}\right), \quad (17)$$

where SVT denotes the singular value thresholding operator,

$$\mathrm{SVT}_\tau(\mathbf{X}) = \mathbf{U}\mathbf{\Lambda}_\tau(\mathbf{\Sigma})\mathbf{V}^\mathrm{T}, \quad (18)$$

where $\mathbf{U},\mathbf{V},\mathbf{\Sigma}$ are from the singular value decomposition of $\mathbf{X}$, i.e., $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathrm{T}$, and

$$\mathbf{\Lambda}_\tau(x) := \mathrm{sign}(x)\max(|x|-\tau, 0) \quad (19)$$

is the soft thresholding operator defined for $\tau\in\mathbb{R}^+$.

The detailed algorithm for solving problem (6) is listed in Table II. The stopping criterion could be either a maximum number of iterations, or the change of $\mathbf{X}^k$ less than a threshold. According to [39] and noticing that the objective function of (8) is closed, proper, and convex, when $\mathbf{X}^k$ is updated with (15), we have that the sequence $\{\mathbf{X}^k,\mathbf{Z}^k,\mathbf{P}^k\}_{k\geq 0}$ generated by LRDS converges to a Karush-Kuhn-Tucker (KKT) point of (9), denoted as $\{\mathbf{X}^*,\mathbf{Z}^*,\mathbf{P}^*\}$.

C. Complexity Analysis

For better understanding the scalability of the proposed LRDS, we provide a brief complexity analysis to it under the asymptotic scenario, where both M and N tend to infinity.

In each iteration of LRDS, $\mathbf{P}$-update involves matrix-matrix summation and matrix-scalar product, and costs $O(MN)$ amount of computational complexity. Compared to $\mathbf{P}$-update, updating $\mathbf{X}$ and $\mathbf{Z}$ is more computation consuming.

Specifically, when updating $\mathbf{X}$ via conjugate gradient method as shown in Table I, the computation is dominated by the gradient computation according to (14). The gradient computation is dominated by the matrix-matrix product $\mathbf{L}\mathbf{X}\mathbf{D}\mathbf{D}^\mathrm{T}$, which consumes $O(N^2M+M^2N)$ floating-point operations (flop). In practice it suffices to compute (13) up to a certain precision, like the machine epsilon. If this precision is considered constant, then the second step takes $O(1/\ln((\sqrt{\kappa(\mathbf{A})}+1)/(\sqrt{\kappa(\mathbf{A})}-1))$ iterations [40], where $\mathbf{A} = \tilde{\mathbf{J}}+\alpha\tilde{\mathbf{L}}\tilde{\mathbf{D}}+\rho\tilde{\mathbf{I}}$ and $\kappa(\mathbf{A})$ denotes its condition number. Then we have that each $\mathbf{X}$-update consumes

$$O\Big(1/\ln\frac{\sqrt{\kappa(\mathbf{A})}+1}{\sqrt{\kappa(\mathbf{A})}-1}\cdot(N^2M+M^2N)\Big) \quad (20)$$

flops. Next, we look into (20) by studying the upper bound on condition number $\kappa(\mathbf{A}) = \sigma_{\max}(\mathbf{A})/\sigma_{\min}(\mathbf{A})$, where $\sigma_{\max}(\cdot)$ and $\sigma_{\min}(\cdot)$ denotes the largest and smallest singular values, respectively.

Proposition 1. *With $d_{\max}$ denoting the largest degree $\max\{d_1,\cdots,d_N\}$ over the set of all N vertices in the graph $\mathcal{G}(\mathcal{V},\mathcal{E},\mathbf{W})$, it holds*

$$\kappa(\mathbf{A}) \leq 1+\frac{1+8\alpha d_{\max}}{\rho}. \quad (21)$$

Proof. We will estimate the smallest and largest singular values of $\mathbf{A}$ separately. Notice that $\tilde{\mathbf{J}}$, $\tilde{\mathbf{L}}$, and $\tilde{\mathbf{D}}$ are all positive semi-definite matrices, and $\alpha,\rho>0$. We have $\sigma_{\min}(\mathbf{A})\geq\rho$. As to the largest one,

$$\sigma_{\max}(\mathbf{A}) \leq \sigma_{\max}(\tilde{\mathbf{J}})+\alpha\sigma_{\max}(\tilde{\mathbf{L}})\sigma_{\max}(\tilde{\mathbf{D}})+\rho. \quad (22)$$

Because $\tilde{\mathbf{J}}$ is a diagonal matrix whose diagonal entries are either 1 or 0, it satisfies $\sigma_{\max}(\tilde{\mathbf{J}})\leq 1$. Besides, according to the definition of $\mathbf{D}$ in (3), it holds

$$\sigma_{\max}(\mathbf{D}) = \|\tilde{\mathbf{D}}\|_2 = \|\mathbf{D}\mathbf{D}^\mathsf{T}\|_2 \leq 4, \quad (23)$$

where the inequality can be simply tested by Gershgorin circle theorem [41]. Again, applying the same theorem on $\mathbf{L} = \mathrm{diag}(d_1,\cdots,d_N)-\mathbf{W}$, where $d_i = \sum_{j=1}^N W_{ij}$ denotes vertex i's degree, we have $\sigma_{\max}(\tilde{\mathbf{L}})\leq 2d_{\max}$. Substituting the upper bounds on $\sigma_{\max}(\tilde{\mathbf{J}}),\sigma_{\max}(\tilde{\mathbf{L}})$ and $\sigma_{\max}(\tilde{\mathbf{D}})$ into (22) and applying $\sigma_{\min}(\mathbf{A})\geq\rho$, we complete the proof. □

Applying the basic inequality $\ln(1+x) \geq x/(1+x), \forall x \geq 0$ and Proposition 1 successively, we have

$$1/\ln \frac{\sqrt{\kappa(\mathbf{A})}+1}{\sqrt{\kappa(\mathbf{A})}-1} \leq \frac{\sqrt{\kappa(\mathbf{A})}+1}{2} \leq 1+\frac{\sqrt{1+8\alpha d_{\max}}}{2\sqrt{\rho}}.$$

By taking α and ρ as constant independent from M and N, we can rephrase the computation consumption of $\mathbf{X}$-update in (20) as

$$O(\sqrt{1+d_{\max}}(N^2M+M^2N)), \tag{24}$$

where the term with $d_{\max}$ is not omitted because it is related to the graph topology and hence probably related to the amount of vertices, N.

As for $\mathbf{Z}$-update, the computation of SVT dominates the computation consumption. Specifically, the matrix-matrix product in (18) consumes $O(\min(MN^2, NM^2))$ flops. To be noted, $\lambda_\tau \leq \min(M,N)$. The SVD of matrix $\mathbf{X}$ also takes the computational cost of $O(\min(MN^2, NM^2))$ flops [42]. Hence, each $\mathbf{Z}$-update consumes $O\Big(\min(MN^2, NM^2)\Big)$ flops.

Jointly applying the above per-update complexity results, we learn that the computation of LRDS is dominated by $\mathbf{X}$-update, and the per-iteration flops of LRDS is (24).

IV. Performance Analysis

In this section, the reconstruction accuracy of the proposed solver is theoretically analyzed. First, a recovery error bounded is provided, which is based on a graph topology and sampling strategy related assumption. Next, this assumption is proved can be easily satisfied and the validity of this error bound is demonstrated. Finally, we discuss about the sampling strategy.

A. Recovery Error Bound

Let $\mathbf{X}$ denote the ground truth spatio-temporal signal to be recovered. We can upper bound the difference of $\mathbf{X}$ and $\mathbf{X}^*$ by Proposition 2.

Proposition 2. *If the following holds true for any $\mathbf{M} \in \mathbb{R}^{N\times M}$ with a constant $\gamma \in [0,1)$,*

$$\|\mathbf{M} - \frac{1}{4}\mathbf{J}\circ\mathbf{M} - \frac{\alpha}{4}\mathbf{L}\mathbf{M}\mathbf{D}\mathbf{D}^\mathsf{T}\|_F \leq \gamma\|\mathbf{M}\|_F, \tag{25}$$

then the recovery error of the proposed method can be upper bounded as shown in (26),

$$\|\mathbf{X}^* - \mathbf{X}\| \leq \frac{\|\mathbf{V}\|_F + \mu\sqrt{\mathrm{rank}(\mathbf{X})} + \alpha\|\mathbf{L}\mathbf{X}\mathbf{D}\mathbf{D}^\mathsf{T}\|_F}{4(1-\gamma)}, \tag{26}$$

where $\mathbf{V}$ is the observation noise in (1).

Proof. Firstly, we upper bound the recovery error

$$\|\mathbf{X}^* - \mathbf{X}\|_F \leq \underbrace{\|\mathbf{X}^* - \mathrm{SVT}_{\frac{\mu}{4}}\left(\mathbf{X} - \frac{\alpha}{4}\mathbf{L}\mathbf{X}\mathbf{D}\mathbf{D}^\mathsf{T}\right)\|_F}_{A} + \underbrace{\|\mathrm{SVT}_{\frac{\mu}{4}}\left(\mathbf{X} - \frac{\alpha}{4}\mathbf{L}\mathbf{X}\mathbf{D}\mathbf{D}^\mathsf{T}\right) - \mathbf{X}\|_F}_{B}. \tag{27}$$

Next, we upper bound terms A and B in (27) respectively. To handle term A, we explore the convergence point $\mathbf{X}^*$. By KKT condition, we have

$$\begin{aligned}\mathbf{X}^* &= \mathbf{Z}^*\\ \mathbf{0} &= \mathbf{J}\circ\mathbf{X}^* - \mathbf{Y} + \alpha\mathbf{L}\mathbf{X}^*\mathbf{D}\mathbf{D}^\mathsf{T} + \mathbf{P}^*\\ \mathbf{0} &\in \mu\partial_{\|\cdot\|_*}(\mathbf{Z}^*) - \mathbf{P}^*.\end{aligned}$$

Jointly applying the above 3 conditions, we have

$$\mathbf{Y} - \mathbf{J}\circ\mathbf{X}^* - \alpha\mathbf{L}\mathbf{X}^*\mathbf{D}\mathbf{D}^\mathsf{T} \in \mu\partial_{\|\cdot\|_*}(\mathbf{X}^*). \tag{28}$$

By (28), it holds

$$\mathbf{X}^* + \frac{1}{4}\left(\mathbf{Y} - \mathbf{J}\circ\mathbf{X}^* - \alpha\mathbf{L}\mathbf{X}^*\mathbf{D}\mathbf{D}^\mathsf{T}\right) \in \frac{\mu}{4}\partial_{\|\cdot\|_*}(\mathbf{X}^*) + \mathbf{X}^*. \tag{29}$$

The reason that we introduce a $1/4$ factor here will be explained in Proposition 3. According to the proof of Theorem 3 of [43], for any $\tau \in \mathbb{R}^+$, we have $\mathrm{SVT}_\tau = (\mathbf{I} + \tau\partial_{\|\|_*})^{-1}$. By (29), with $\tau = \mu/4$, we have

$$\mathbf{X}^* = \mathrm{SVT}_{\frac{\mu}{4}}\left(\mathbf{X}^* - \frac{1}{4}\left(\mathbf{J}\circ\mathbf{X}^* - \mathbf{Y}\right) - \frac{\alpha}{4}\mathbf{L}\mathbf{X}^*\mathbf{D}\mathbf{D}^\mathsf{T}\right). \tag{30}$$

By the equivalent form of the operator SVT_τ in (16), it is a proximal operator of the nuclear norm. The nonexpansiveness of this proximal operator is verified in Lemma 1 [43], that is, for any two matrices $\mathbf{M}_1$ and $\mathbf{M}_2$, and any $\tau \in \mathbb{R}^+$, we have

$$\|\mathrm{SVT}_\tau(\mathbf{M}_1) - \mathrm{SVT}_\tau(\mathbf{M}_2)\|_F \leq \|\mathbf{M}_1 - \mathbf{M}_2\|_F. \tag{31}$$

Substituting $\mathbf{X}^*$ with the right hand side of (30) into term A of (27), and applying (31), we have

$$A \leq \|\mathbf{X}^* - \mathbf{X} - \frac{\alpha}{4}\mathbf{L}(\mathbf{X}^* - \mathbf{X})\mathbf{D}\mathbf{D}^\mathsf{T} - \frac{1}{4}\left(\mathbf{J}\circ\mathbf{X}^* - \mathbf{Y}\right)\|_F.$$

Further by the observation model (1) and the condition in (25), by taking $\mathbf{M}$ as $\mathbf{X}^* - \mathbf{X}$, we have

$$A \leq \gamma\|\mathbf{X}^* - \mathbf{X}\|_F + \frac{1}{4}\|\mathbf{J}\circ\mathbf{X} - \mathbf{Y}\|_F = \gamma\|\mathbf{X}^* - \mathbf{X}\|_F + \frac{1}{4}\|\mathbf{V}\|_F, \tag{32}$$

Then, we consider the term B in (27). It holds

$$\begin{aligned}B \leq& \|\mathrm{SVT}_{\frac{\mu}{4}}\left(\mathbf{X} - \frac{\alpha}{4}\mathbf{L}\mathbf{X}\mathbf{D}\mathbf{D}^\mathsf{T}\right) - \mathbf{X} - \frac{\alpha}{4}\mathbf{L}\mathbf{X}\mathbf{D}\mathbf{D}^\mathsf{T}\|_F\\ &+ \frac{\alpha}{4}\|\mathbf{L}\mathbf{X}\mathbf{D}\mathbf{D}^\mathsf{T}\|_F\\ \overset{(a)}{\leq}& \|\mathbf{\Lambda}_{\frac{\mu}{4}}(\hat{\mathbf{\Sigma}}) - \hat{\mathbf{\Sigma}}\|_F + \frac{\alpha}{4}\|\mathbf{L}\mathbf{X}\mathbf{D}\mathbf{D}^\mathsf{T}\|_F\\ \overset{(b)}{\leq}& \frac{\mu}{4}\sqrt{\mathrm{rank}(\mathbf{X})} + \frac{\alpha}{4}\|\mathbf{L}\mathbf{X}\mathbf{D}\mathbf{D}^\mathsf{T}\|_F,\end{aligned} \tag{33}$$

where the diagonal matrix $\hat{\mathbf{\Sigma}}$ collects the singular values of $\mathbf{X} - \frac{\alpha}{4}\mathbf{L}\mathbf{X}\mathbf{D}\mathbf{D}^\mathsf{T}$ as its diagonal elements, the inequality (a) follows from the definition of SVT_τ in (18) and the fact that Frobenius norm is not affected by the multiplication with orthonormal matrix, and (b) comes from (19). Substituting (32) and (33) into (27) proves (26) □

Remark 2. *The three terms composing the numerator of the right hand side of (26) jointly measures the fidelity of the ground true saptio-temporal data to the proposed model. Concretely speaking, $\|\mathbf{V}\|_F$ measures the noise level of the additive noise $\mathbf{V}$, while $\sqrt{\mathrm{rank}(\mathbf{X})}$ refers to the global*

correlated property of the spatio-temporal signals, and by (34), the term $\|\mathbf{LXDD}^\mathsf{T}\|_F$ measures the violation to the assumed differential smoothness (i.e. Assumption 3).

$$\|\mathbf{LXDD}^\mathsf{T}\|_F^2=\sum_{t=0}^{M}\sum_{i=1}^{N}\sum_{j\in\mathcal{N}_i} W_{ij}^2(2x_{t,j}-x_{t-1,j}-x_{t+1,j})^2, \quad (34)$$

where $x_{t,j}$ denotes the signal value on the j-th vertex at the t-th time instant, and $\mathbf{x}_{-1}=\mathbf{x}_0$, $\mathbf{x}_{M+1}=\mathbf{x}_M$.

B. Successful Recovery Condition

Based on Proposition 2, the reconstruction error bound provided in (26) is only valid when the condition (25) holds. Similar conditions have been considered in [44] and [45]. However, it has not been analyzed whether this condition is easy to be satisfied. Next, we for the first time analyze this condition. Specifically, in Remark 3 we find the bond between parameter γ in (25) and the graph topology labelled by $\mathbf{L}$ and the temporal difference operator $\mathbf{D}$. Proposition 3 provides a sufficient condition on sampling set for (25) holding true. Remark 4 and Proposition 4 further link this condition on sampling set to specific sampling strategies.

For simplicity, we first rephrase this condition in a vector form, i.e., for any matrix $\mathbf{M}\in\mathbb{R}^{N\times M}$, it holds

$$\left\|\left(\tilde{\mathbf{I}}-\frac{1}{4}\tilde{\mathbf{J}}-\frac{\alpha}{4}\tilde{\mathbf{L}}\tilde{\mathbf{D}}\right)\mathrm{vec}(\mathbf{M})\right\|_2\leq\gamma\|\mathrm{vec}(\mathbf{M})\|_2, \quad (35)$$

where the matrix with $\tilde{(\cdot)}$ are all square matrices of $MN\times MN$ dimensional. Specifically, $\tilde{\mathbf{I}}$ is an MN dimensional identity matrix,

$$\tilde{\mathbf{J}}=\mathrm{diag}(\mathrm{vec}(\mathbf{J})),\quad \tilde{\mathbf{L}}=\mathbf{I}\otimes\mathbf{L},\quad \tilde{\mathbf{D}}=(\mathbf{DD}^\mathsf{T})\otimes\mathbf{I},$$

where $\otimes$ denotes Kronecker product. To achieve (35) holding for any matrix $\mathbf{M}$, it suffices to prove that there exists constant $\gamma\in[0,1)$ such that (36) holds,

$$\tilde{\mathbf{I}}\succeq\frac{1}{4}\tilde{\mathbf{J}}+\frac{\alpha}{4}\tilde{\mathbf{L}}\tilde{\mathbf{D}}\succeq(1-\gamma)\tilde{\mathbf{I}}. \quad (36)$$

Remark 3. *We consider (36), the sufficient condition of (25). It can be easily checked that the first relationship in (25) can be guaranteed by upper bounding the parameter $\alpha\leq 3/\|\tilde{\mathbf{L}}\tilde{\mathbf{D}}\|_2$. The second relationship in (25) implies that matrix $\tilde{\mathbf{J}}+\alpha\tilde{\mathbf{L}}\tilde{\mathbf{D}}$ needs to be positive definite. Notice that $\tilde{\mathbf{J}}$, $\tilde{\mathbf{L}}$ and $\tilde{\mathbf{D}}$ are all positive semi-definite matrices, this requires the null spaces of $\tilde{\mathbf{J}}$ and $\tilde{\mathbf{L}}\tilde{\mathbf{D}}$ only intersects at the original point $\mathbf{0}$. Since increasing the sampling positions leads to the shrinkage of the null space of $\tilde{\mathbf{J}}$, a denser sampling set is preferred to have $\tilde{\mathbf{J}}+\alpha\tilde{\mathbf{L}}\tilde{\mathbf{D}}$ being positive definite. Furthermore, when $\tilde{\mathbf{J}}+\alpha\tilde{\mathbf{L}}\tilde{\mathbf{D}}\succ 0$ holds, by simple derivation, we have*

$$1-\gamma=\min\left(\alpha\sigma(\tilde{\mathbf{L}}\tilde{\mathbf{D}}),1\right)/4$$

where $\sigma(\mathbf{A})=\min_{\mathbf{v}\in\mathbb{R}^{MN},\mathbf{Av}\neq 0}\|\mathbf{Av}\|/\|\mathbf{v}\|$ denotes the smallest non-zero singular value of a positive semi-definite matrix $\mathbf{A}$. It holds $\sigma(\tilde{\mathbf{L}}\tilde{\mathbf{D}})\geq\sigma(\mathbf{L})\sigma(\mathbf{DD}^\mathsf{T})$. Then the error bound in (26) can be further relaxed as

$$\|\mathbf{X}^*-\mathbf{X}\|\leq\frac{\|\mathbf{V}\|_F+\mu\sqrt{\mathrm{rank}(\mathbf{X})}+\alpha\|\mathbf{LXDD}^\mathsf{T}\|_F}{\min\left(\alpha\sigma(\mathbf{L})\sigma(\mathbf{DD}^\mathsf{T}),1\right)}.$$

To have a smaller error upper bound, $\sigma(\mathbf{L})$ is preferred to be larger.

For better understanding of (25), we provide an explicit sufficient condition for it.

Proposition 3. *Assume that the generated graph is connected and the regularization parameter α in (6) satisfies $\alpha\in(0,\frac{3}{4\|\mathbf{L}\|_2}]$. Let $S:=\{(n,m):J_{n,m}=1\}$ denote the sampling set. Once there exists a split of $S=S_1\cup S_2, S_1\cap S_2=\phi$, such that the following two criteria hold true:*

1) *for each time instant $m\in\{1,\cdots,M\}$, there exists at least one sampling index of the form $(\cdot,m)$ belonging to S_1;*
2) *for each vertex $n\in\{1,\cdots,N\}$, there exists at least one sampling index of the form $(n,\cdot)$ belonging to S_2,*

then we claim that there exists a constant $\gamma\in[0,1)$ such that (25) is satisfied by any matrix $\mathbf{M}\in\mathbb{R}^{N\times M}$.

Proof. The first relationship in (36) can be checked by the upper bound on α. One has

$$\frac{1}{4}\|\tilde{\mathbf{J}}+\alpha\tilde{\mathbf{L}}\tilde{\mathbf{D}}\|_2\leq\frac{1}{4}\|\mathbf{J}\|_2+\frac{\alpha}{4}\|\mathbf{DD}^\mathsf{T}\|_2\|\mathbf{L}\|_2\leq 1, \quad (37)$$

where the last inequality comes from inequality (23). The inequality (37) implies the first relationship in (36).

To prove the second relationship in (36), we deal with the sampling set S. According to the sampling set S_1 (resp. S_2), we can define a corresponding sampling operator $\mathbf{J}^1$ (resp. $\mathbf{J}^2$), with

$$J^i_{n,m}=\begin{cases}1 & (n,m)\in S_i\\ 0 & (n,m)\notin S_i\end{cases},\forall i\in\{1,2\},$$

and we can define $\tilde{\mathbf{J}}^1$ (resp. $\tilde{\mathbf{J}}^2$) as $\mathrm{diag}(\mathrm{vec}(\mathbf{J}^1))$ (resp. $\mathrm{diag}(\mathrm{vec}(\mathbf{J}^2))$). A simple fact is that $\tilde{\mathbf{J}}=\tilde{\mathbf{J}}^1+\tilde{\mathbf{J}}^2$.

$\tilde{\mathbf{J}}^1$ is positive semi-definite, since it's a diagonal matrix with nonnegative diagonal elements. By Remark 3, we learn that $\tilde{\mathbf{L}}$ is also positive semi-definite. Hence, the matrix $\frac{1}{4}\tilde{\mathbf{J}}^1+\alpha\tilde{\mathbf{L}}$ is also positive semi-definite. To further show that this matrix is positive definite, it suffices to prove that the null spaces of $\tilde{\mathbf{J}}^1$ and $\tilde{\mathbf{L}}$ only intersects at the original point $\mathbf{0}$. Noticing that the null space of Laplacian matrix $\mathbf{L}$ is $\{t\mathbf{1}_N:t\in\mathbb{R}\}$, we have that the null space of $\tilde{\mathbf{L}}$ is composed of piecewise constant vectors, i.e., $\mathcal{K}(\tilde{\mathbf{L}})=\{\mathbf{t}\otimes\mathbf{1}_N:\mathbf{t}\in\mathbb{R}^{M\times 1}\}$, where $\mathcal{K}(\tilde{\mathbf{L}})$ denotes the null space of matrix $\tilde{\mathbf{L}}$. For any $\mathbf{v}\in\mathcal{K}(\tilde{\mathbf{L}})$, it holds

$$\mathbf{v}^\mathsf{T}\tilde{\mathbf{J}}^1\mathbf{v}=\sum_{(n,m)\in S_1}t_m^2\geq\|\mathbf{t}\|^2\geq 0,$$

where the first inequality comes from the first criterion in Proposition 3. $\mathbf{v}^\mathsf{T}\tilde{\mathbf{J}}^1\mathbf{v}=0$ holds if and only if $\mathbf{v}$ is an all zero vector. Hence there exists a constant $\gamma_1\in(0,1/4)$ such that

$$\frac{1}{4}\tilde{\mathbf{J}}^1+\alpha\tilde{\mathbf{L}}\succeq\gamma_1\tilde{\mathbf{I}}. \quad (38)$$

Following the similar deduction, the property of S_2 guarantees that the only intersection of the null spaces of the positive semi-definite matrices $\tilde{\mathbf{D}}$ and $\tilde{\mathbf{J}}^2$ is $\mathbf{0}$. Therefore, there exists a constant $\gamma_2\in(0,1/4)$ such that

$$\frac{1}{4}\tilde{\mathbf{J}}^2+\frac{\gamma_1}{4}\tilde{\mathbf{D}}\succeq\gamma_2\tilde{\mathbf{I}}. \quad (39)$$

Hence, it holds

$$\frac{1}{4}\tilde{\mathbf{J}}+\frac{\alpha}{4}\tilde{\mathbf{L}}\tilde{\mathbf{D}} \overset{(a)}{\succeq} \frac{1}{4}\tilde{\mathbf{J}}^2+\frac{\|\tilde{\mathbf{D}}\|_2}{16}\tilde{\mathbf{J}}^1+\frac{\alpha}{4}\tilde{\mathbf{L}}\tilde{\mathbf{D}} \succeq \frac{1}{4}\tilde{\mathbf{J}}^2+\frac{1}{4}\left(\frac{1}{4}\tilde{\mathbf{J}}^1+\alpha\tilde{\mathbf{L}}\right)\tilde{\mathbf{D}} \overset{(b)}{\succeq} \gamma_2\tilde{\mathbf{I}} \quad (40)$$

where (a) comes from (23), and (b) follows from (38) and (39). To be noted, as the upper bound on $\|\tilde{\mathbf{D}}\|_2$ is 4, to get the splitting in (40), we introduce a factor $1/4$ in (29). By taking γ as $1-\gamma_2$, (36) is proved and hence (35) is satisfied, so is (25). □

Remark 4. *Proposition 3 provides a sufficient but not necessary condition for (25). As shown in Proposition 3, such sufficient condition is tightly related to the sampling strategy* $\mathbf{J}$*. The only constraint on the graph topology is that it is connected. The proposed condition on* $\mathbf{J}$ *can be easily satisfied. Specifically speaking, to sample a spatio-temporal signals with* N *observation sites over* M *time instants, there exists a sampling strategy taking only* $M+N$ *samples that satisfies the condition on* $\mathbf{J}$ *stated in Proposition 3.*

C. Sampling Strategy

Now we consider the simple uniform sampling strategy. With a fixed sampling rate $r \in (0,1)$, at each time instance $m \in \{1,\cdots,M\}$, a different set of $\lceil rN \rceil$ sampling vertices are uniformly chosen from all the N vertices, corresponding to $\lceil rN \rceil$ non-zero elements in each column of $\mathbf{J}$. Such sampling is independent over the temporal dimension. For any time instant $m \in \{1,\cdots,M\}$, each possible sampling state, $J_{:,m}$ with $|J_{:,m}| = \lceil rN \rceil$, occurs with probability $1/\binom{N}{\lceil rN \rceil}$.

Proposition 4. *When uniform sampling strategy with sampling rate* $r \geq 2/N$ *is applied, the probability that the generated sampling set satisfies the two criteria of Proposition 3 is at least*

$$1-N(1+(M-1)r)(1-r)^{M-1}. \quad (41)$$

Proof. The condition $r \geq 2/N$ ensures that there are at least two samples taken at each time instant. For any vertex j, we define the event

$$\mathcal{A}_j = \{\exists t_1 \neq t_2 \in \{1,\cdots,M\},\ \text{s.t. } J_{j,t_1} = J_{j,t_2} = 1\},$$

which refers to the case that the observations at vertex j are sampled at least twice. With uniform sampling, it holds

$$\Pr(\mathcal{A}_j) \geq 1-(1-r)^M - Mr(1-r)^{M-1}. \quad (42)$$

It can be simply verified that once the intersected event $\cap_{m=1}^{N}\mathcal{A}_j$ holds true, we can always find a split of the sampling set satisfying the two requirements of Proposition 3. Hence the probability that the generated sampling set satisfies Proposition 3 is at least

$$1-\sum_{j=1}^{N}(1-\Pr(\mathcal{A}_j)). \quad (43)$$

The proof is accomplished by substituting (42) into (43). □

For better understanding of Proposition 4, we give a quantitative example with $N=20$, $M=40$ and $r=0.2$, which corresponds to the probability lower bound in (41) as 0.9708.

Due to the space limitation, here only uniform sampling is discussed. Via assigning different sampling weights to the set of vertices, importance sampling, also known as leveraged sampling [46], provides more flexibility in choosing sampling vertices. Intuitively, by properly tuning the sampling weights at a certain time instance according to the previous sampling states, one could avoid a vertex being repeatedly sampled and expect a more balanced sampling result. However, it is still an open problem regarding which is the optimal sampling scheme for time-varying graph signal reconstruction problem.

V. Comparison with Related Woks

To draw a clearer picture of the rationality and benefits of the proposed model in (5), we compare it with the state-of-the-art recovery models for spatio-temporal signal and time-varying graph signals.

A. Fixed Point Continuation with Approximate SVD (FPCA)

In light of the global correlated property, the recovery of spatio-temporal signals can be described as low-rank matrix completion problem [47], [36]

$$\min_{\mathbf{X}}\ \text{rank}(\mathbf{X}) \quad \text{s.t.} \quad \|\mathbf{J}\circ\mathbf{X}-\mathbf{Y}\|_F^2 \leq \epsilon, \quad (44)$$

where ϵ controls the noise level. FPCA is proposed in [36] to solve (44). To be noted, the low-rank completion problem is applicable in many problems but not designed specifically for the spatio-temporal signal recovery problem. Therefore the local correlated property of spatio-temporal signals is not involved in this model.

B. Graph signal Matrix Completion via total variation Regularization (GMCR)

In [30], the matrix completion problem is studied in the graph signal processing background. Specifically speaking, the aim is to recover the missing entries of a matrix $\mathbf{X}$, whose columns are all smooth graph signals attached to a same graph $\mathcal{G}(\mathcal{V},\mathcal{E},\mathbf{W})$. GMCR is proposed in [30] to recover $\mathbf{X}$ by solving the optimization problem

$$\min_{\mathbf{X}}\ \mu\|\mathbf{X}\|_* + \lambda S_2(\mathbf{X}) \quad \text{s.t.} \quad \|\mathbf{J}\circ\mathbf{X}-\mathbf{Y}\|_F^2 \leq \epsilon, \quad (45)$$

where μ and λ are the regularization parameters, and similar to $S(\cdot)$ as defined in (2), $S_2(\mathbf{X}) := \|(\mathbf{I}-\frac{\mathbf{W}}{|\lambda_{\max}(\mathbf{W})|})\mathbf{X}\|_F^2$ measures the variation of the graph signals, where $\lambda_{\max}(\mathbf{W})$ is the maximum eigenvalue of adjacency matrix $\mathbf{W}$.

GMCR can also be implemented to recover spatio-temporal signals by constructing graphs of observation sites via k-nearest neighbors method. GMCR is spatio-temporal signal recovery method based on both the global correlation and the spatial smoothness (Assumption 1). As GMCR is not specialized to recover spatio-temporal signals, temporal smoothness as shown in Assumption 2 is not taken into consideration.

C. Spatio-Temporal constrained Low Rank Matrix Approximation (ST-LRMA)

In [12], jointly applying the global correlation and local correlation, the spatio-temporal signal recovery method ST-LRMA is proposed, which recovers the target signal by solving the optimization problem

$$\min_{\mathbf{X}} \ \operatorname{rank}(\mathbf{X}) + \lambda_1 S(\mathbf{X}) + \lambda_2 \|\mathbf{X}\mathbf{D}\|_F^2 \quad \text{s.t.} \quad \|\mathbf{J} \circ \mathbf{X} - \mathbf{Y}\|_F^2 \leq \epsilon, \tag{46}$$

where λ_1 and λ_2 are regularization parameters. With the term $S(\mathbf{X})$ inducing spatial smoothness, and the term $\|\mathbf{X}\mathbf{D}\|_F^2$ inducing the temporal smoothness of the recovered signal, the model implements both Assumptions 1 and 2 to capture local correlated property of spatio-temporal signals.

As a comparison, the proposed model in (5) implements Assumption 3 to capture the local correlated property. Next, we will show that Assumption 3 is satisfied by a wider class of spatio-temporal signals compared to those satisfying Assumption 1 and 2. Firstly, we provide mathematical expressions of Assumptions 1 and 2.

1) If it holds $S(\mathbf{X}) = \|\mathbf{L}^{\frac{1}{2}}\mathbf{X}\|_F^2 \leq \epsilon_1$, we say that $\mathbf{X}$ satisfies Assumption 1 with parameter ϵ_1.
2) If it holds $\|\mathbf{X}\mathbf{D}\|_F^2 \leq \epsilon_2$, we say that $\mathbf{X}$ satisfies Assumption 2 with parameter ϵ_2.

Remark 5. *With the mathematical expression of Assumption 3 given in (4), we have that any spatio-temporal signal satisfying Assumption 1 and 2 with positive parameters ϵ_1 and ϵ_2 respectively also satisfies Assumption 3 with $\epsilon = \min\{4\epsilon_1, 2\epsilon_2 \max_i d_i\}$. Recall that d_i denotes the degree of the i-th vertex.*

Proof. Provided that $\mathbf{X}$ satisfies Assumption 1 with parameter ϵ_1, it holds

$$S(\mathbf{X}\mathbf{D}) = \|\mathbf{L}^{\frac{1}{2}}\mathbf{X}\mathbf{D}\|_F^2 \leq \|\mathbf{D}^\mathsf{T}\|_2^2 \|\mathbf{L}^{\frac{1}{2}}\mathbf{X}\|_F^2 \overset{\text{(a)}}{\leq} \|\mathbf{D}\mathbf{D}^\mathsf{T}\|_2 \epsilon_1 \overset{\text{(b)}}{\leq} 4\epsilon_1,$$

where (a) follows from Assumption 1, and (b) comes from (23). In addition, $\mathbf{X}$ satisfying Assumption 2 with ϵ_2 gives

$$S(\mathbf{X}\mathbf{D}) = \|\mathbf{L}^{\frac{1}{2}}\mathbf{X}\mathbf{D}\|_F^2 \leq \|\mathbf{L}\|_2 \epsilon_2 \leq 2\epsilon_2 \max_i d_i,$$

where the last inequality comes from Gershgorin circle theorem [41]. □

However, the reverse of Remark 5 is not true. That is, there exists spatio-temporal signals that satisfies Assumption 3 but do not satisfy Assumption 1 or Assumption 2. We provide two such examples as follows by simply manipulating a spatio-temporal signal $\mathbf{X}$ that satisfies both Assumptions 1 and 2.

- By adding a non-smooth signal $\mathbf{f}$ to each of its columns, denoted as $\widetilde{\mathbf{x}}_t = \mathbf{x}_t + \mathbf{f}$, the time difference of the new signal is the same as that of the original signal. So $\widetilde{\mathbf{X}}$ still satisfies Assumptions 2 and 3, but does not satisfy Assumption 1 because of the non-smoothness of $\mathbf{f}$.
- Adding a DC signal to $\widetilde{\mathbf{x}}_t$, denoted as $\overline{\mathbf{x}}_t = \widetilde{\mathbf{x}}_t + \delta_t \mathbf{1}_N$, where δ_t is a random variable, and $\mathbf{1}_N$ is a column vector consisting of all ones. As $\|\mathbf{L}^{\frac{1}{2}}\mathbf{1}\|_2 = 0$, $\overline{\mathbf{X}}$ still satisfies Assumption 3, but do not satisfy Assumption 2 because the energy of the time difference can be greatly increased if δ_t takes a large value.

We conclude that compared to applying Assumptions 1 and 2 simultaneously, Assumption 3 is applicable on a wider range of spatio-temporal signals.

D. Batch Reconstruction of Time-Varying Graph Signals (BR-TVGS)

Based on the observation that many time-varying graph signals exhibit differential smoothness (Assumption 3), in [31], the time-varying graph signal reconstruction problem is formulated as

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{J} \circ \mathbf{X} - \mathbf{Y}\|_F^2 + \lambda S(\mathbf{X}\mathbf{D}). \tag{47}$$

As is stated in (4), the regularizer $S(\mathbf{X}\mathbf{D}) = \operatorname{tr}\left(\mathbf{D}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{L}\mathbf{X}\mathbf{D}\right)$ induces the temporal difference of the recovered time-varying graph signal to be smooth with respect to the graph topology. We have analyzed in detail the benefits of applying Assumption 3 in Section V-C. BR-TVGS is proposed in [31] to solve problem (47). But as BR-TVGS is not designed specifically to recover spatio-temporal signals, the global correlated property is not involved in model (47).

VI. Prior LRDS (PLRDS)

In some practical cases, prior information about global pattern of the spatio-temporal signal can be learned or transferred from signals collected at geometrically similar locations and comparable time period. For example, consider that the spatio-temporal signal to be reconstructed denotes the temporal in a specific area recorded through a year. If one possesses the temperature data of the previous year, in the same spatial range, then one can implement principle component analysis to a priori known data. Its principal column space and row space usually carries the transferrable information, and hence can be taken as prior information.

When the global pattern of the spatio-temporal signal is known a priori, we can take advantage of it by applying the idea provided in [48]. Specifically, the global pattern refers to the column space spanned by columns of $\mathbf{U}$ and row space spanned by columns of $\mathbf{V}$ that the spatio-temporal signal is more likely to reside in. Inspired by the matrix completion with prior method in [48], we refine the global correlated property inducing term in (6) and formulate the problem as

$$\min_{\mathbf{X}} \ \frac{1}{2}\|\mathbf{J} \circ \mathbf{X} - \mathbf{Y}\|_F^2 + \frac{\alpha}{2}\operatorname{tr}\left(\mathbf{D}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{L}\mathbf{X}\mathbf{D}\right) + \mu\|\mathbf{Q}_1\mathbf{X}\mathbf{Q}_2\|_*, \tag{48}$$

where $\mathbf{Q}_1 = \mathbf{I}_N - w_1\mathbf{U}\mathbf{U}^\mathsf{T}$ and $\mathbf{Q}_2 = \mathbf{I}_M - w_2\mathbf{V}\mathbf{V}^\mathsf{T}$, with parameters $w_1, w_2 \in (0, 1)$. Intuitively, the new regularizer penalizes the contradiction to the prior global pattern. The parameters w_1 and w_2 reflect the confidence on the prior knowledge. When $w_1 = w_2 = 0$, problem (48) reduces to the originally proposed problem (6).

Problem (48) can be solved by applying ADMM on the augmented Lagrangian problem

$$\tilde{\mathcal{L}}_\rho(\mathbf{X}, \mathbf{Z}, \mathbf{P}) = \frac{1}{2}\|\mathbf{J} \circ \mathbf{X} - \mathbf{Y}\|_F^2 + \frac{\alpha}{2}\operatorname{tr}\left(\mathbf{D}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{L}\mathbf{X}\mathbf{D}\right) + \mu\|\mathbf{Z}\|_* + \langle \mathbf{P}, \mathbf{Q_1}\mathbf{X}\mathbf{Q_2} - \mathbf{Z}\rangle + \frac{\rho}{2}\|\mathbf{Q}_1\mathbf{X}\mathbf{Q}_2 - \mathbf{Z}\|_F^2, \tag{49}$$

Noticing the similarity between (9) and (49), we arrive at a solver to problem (48) by simply modifying LRDS proposed in Section III-B. Concretely speaking, the same as LRDS in Table II, we iteratively update $\mathbf{X}^k, \mathbf{Z}^k$ and $\mathbf{P}^k$.

When updating $\mathbf{X}^k$, we need to solve the subproblem

$$\mathbf{X}^{k+1} = \arg\min_{\mathbf{X}} \ \tilde{\mathcal{L}}_\rho \left(\mathbf{X}, \mathbf{Z}^k, \mathbf{P}^k\right),$$

which equivalents to

$$\mathbf{X}^{k+1} = \arg\min_{\mathbf{X}} \ \frac{1}{2}\|\mathbf{J} \circ \mathbf{X} - \mathbf{Y}\|_F^2 + \frac{\alpha}{2}\mathrm{tr}\left(\mathbf{D}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{L}\mathbf{X}\mathbf{D}\right) + \frac{\rho}{2}\|\mathbf{Q}_1\mathbf{X}\mathbf{Q}_2 - \mathbf{Z}^k + \frac{\mathbf{P}^k}{\rho}\|_F^2.$$

Denote the objective function as $\tilde{f}$. This subproblem can be efficiently solved by implementing the procedure in Table I, with all f replaced by $\tilde{f}$, and the stepsize decision step replaced by

$$\tau = \frac{-\left\langle \Delta\mathbf{X}_i, \nabla\tilde{f}\left(\mathbf{X}_i\right)\right\rangle}{\left\langle \Delta\mathbf{X}_i, \nabla\tilde{f}\left(\Delta\mathbf{X}_i\right) + \mathbf{Y} + \rho\mathbf{Q}_1\left(\mathbf{Z}^k - \frac{\mathbf{P}^k}{\rho}\right)\mathbf{Q}_2\right\rangle}.$$

The update equation of $\mathbf{Z}^k$ and $\mathbf{P}^k$ can be simply achieved by replacing $\mathbf{X}^{k+1}$ with $\mathbf{Q}_1\mathbf{X}^{k+1}\mathbf{Q}_2$ in (17) and (12), respectively. We refer this modified LRDS method as prior LRDS (PLRDS).

VII. Experiments

To test the recovery performance of the proposed methods, we conduct experiments on one synthetic dataset and four real-world datasets, including the sea surface temperature data from Earth System Research Laboratory [49], the global sea-level pressure data from Joint Institute for the Study of the Atmosphere and Ocean [50], the temperature data of the sensor network from Intel Berkeley Research Lab [51], and the daily mean PM2.5 Concentration of California from US Environmental Protection Agency [52].

The proposed methods LRDS and PLRDS are compared with Natural Neighbor Interpolation (NNI) [53], FPCA [36], GMCR [30], ST-LRMA [12], and BR-TVGS [31]. NNI only utilizes the local correlated property of spatio-temporal signals. The models of the other compared methods are reviewed in Section V.

In each experiment, an undirected weighted graph is constructed by k-nearest neighbor method to characterize the geographical adjacency of the observation sites. In order to achieve a connected graph with few mal-connections, we heuristically take $k = 5$. The mean absolute error (MAE) is used to evaluate the recovery result, denoted as $\mathrm{MAE} = \|\mathbf{x}^* - \mathbf{x}\|_1/N_x$, where $\mathbf{x}$ is the ground-true signal, $\mathbf{x}^*$ is the recovered signal, and N_x is the length of the signals. We adopt uniformly random sampling scheme to select vertices. At each time slot, sampled vertices are randomly chosen following certain sampling rate[2]. 50 independently random sampling patterns are tested to give a convincing comparison result, and the MAE of each time instant is an average of 50 tests.

[2]For real-world datasets, the sampled vertices are constrained to be among those with valid data.

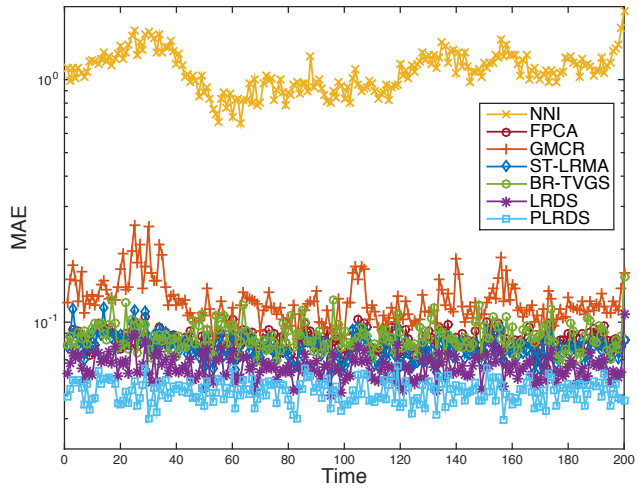


Fig. 3: Synthetic data $\mathbf{X}_1$: MAEs versus time given by different methods under the sampling rate of 40%.

A. Experiments on synthetic data

A 50×200 spatio-temporal signal $\mathbf{X}_1$ on the graph shown in Fig. 1 is generated. The generated signal satisfies $\mathbf{X}_1 = \tilde{\mathbf{X}}+\mathbf{N}$, where $\tilde{\mathbf{X}}$ enjoys both spatial and temporal smoothness, and $\mathbf{N}$ denotes the perturbation part. Specifically, the initial signal of $\tilde{\mathbf{X}}$, $\tilde{\mathbf{x}}_0$, is a smooth graph signal residing on $\mathcal{S}_3$, which is the eigenspace corresponding to the 3 smallest eigenvalues of the graph Laplacian $\mathbf{L}$, and the difference signals $\tilde{\mathbf{x}}_t - \tilde{\mathbf{x}}_{t-1}$'s are all smooth graph signals residing on the same subspace $\mathcal{S}_3$. The signal strength satisfies $\|\tilde{\mathbf{x}}_0\|_2 = 100$, $\|\tilde{\mathbf{x}}_t - \tilde{\mathbf{x}}_{t-1}\| = 10, \forall t$. As for the perturbation part, the elements of $\mathbf{N}$ are all identically independent distributed zero-mean Gaussian random variables with 0.01 variance.

As the strength of the perturbation term $\mathbf{N}$ is relatively small comparing to the unperturbed term $\tilde{\mathbf{X}}$, the generated spatio-temporal signal $\mathbf{X}_1$ is approximately low-rank, and each column of $\mathbf{X}_1$ is smooth with respect to the underlying graph topology, so are the temporal difference signals. Hence, we have $\mathbf{X}_1$ satisfies all the three assumptions.

When the sampling rate is 40%, the recovery errors of all the methods measured by MAE per time instant are displayed in Fig. 3. The average MAEs over time of the comparison methods NNI, FPCA, GMCR, ST-LRMA, BR-TVGS, and the proposed LRDS are 1.095, 0.084, 0.123, 0.079, 0.088 and 0.065 respectively. By taking $\mathbf{U}^\mathsf{T}\mathbf{U}$ as the projection matrix of $\mathcal{S}_3$ and $\mathbf{V}\mathbf{V}^\mathsf{T}$ as $\mathbf{1}^\mathsf{T}\mathbf{1}/M$ which implies the temporal smooth property, we apply the PLRDS method and achieve the average MAE 0.052. As can be seen, compared with NNI, the other recovery methods show great improvement on recovery accuracy, which is typically because the proper modeling of global and local correlation. Meanwhile, the proposed LRDS method achieves better performance than the other methods other than PLRDS. By properly utilizing the prior information on the column space of $\mathbf{X}$, PLRDS exhibits the best reconstruction performance.

Next we change signal $\mathbf{X}_1$ as follows:

TABLE III: Assumptions that the synthetic signals satisfied

Signals	Assumption 1	Assumption 2	Assumption 3
$\mathbf{X}_1$	√	√	√
$\mathbf{X}_2$	×	√	√
$\mathbf{X}_3$	√	×	√
$\mathbf{X}_4$	×	×	√

TABLE IV: Recovery error comparison

	NNI	FPCA	GMCR	ST-LRMA	BR-TVGS	LRDS	PLRDS
$\mathbf{X}_1$	1.10	0.084	0.122	0.079	0.088	**0.065**	*0.052*
$\mathbf{X}_2$	6.12	0.086	0.287	0.134	0.088	**0.068**	*0.052*
$\mathbf{X}_3$	4.55	0.084	0.140	0.119	0.087	**0.065**	*0.052*
$\mathbf{X}_4$	9.25	0.086	1.371	0.572	0.087	**0.067**	*0.052*

1) Add an identical non-smooth signal to each column of $\mathbf{X}_1$, getting signal $\mathbf{X}_2$. The non-smooth signal is of strength 100 and resides on the orthogonal complementary subspace of $\mathcal{S}_3$.
2) Add a DC signal with random amplitude to each column of $\mathbf{X}_1$, getting signal $\mathbf{X}_3$. The amplitude follows zero-mean Gaussian distribution, with variance equal to 400.
3) Carry out the above two operations at the same time, getting signal $\mathbf{X}_4$.

The properties that these 4 synthetic signals exhibit are compared in Table III.

When the sampling rate is 40%, the average MAEs over time of all methods with respect to spatio-temporal signals $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$ and $\mathbf{X}_4$ are listed in Table IV. For GMCR, because signals $\mathbf{X}_2$ and $\mathbf{X}_4$ do not satisfy Assumption 1 any more, the term $S_2(\mathbf{X})$ in model (45) is not proper any more, so when recovering $\mathbf{X}_2$ and $\mathbf{X}_4$, the performance of GMCR is not as good as recovering $\mathbf{X}_1$ and $\mathbf{X}_3$. For ST-LRMA, because signal $\mathbf{X}_4$ does not satisfy Assumption 1 or 2 any more, the terms $S(\mathbf{X})$ and $\|\mathbf{XD}\|_F^2$ in model (46) are improper, so its performance on recovering $\mathbf{X}_4$ is not competitive. Among the methods not implementing a priori information $\mathcal{S}_3$, the proposed LRDS achieves the best recovery accuracy when recovering any of the four signals. The improvement of LRDS compared with BR-TVGS is due to the exploitation of global correlation, i.e., low rank. The improvement of LRDS over GMCR, FPCA and ST-LRMA comes from the proper modeling of local correlation, which also shows the benefits of applying differential smooth prior (Assumption 3). By properly implementing a priori information of $\mathcal{S}_3$, PLRDS shows better performance than LRDS.

B. The sea surface temperature dataset

The sea surface temperature data [49] is collected monthly from 1870 to 2014 with a spatial resolution of 1° latitude ×1° longitude global grid. We randomly select 100 locations on the Pacific for simulation, with the observing duration of 500 months. The selected data ranges from 0.02°C to 30.72°C, and the mean temperature is 19.14°C. Due to the limited knowledge of global pattern in many real-world applications, PLRDS is not tested in this and the following experiments on real-world datasets.

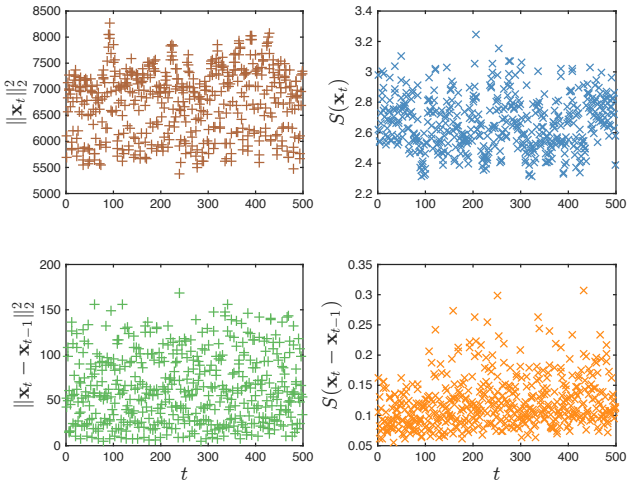


Fig. 4: The sea surface temperature dataset: energy and spatial smoothness comparison of the original signal and the temporal difference signal.

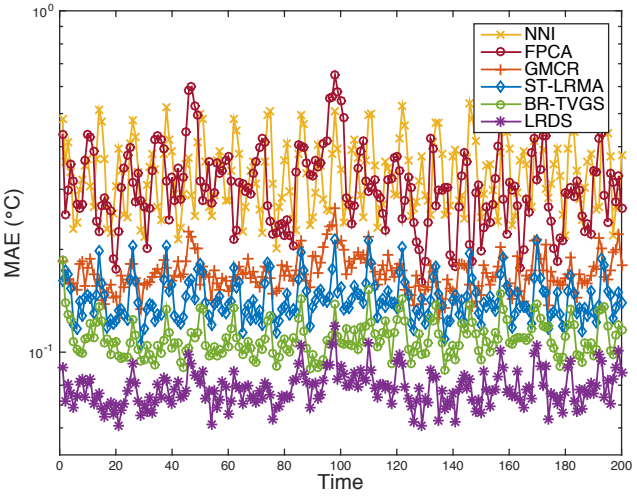


Fig. 5: The sea surface temperature dataset: MAEs versus time given by different methods under the sampling rate of 40%.

Fig. 4 shows the smoothness comparison of the original signal and the temporal difference signal calculated by columns. The left column compares the energy of the original signal and the temporal difference signal, whose averages are 6.77×10^3 and 63.20 respectively, where the DC component has been removed from the original signal. We can see that the energy of the temporal difference signal is much smaller than the original signal, so the original signal satisfies temporal smooth prior. The right column compares the spatial smoothness measured by $S(\cdot)$ of the original signal and the temporal difference signal, whose averages are 2.66 and 0.12 respectively. It shows that the spatial smoothness of the temporal difference signal is more obvious than the original signal.

Firstly, we set the sampling rate as 40%. The MAEs of all the methods are displayed in Fig. 5, and only the first 200 time instants are drawn to facilitate display. The average MAEs over time of the comparison methods NNI, FPCA, GMCR, ST-LRMA, BR-TVGS, and the proposed LRDS are

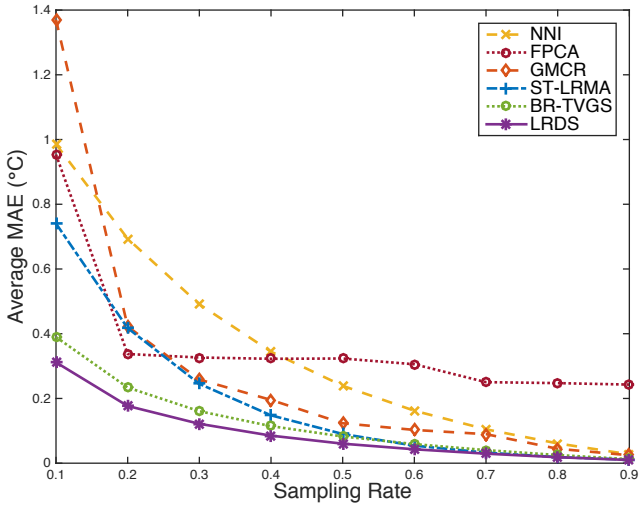


Fig. 6: The sea surface temperature dataset: the average MAEs under different sampling rates.

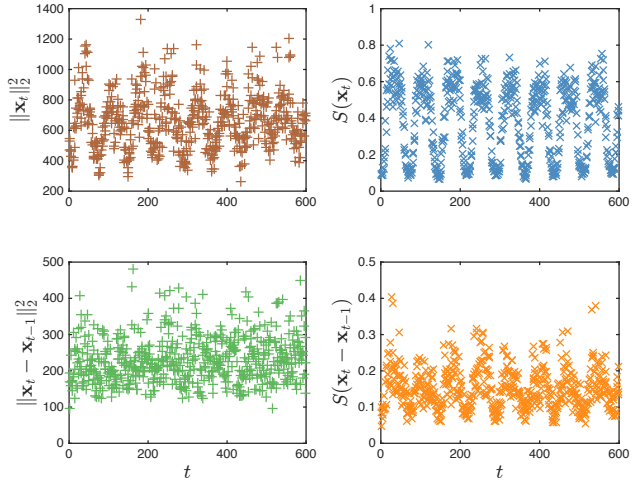


Fig. 7: The global sea-level pressure dataset: energy and spatial smoothness comparison of the original signal and the temporal difference signal.

0.34, 0.32, 0.18, 0.15, 0.12 and 0.081, respectively. As can be seen, the performance of the proposed method is better than all references. Then we test the performance of all the methods under different sampling rates, and the results are displayed in Fig. 6. One can read from the results that the recovery error of all the methods decreases with the increasing of sampling rate, and the proposed method is always superior to the others.

C. The global sea-level pressure dataset

The global sea-level pressure data [50] is collected from 1948 to 2010. The spatial resolution is 2.5° latitude $\times$2.5° longitude global grid, and the temporal resolution is five days. We randomly select 500 locations worldwide over a time period of 600. The selected data ranges from 96.22kPa to 110.06kPa, and the average is 101.22kPa.

Fig. 7 shows the energy and smoothness comparison of the original signal and the temporal difference signal calculated by

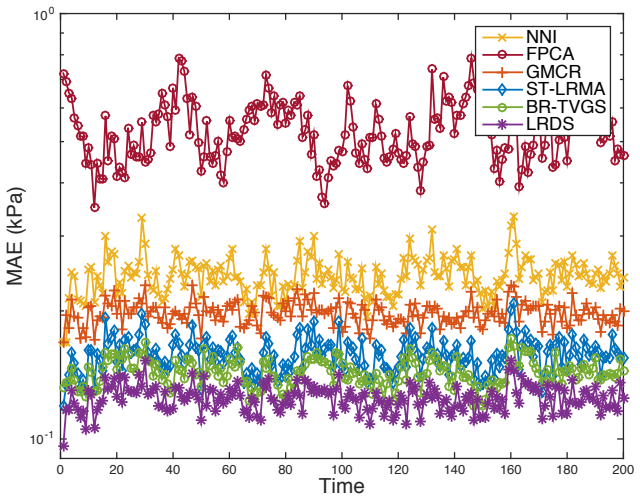


Fig. 8: The global sea-level pressure dataset: MAEs versus time given by different methods under the sampling rate of 40%.

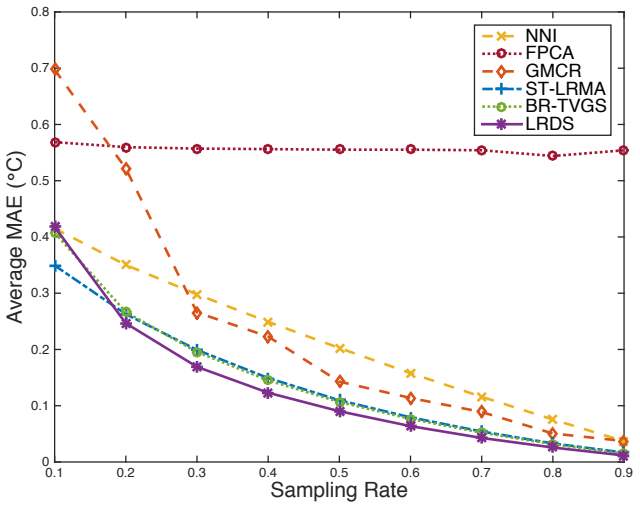


Fig. 9: The global sea-level pressure dataset: the average MAEs under different sampling rates.

columns. The left column compares the energy the original signal and the temporal difference signal, whose averages are 6.63×10^2 and 2.30×10^2 respectively, where the DC component has been removed from the original signal. We can see that the energy of the temporal difference signal is smaller than the original signal, so the original signal is smooth in the temporal direction to some extend. The right column compares the $S(\cdot)$ value of the original signal and the temporal difference signal, whose averages are 0.39 and 0.15 respectively. It can be seen that the spatial smoothness of the temporal difference signal is more obvious than the original signal.

When the sampling rate is 40%, the MAEs of all the methods are displayed in Fig. 8, and only the first 200 time instants are drawn to facilitate display. The average MAEs over time of the comparison methods NNI, FPCA, GMCR, ST-LRMA, BR-TVGS, and the proposed LRDS are 0.25, 0.55, 0.20, 0.16, 0.15 and 0.13, respectively. The average MAEs for

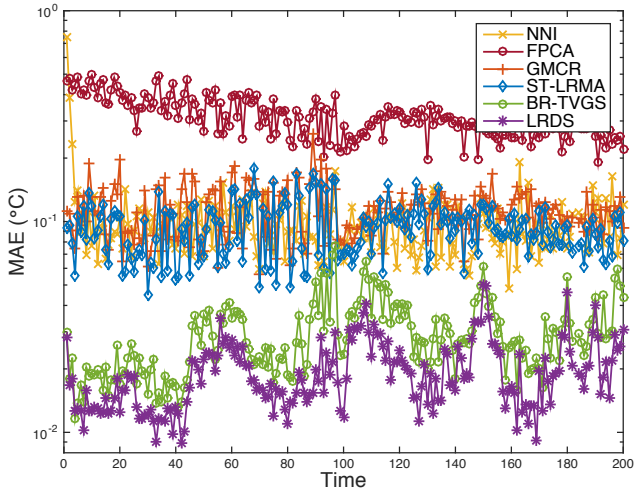


Fig. 10: The temperature sensor network dataset: MAEs versus time given by different methods under the sampling rate of 40%.

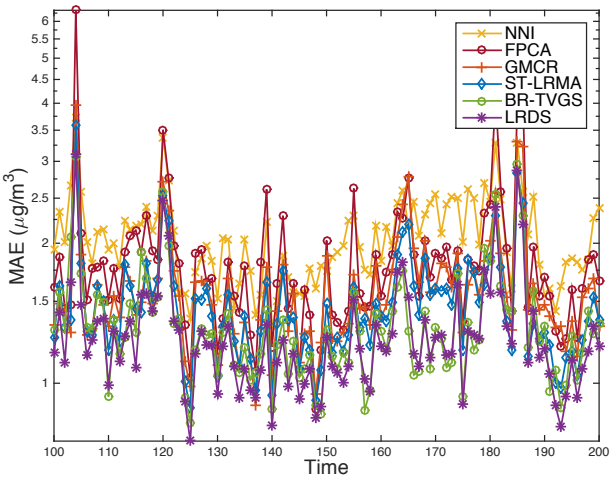


Fig. 12: Mean reconstruction error for the last 100 moments on daily mean PM2.5 concentration data, where the mean is taken over vertices.

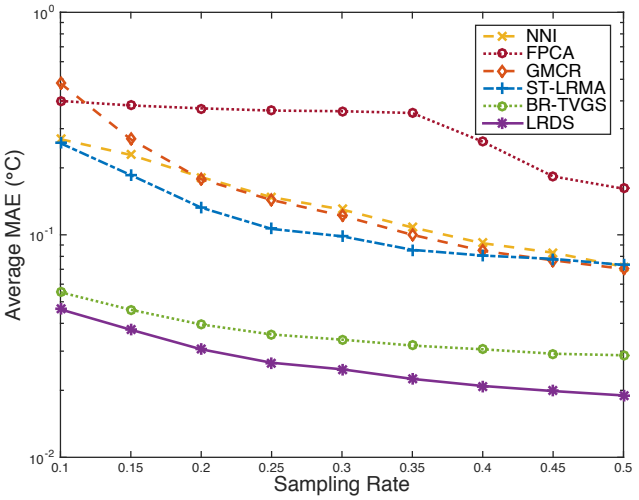


Fig. 11: The temperature sensor network dataset: the average MAEs under different sampling rates.

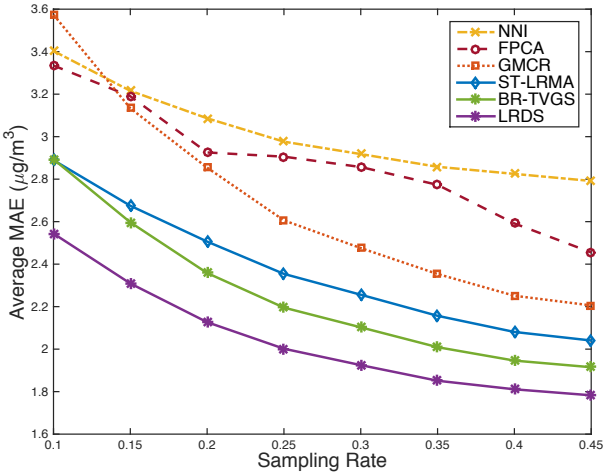


Fig. 13: Mean reconstruction error under different sampling rates on daily mean PM2.5 concentration data, where the mean is taken over both time, all the vertices and 50 independent trials.

all the methods under different sampling rates are displayed in Fig. 9. We can draw a similar conclusion as the previous experiment that the proposed method LRDS is superior to the others except for 0.1 sampling rate.

D. The temperature sensor network dataset

The temperature data of the sensor network is collected by 54 sensors distributed in the laboratory [51] every 30 seconds from February 28th, 2004, and the data between 09:26 and 13:36 on February 28th, 2004 is selected for simulation. The dimension of the selected data is 53×500, excluding one fully damaged sensor. There are missing values due to various reasons such as sensor malfunctions or communication failure. The valid data ranges from $14.01°\mathrm{C}$ to $28.0°\mathrm{C}$.

In this experiment, sampling is among the valid data, and the recovery error is evaluated among the unsampled valid data. When the sampling rate is 40%, the MAEs of all the compared methods are displayed in Fig. 10, and only the first 200 time instants are drawn to facilitate display. The average MAEs over time of the comparison methods NNI, FPCA, GMCR, ST-LRMA, BR-TVGS, and the proposed LRDS are 0.090, 0.26, 0.093, 0.080, 0.029 and 0.021, respectively. The average MAEs for all the methods under different sampling rates are displayed in Fig. 11. These results show that the proposed method generally exhibits better performance than the comparison methods on this temporal sensor network dataset.

E. The daily mean PM2.5 concentration dataset

The California daily mean PM2.5 concentration data [52] are collected from 93 observation sites over 200 days starting from January 1, 2015, and the size of the data is 93×200. Not all of these sites collected valid data everyday, and the

percentages of valid data collected each day roughly range from 90 to 45. The valid data range from 0.1μg/m^3 to 102.7μg/m^3.

In this experiment, sampling is among the valid data, and the reconstruction error is evaluated among the unsampled valid data. When the sampling rate is 40%, the MAEs of all the methods are displayed in Fig. 12, and only the first 100 time instants are drawn to facilitate the display. The average MAEs over time of the comparison methods NNI, FPCA, GMCR, ST-LRMA, BR-TVGS, and the proposed LRDS are 2.74, 2.50, 2.14, 2.00, 1.86, and 1.74, respectively. The average MAEs for all the methods under different sampling rates are displayed in Fig. 13. These results show that the proposed LRDS outperforms the compared methods on this dataset.

VIII. Conclusions

This paper studies the spatio-temporal signal recovery problem. By exploring the low-rank property of the original signal, as well as the smoothness of the temporal difference signal with respect to the graph characterizing the geographical adjacency of the observation sites, we formulate the spatio-temporal signal recovery problem as an unconstrained convex optimization problem. A new spatio-temporal signal recovery method, LRDS, is proposed via applying the alternating direction method of multipliers to solve the proposed optimization problem. Theoretical analysis on the recovery performance is provided. By modifying LRDS, we propose PLRDS to utilize the possible a priori information on the global pattern. Experiments on synthetic datasets test both LRDS and PLRDS, and verify their correctness and performance improvement over the state-of-the-art spatio-temporal signal recovery methods. Experiments on four different real-world datasets demonstrate that LRDS outperforms all the compared methods.

References

[1] N. Cressie and C. K. Wikle, *Statistics for spatio-temporal data*. John Wiley & Sons, 2011.

[2] J. M. Xu, A. Bhargava, R. Nowak, and X. Zhu, "Socioscope: Spatio-temporal signal recovery from social media," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2012, pp. 644–659.

[3] C. Sengstock, M. Gertz, F. Flatow, and H. Abdelhaq, "A probablistic model for spatio-temporal signal extraction from social media," in *ACM Sigspatial International Conference on Advances in Geographic Information Systems*, 2013, pp. 274–283.

[4] H. Wang, "Principal component analysis on temporal-spatial variations of sea level anomalies from t/p satellite altimeter data over the northwest pacific," in *Gravity, Geoid and Geodynamics 2000*. Springer, 2001, pp. 165–170.

[5] T. M. Smith and R. W. Reynolds, "Improved extended reconstruction of SST (1854-1997)," *Journal of climate*, vol. 17, no. 12, pp. 2466–2477, 2004.

[6] C. Olives, L. Sheppard, J. Lindstrm, P. D. Sampson, J. D. Kaufman, and A. A. Szpiro, "Reduced-rank spatio-temporal modeling of air pollution concentrations in the multi-ethnic study of atherosclerosis and air pollution," *Ann Appl Stat.*, vol. 8, no. 4, pp. 2509–2537, 2015.

[7] A. M. Mosammam and J. T. Kent, "Estimation and testing for covariance-spectral spatial-temporal models," *Environmental and ecological statistics*, vol. 23, no. 1, pp. 43–64, 2016.

[8] H. Wang, F. Nie, and H. Huang, "Low-rank tensor completion with spatio-temporal consistency," in *AAAI*, 2014, pp. 2846–2852.

[9] C. Ulas, P. A. Gómez, J. I. Sperl, C. Preibisch, and B. H. Menze, "Spatio-temporal mri reconstruction by enforcing local and global regularity via dynamic total variation and nuclear norm minimization," in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*. IEEE, 2016, pp. 306–309.

[10] M. Balazinska, A. Deshpande, M. J. Franklin, P. B. Gibbons, J. Gray, M. Hansen, M. Liebhold, S. Nath, A. Szalay, and V. Tao, "Data management in the worldwide sensor web," *IEEE Pervasive Computing*, vol. 6, no. 2, 2007.

[11] L. Kong, M. Xia, X. Y. Liu, M. Wu, and X. Liu, "Data loss and reconstruction in sensor networks," in *INFOCOM, 2013 Proceedings IEEE*, 2013, pp. 1654–1662.

[12] X. Piao, Y. Hu, Y. Sun, B. Yin, and J. Gao, "Correlated spatio-temporal data collection in wireless sensor networks based on low rank matrix approximation and optimized node sampling," *Sensors*, vol. 14, no. 12, pp. 23 137–23 158, 2014.

[13] K. Wang, J. Xia, C. Li, L. V. Wang, and M. A. Anastasio, "Low-rank matrix estimation-based spatio-temporal image reconstruction for dynamic photoacoustic computed tomography," in *SPIE BiOS*. International Society for Optics and Photonics, 2014, pp. 89 432I–89 432I.

[14] R. Fang, J. Huang, and W. M. Luh, "A spatio-temporal low-rank total variation approach for denoising arterial spin labeling mri data," in *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*. IEEE, 2015, pp. 498–502.

[15] J. Cheng, Q. Ye, H. Jiang, D. Wang, and C. Wang, "STCDG: An efficient data gathering algorithm based on matrix completion for wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 2, pp. 850–861, 2013.

[16] Y. Zhang, M. Roughan, W. Willinger, and L. Qiu, "Spatio-temporal compressive sensing and internet traffic matrices," in *ACM SIGCOMM 2009 Conference on Data Communication*, 2009, pp. 267–278.

[17] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process*, vol. 61, no. 7, pp. 1644–1656, 2013.

[18] X. Wang, P. Liu, and Y. Gu, "Local-set-based graph signal reconstruction," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2432–2444, 2015.

[19] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6510–6523, 2015.

[20] X. Wang, M. Wang, and Y. Gu, "A distributed tracking algorithm for reconstruction of graph signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 4, pp. 728–740, 2015.

[21] P. Liu, X. Wang, and Y. Gu, "Coarsening graph signal with spectral invariance," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1070–1074.

[22] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, 2013.

[23] H. E. Egilmez, E. Pavez, and A. Ortega, "Graph learning from data under structural and laplacian constraints," *arXiv preprint arXiv:1611.05181*, 2016.

[24] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology identification from spectral templates," in *Statistical Signal Processing Workshop (SSP), 2016 IEEE*. IEEE, 2016, pp. 1–5.

[25] I. Pesenson, "Sampling in paley-wiener spaces on combinatorial graphs," *Trans. Amer. Math. Soc.*, vol. 360, no. 10, pp. 5603–5627, 2008.

[26] X. Wang, J. Chen, and Y. Gu, "Local measurement and reconstruction for noisy bandlimited graph signals," *Signal Processing*, vol. 129, no. 5, pp. 119–129, 2015.

[27] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Sampling of graph signals with successive local aggregations," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1832–1843, 2016.

[28] M. T. Bahadori, Q. R. Yu, and Y. Liu, "Fast multivariate spatio-temporal analysis via low rank tensor learning," in *Advances in neural information processing systems*, 2014, pp. 3491–3499.

[29] Y. Liu, "Scalable multivariate time-series models for climate informatics," *Computing in Science & Engineering*, vol. 17, no. 6, pp. 19–26, 2015.

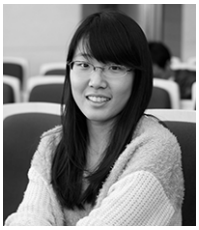
[30] S. Chen, A. Sandryhaila, J. M. Moura, and J. Kovačević, "Signal recovery on graphs: Variation minimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 17, pp. 4609–4624, 2015.

[31] K. Qiu, X. Mao, X. Shen, X. Wang, T. Li, and Y. Gu, "Time-varying graph signal reconstruction," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 870–883, 2017.

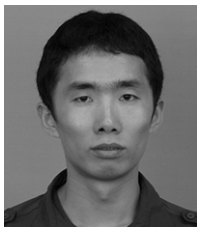
[32] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[33] M. E. El-Telbany and M. A. Maged, "Exploiting sparsity in wireless sensor networks for energy saving: A comparative study," *International*

Journal of Applied Engineering Research, vol. 12, no. 4, pp. 452–460, 2017.
[34] X. Piao, Y. Hu, Y. Sun, and B. Yin, "Efficient data gathering in wireless sensor networks based on low rank approximation," in *Green Computing and Communications*, 2013, pp. 699–706.
[35] I. Pesenson, "Sampling in paley-wiener spaces on combinatorial graphs," *Transactions of the American Mathematical Society*, vol. 360, no. 10, pp. 5603–5627, 2008.
[36] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and bregman iterative methods for matrix rank minimization," *Mathematical Programming*, vol. 128, no. 1-2, pp. 321–353, 2011.
[37] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, PhD thesis, Stanford University, 2002.
[38] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers & Mathematics with Applications*, vol. 2, no. 1, pp. 17–40, 1976.
[39] J. Eckstein and W. Yao, "Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives," *Pac. J. Optim.*, vol. 11, no. 4, pp. 619–644, 2015.
[40] W. Hackbusch, *Iterative solution of large sparse systems of equations*. Springer, 1994, vol. 95.
[41] H. E. Bell, "Gershgorin's theorem and the zeros of polynomials," *The American Mathematical Monthly*, vol. 72, no. 3, pp. 292–295, 1965.
[42] L. N. Trefethen and D. Bau III, *Numerical linear algebra*. Siam, 1997, vol. 50.
[43] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and bregman iterative methods for matrix rank minimization," *Mathematical Programming*, vol. 128, no. 1, pp. 321–353, 2011.
[44] P. Giménez-Febrer and A. Pages-Zamora, "Matrix completion of noisy graph signals via proximal gradient minimization," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 4441–4445.
[45] K. Yi, J. Wan, T. Bao, and L. Yao, "A dct regularized matrix completion algorithm for energy efficient data gathering in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 11, no. 7, p. 272761, 2015.
[46] A. Eftekhari, M. B. Wakin, and R. A. Ward, "Mcˆ 2: A two-phase algorithm for leveraged matrix completion," *Information and Inference: A Journal of the IMA*, 2016.
[47] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.
[48] A. Eftekhari, D. Yang, and M. B. Wakin, "Weighted matrix completion and recovery with prior subspace information," *IEEE Transactions on Information Theory*, 2018.
[49] "Sea surface temperature (sst) v2," Dec. 18, 2015. [Online]. Available: http://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html
[50] "Sea-level pressure, 1948 - 2010," May 25, 2016. [Online]. Available: http://research.jisao.washington.edu/data_sets/reanalysis/
[51] "Sensor network data," March 30, 2016. [Online]. Available: http://www.cs.cmu.edu/~guestrin/Research/Data/
[52] "Air quality data," https://www.epa.gov/outdoor-air-quality-data, May 19, 2016.
[53] S. Robert, "A brief description of natural neighbour interpolation," *Interpreting multivariate data*, vol. 21, pp. 21–36, 1981.

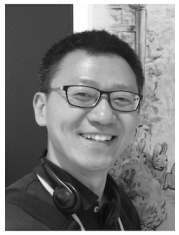


Xianghui Mao received the B.E. degree with honor from the Electronic Engineering Department, Tsinghua University, Beijing, China, in 2014. She is currently working towards the Ph.D. degree in Electronic Engineering Department, Tsinghua University, and her advisor is Professor Yuantao Gu. She was a Visiting Graduate Researcher at University of California, Los Angeles from 2017.03 to 2017.09. Her recent research interests include optimization related topics and graph signal processing.



Kai Qiu received the Ph.D. degree from the Electronic Engineering Department, Tsinghua University, Beijing, China, in 2017. He is currently an associate researcher at Academy of Military Sciences PLA China, Beijing, China. His research interests include Magnetic resonance imaging and graph signal processing.

Tiejian Li received the Ph.D. degree from Tsinghua University, Beijing, China, in 2008. He is currently an associate professor at Tsinghua University and an adjunct professor at Qinghai University. His research interests are in the fields of river basin modeling, parallel computing in hydrological models, and data mining algorithms in hydroscience.



Yuantao Gu (S'02-M'03-SM'17) received the B.E. degree from Xi'an Jiaotong University in 1998, and the Ph.D. degree with honor from Tsinghua University in 2003, both in Electronic Engineering. He joined the faculty of Tsinghua University in 2003 and is now a Tenured Associate Professor with Department of Electronic Engineering. He was a visiting scientist at Microsoft Research Asia during 2005 to 2006, Research Laboratory of Electronics at Massachusetts Institute of Technology during 2012 to 2013, and Department of Electrical Engineering and Computer Science at the University of Michigan in Ann Arbor during 2015. His research interests include high-dimensional statistics, sparse signal recovery, temporal-space and graph signal processing, related topics in wireless communications and information networks.

He has been an Associate Editor of the IEEE Transactions on Signal Processing since 2015, a Handling Editor for EURASIP Digital Signal Processing from 2015 to 2017, and an Elected Member of the IEEE Signal Processing Theory and Methods (SPTM) Technical Committee since 2017. He is a senior member of IEEE.

He received the Best Paper Award of IEEE Global Conference on Signal and Information Processing (GlobalSIP) in 2015, the Award for Best Presentation of Journal Paper of IEEE International Conference on Signal and Information Processing (ChinaSIP) in 2015, and Zhang Si-Ying (CCDC) Outstanding Youth Paper Award (with his student) in 2017.