

# Laplacian Convolutional Representation for Traffic Time Series Imputation

Xinyu Chen, Zhanhong Cheng, HanQin Cai, Nicolas Saunier, Lijun Sun, *Senior Member, IEEE*

**Abstract**—Spatiotemporal traffic data imputation is of great significance in intelligent transportation systems and data-driven decision-making processes. To perform efficient learning and accurate reconstruction from partially observed traffic data, we assert the importance of characterizing both global and local trends in time series. In the literature, substantial works have demonstrated the effectiveness of utilizing the low-rank property of traffic data by matrix/tensor completion models. In this study, we first introduce a Laplacian kernel to temporal regularization for characterizing local trends in traffic time series, which can be formulated as a circular convolution. Then, we develop a low-rank Laplacian convolutional representation (LCR) model by putting the circulant matrix nuclear norm and the Laplacian kernelized temporal regularization together, which is proved to meet a unified framework that has a fast Fourier transform (FFT) solution in log-linear time complexity. Through extensive experiments on several traffic datasets, we demonstrate the superiority of LCR over several baseline models for imputing traffic time series of various time series behaviors (e.g., data noises and strong/weak periodicity) and reconstructing sparse speed fields of vehicular traffic flow. The proposed LCR model is also an efficient solution to large-scale traffic data imputation over the existing imputation models.

**Index Terms**—Spatiotemporal traffic data, time series imputation, low-rank models, circulant matrix nuclear norm, Laplacian kernelized regularization, circular convolution, discrete Fourier transform, fast Fourier transform

## 1 INTRODUCTION

MISSING data imputation is a fundamental component to a wide range of applications in intelligent transportation systems (ITS), including route planning, travel time estimation, and traffic flow forecasting. Typically, traffic data can be collected by sensors (e.g., loop detectors and video cameras) on a continuous basis, producing a sequence of traffic flow time series such as speed and volume measurements. However, the real-world ITS often suffers from various operational issues such as sensor failure and network communication disorder, leading to data corruption and sparsity. Making accurate recovery of these data is vital for supporting ITS, but it still demands appropriate imputation approaches.

The basic modeling idea of missing data imputation in traffic time series is to exploit complicated spatial and temporal correlations/dependencies from partial observations, consequently leading to extensive data-driven approaches such as low-rank models [1] and deep learning methods [2], [3]. Typically, traffic flow data always show strong global and local trends with long- and short-term patterns [4]. The global trends usually refer to as certain periodic and

cyclical patterns, which can be well characterized by low-rank models. Unfortunately, conventional low-rank models fail to characterize the time series dynamics because the reconstruction of low-rank models (e.g., low-rank matrix completion (LRMC) [5], [6]) is invariant to the permutation of rows and columns. Thus, recent studies presented low-rank time series completion models based on certain algebraic structures, including Hankel matrices/tensors [7], [8], [9], circulant matrices/tensors [10], [11], and convolution matrices [11], [12]. The sequential dependencies are implicitly captured by these structures when characterizing the low-rank property of time series. However, low-rank Hankel/convolution models are always limited to small- or middle-scale problems due to the large size of algebraic structures. Although circulant matrix nuclear norm minimization (CircNNM) can be efficiently solved through the fast Fourier transform (FFT), circulant matrices are restricted and fail to characterize the local trends of time series [11].

The default structure of most low-rank models—rank minimization whose objective is essentially converted into the nuclear norm minimization and takes a singular value thresholding [5], [6]—does not ensure local smoothness. Thus, it requires us to model both global and local trends in a unified framework. In the literature, there are several ways to characterize the local spatial and temporal dependencies in data-driven machine learning models. For example, on the spatial dimension, Laplacian regularization has become a standard technique to impose local consistency (see e.g., [13], [14]). On the temporal dimension, the local smoothness is often characterized using time series smoothing and autoregression explicitly (see e.g., [15], [16], [17]). Considering the importance of temporal regularization in regulating the behavior of global low-rank models, we are inspired to develop a tailored regularization for CircNNM to reinforce

- X. Chen and N. Saunier are with the Civil, Geological and Mining Engineering Department, Polytechnique Montreal, Montreal, QC H3T 1J4, Canada. E-mail: chenxy346@gmail.com (X. Chen), nicolas.saunier@polymtl.ca (N. Saunier).
- Z. Cheng and L. Sun are with the Department of Civil Engineering, McGill University, Montreal, QC H3A 0C3, Canada. E-mail: zhanhong.cheng@mail.mcgill.ca (Z. Cheng), lijun.sun@mcgill.ca (L. Sun).
- H. Cai is with the Department of Statistics and Data Science and Department of Computer Science, University of Central Florida, Orlando, FL 32816, USA. E-mail: hqcai@ucf.edu

(Corresponding author: Nicolas Saunier)

local trends while maintaining the algorithm's efficiency through FFT. Thus, using the fact that the Laplacian matrix of a circulant graph is a circulant matrix, we first introduce a novel temporal regularization to CircNNM. Next, we develop an efficient algorithm to solve the proposed Laplacian convolutional representation (LCR) using the alternating direction method of multipliers (ADMM). The contribution of this work is three-fold:

- We introduce a circular Laplacian kernel and use it to define a temporal regularization for characterizing the local trends in time series. By doing so, the temporal regularization can be formulated with circular convolution and draw the connection with FFT.
- We propose a low-rank completion model—LCR—by characterizing global trends of sparse traffic time series as the nuclear norm of a circulant matrix and modeling local trends by the temporal regularization simultaneously. According to the properties of the circulant matrix and circular convolution, we present a fast implementation of LCR through FFT.
- We empirically verify the importance of temporal regularization in LCR. Further, due to the fast implementation via FFT with log-linear time complexity, LCR is scalable to large imputation problems. The experimental results demonstrate that LCR performs better than the state-of-the-art baseline methods in terms of both accuracy and efficiency.

In practice, there is often a significant gap between accurate estimation and efficient implementation in the substantial imputation methods due to methodological challenges. For instance, while introducing algebraic structures [7], [10], [12], [11] or spatiotemporal smoothness [18], [16] usually enhances accuracy, these approaches often come with a significantly higher computational cost. To address this gap, this work advances low-rank completion methods in the following ways. (i) The basic modeling idea of LCR stems from CircNNM [11], [12], but LCR reinforces local time series trends with temporal regularization. (ii) LCR advances convolution nuclear norm minimization (ConvNNM, capable of global/local trend modeling depending on the kernel size) [11], [12] with a flexible time series modeling mechanism. (iii) LCR follows an efficient FFT implementation as CircNNM, while FFT cannot be used in ConvNNM (see e.g., Fig. 3(a) for the computational cost). (iv) The flipping operation (see Fig. 7) in LCR addresses the issue caused by the correlation between the start and end data points of the time series.

The remainder of this paper is structured as follows. Section 2 and 3 introduce the related work and some basic concepts, respectively. In Section 4, we integrate the temporal regularization into low-rank models for characterizing both global and local trends in traffic time series. Section 5 and Section 6 conduct imputation experiments on several real-world datasets. Finally, we conclude this study in Section 7.

## 2 RELATED WORK

### 2.1 Low-Rank Completion with Algebraic Structures

Recent studies show great interest in time series completion with certain algebraic structures, e.g., the Hankel and circulant matrices. These approaches overcome some critical

limitations of pure LRMC models, e.g., (i) LRMC is incapable of handling the entire row/column missing, (ii) LRMC is invariant to the permutation of rows/columns, and (iii) LRMC is not applicable to the case of univariate time series. For example, the model developed by Yokota *et al.* [7] can recover the missing slices of tensor using Hankel structure. Sedighin *et al.* [19] applied the tensor train decomposition to tensors obtained by multi-way Hankel structures and found better completion performance. In both studies, the low-rank methods on Hankel structures—replicating the data with certain sliding rules—are well-suited to modeling spatiotemporal dependencies and learning from sparse data.

A critical property of a circulant matrix is that its nuclear norm can be efficiently obtained via FFT. Using this property, Yamamoto *et al.* [10] proposed a fast tensor completion method; Liu and Zhang [11] used the nuclear norm minimization of circulant matrices for missing data recovery and time series forecasting. Despite the fast algorithm for the circulant matrix, circulant-matrix-based models are inadequate in capturing the local trend/continuity in time series. Therefore, the ConvNNM model shows better local trends modeling if the kernel size of the convolution matrix is set as a relatively small value [11]. Further, Liu [12] proposed a learnable and orthonormal transformation for ConvNNM to reinforce its modeling ability when the convolutional low-rankness condition is not fully satisfied.

### 2.2 Imputation with Temporal Modeling

In the literature, considerable research has leveraged temporal dynamics in low-rank models for time series imputation. A common assumption among these models is that time series and their low-rank factors have local dependencies. Chen and Cichocki [20] proposed a Toeplitz-matrix-based regularization to impose temporal smoothness in matrix factorization; the regularizer handles the difference between the low-rank factors of consecutive times. A similar regularizer based on the Toeplitz matrix was used by [21] for traffic data reconstruction. Chen *et al.* [22] applied a quadratic variation (QV) to a traffic tensor completion problem to ensure temporal smoothness. For modeling low-dimensional temporal dynamics, Xiong *et al.* [15] formulated a Bayesian tensor factorization with first-order Markovian assumptions on the temporal factors. Yu *et al.* [16] developed a temporal autoregressive regularizer in matrix factorization. Chen *et al.* [23] developed a low-rank autoregressive tensor completion model for traffic data imputation. While these two works assume the univariate autoregression, Chen and Sun [1] applied a vector autoregression on the latent temporal factors and developed a fully Bayesian model for multidimensional and sparse time series prediction. As Laplacian regularization is of broad use in graph modeling, it is also applicable to temporal modeling. For example, Rao *et al.* [24] proposed a matrix completion algorithm with Laplacian regularization. To the best of our knowledge, we are the first to present a Laplacian kernelized temporal regularization with circular convolution which consequently leads to the use of FFT.

## 3 PRELIMINARIES

In this section, we introduce the basic definitions of the circulant matrix, convolution matrix, and circular convolution,

in the meanwhile summarizing their relationships.

### 3.1 Circulant Matrix

The circulant matrix is an important structure that shows broad use in the field of signal processing [25], [26]. By definition, for any vector  $\mathbf{x} = (x_1, x_2, \dots, x_T)^\top \in \mathbb{R}^T$ , the circulant matrix can be written as follows,

$$\mathcal{C}(\mathbf{x}) \triangleq \begin{bmatrix} x_1 & x_T & x_{T-1} & \cdots & x_2 \\ x_2 & x_1 & x_T & \cdots & x_3 \\ x_3 & x_2 & x_1 & \cdots & x_4 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_T & x_{T-1} & x_{T-2} & \cdots & x_1 \end{bmatrix} \in \mathbb{R}^{T \times T}, \quad (1)$$

where  $\mathcal{C} : \mathbb{R}^T \rightarrow \mathbb{R}^{T \times T}$  denotes the circulant operator. The first column of  $\mathcal{C}(\mathbf{x})$  is the vector  $\mathbf{x}$ , and the diagonal entries of  $\mathcal{C}(\mathbf{x})$  are all equal to  $x_1$ .

### 3.2 Convolution Matrix

Convolution is vital to a variety of machine learning problems. By definition, for any vectors  $\mathbf{x} = (x_1, x_2, \dots, x_T)^\top \in \mathbb{R}^T$  and  $\mathbf{y} = (y_1, y_2, \dots, y_{\tilde{T}})^\top \in \mathbb{R}^{\tilde{T}}$  with  $\tilde{T} \leq T$ , the circular convolution of two vectors is  $\mathbf{z} = \mathbf{x} \star \mathbf{y} \in \mathbb{R}^T$  [27], denoting the operator with the symbol  $\star$ ; element-wise, we have

$$z_t = \sum_{k=1}^{\tilde{T}} x_{t-k+1} y_k, \quad \forall t \in \{1, 2, \dots, T\}, \quad (2)$$

where  $z_t$  is the  $t$ th entry of  $\mathbf{z}$  and  $x_{t-k+1} = x_{t-k+1+T}$  for  $t+1 \leq k$ . In particular, circular convolution is a linear operator that can be expressed  $\mathbf{x} \star \mathbf{y} \equiv \mathcal{C}_{\tilde{T}}(\mathbf{x})\mathbf{y}$  where  $\mathcal{C}_{\tilde{T}} : \mathbb{R}^T \rightarrow \mathbb{R}^{T \times \tilde{T}}$  denotes the convolution operator with kernel size  $\tilde{T}$ . The resultant convolution matrix consists of the first  $\tilde{T}$  columns of the circulant matrix  $\mathcal{C}(\mathbf{x})$  [11], [12]. Given any vectors  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^T$  of the same length, then we have  $\mathbf{x} \star \mathbf{y} \equiv \mathcal{C}(\mathbf{x})\mathbf{y}$ .

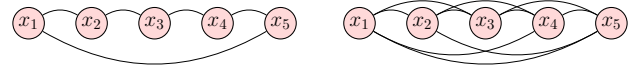
## 4 METHODOLOGY

In this section, we introduce an efficient LCR model for imputing sparse traffic time series. To resolve the optimization problem, we seek an FFT implementation in the frequency domain within the ADMM, instead of the time domain.

### 4.1 Laplacian Kernel

Laplacian matrix is a classical structure for representing the links among nodes in a graph. In this work, we extract the temporal dependencies of time series through undirected and circulant graphs. Recall that the Laplacian matrix by definition takes  $\mathbf{L} = \mathbf{D} - \mathbf{A}$  in which  $\mathbf{D}$  and  $\mathbf{A}$  are the (diagonal) degree matrix and adjacency matrix, respectively [28]. In Fig. 1, the Laplacian matrices of both graphs are circulant matrices, and their first columns are  $\boldsymbol{\ell} = (2, -1, 0, 0, -1)^\top$  and  $\boldsymbol{\ell} = (4, -1, -1, -1, -1)^\top$ , respectively.

In this work, we introduce a Laplacian kernel as described in Definition 1, allowing one to characterize temporal dependencies of time series. In aforementioned cases as shown in Fig. 1, the first column of the Laplacian matrix  $\mathbf{L}$  is indeed a simple example of Laplacian kernel.



(a) Circulant graph with degree 2. (b) circulant graph with degree 4.

Fig. 1: Undirected and circulant graphs on the relational data samples  $\{x_1, x_2, \dots, x_5\}$  with certain degrees.

**Definition 1** (Laplacian Kernel). *Given any time series  $\mathbf{x} \in \mathbb{R}^T$ , suppose  $\tau \in \mathbb{Z}^+$  ( $\tau \leq \frac{1}{2}(T-1)$ ) be the kernel size of an undirected and circulant graph, then the Laplacian kernel is defined as*

$$\boldsymbol{\ell} \triangleq (2\tau, \underbrace{-1, \dots, -1}_{\tau}, 0, \dots, 0, \underbrace{-1, \dots, -1}_{\tau})^\top \in \mathbb{R}^T, \quad (3)$$

which is also the first column of the Laplacian matrix and the inherent degree matrix is diagonalized with entries  $2\tau$ .

**Remark 1.** *The circulant operation assumes that the start data points and the end data points are connected, which is a disadvantage in real-world data analysis. To overcome this issue, on any time series  $\mathbf{x} \in \mathbb{R}^T$ , one can construct the following vector:*

$$\mathbf{x}_{\text{new}} = \begin{bmatrix} \mathbf{x} \\ \mathbf{J}_T \mathbf{x} \end{bmatrix} = (x_1, \dots, x_T, x_T, \dots, x_1)^\top \in \mathbb{R}^{2T}, \quad (4)$$

where  $\mathbf{J}_T \in \mathbb{R}^{T \times T}$  is the exchange matrix whose antidiagonal entries are one and other entries are zero.

Typically, the temporal regularization calculates how values of  $\mathbf{x}$  differ from their adjacent values, and can thus be used as a regularization of local temporal smoothness. According to the definitions of Laplacian kernel and circulant matrix and the relationship between circular convolution and circulant matrix, we declare the following form:

$$\mathcal{R}_\tau(\mathbf{x}) = \frac{1}{2} \|\mathbf{L}\mathbf{x}\|_2^2 = \frac{1}{2} \|\mathcal{C}(\boldsymbol{\ell})\mathbf{x}\|_2^2 = \frac{1}{2} \|\boldsymbol{\ell} \star \mathbf{x}\|_2^2. \quad (5)$$

where  $\|\cdot\|_2$  denotes the  $\ell_2$ -norm of a vector. As can be seen, the Laplacian kernel can represent the graphical relationship in the Laplacian matrix, showing no need for constructing the Laplacian matrix anymore. The setting of the degree  $2\tau$  depends on the strength of local dependencies and the missing data scenarios (e.g., missing rate).

**Remark 2.** *Definition 1 allows one to obtain a more flexible design of the kernel  $\boldsymbol{\ell}$ . For example, if we introduce a directed Laplacian kernel in the form of random walk [15], [13] (i.e.,  $\boldsymbol{\ell} = (1, 0, \dots, 0, -1)^\top \in \mathbb{R}^T$ ), then the temporal regularization is equivalent to the QV regularization, namely,*

$$\frac{1}{2} \|\boldsymbol{\ell} \star \mathbf{x}\|_2^2 = \frac{1}{2} \mathbf{x}^\top \tilde{\mathbf{L}} \mathbf{x}, \quad (6)$$

where  $\tilde{\mathbf{L}} \in \mathbb{R}^{T \times T}$  is a circulant matrix with the degree 2.

**Theorem 1** (Convolution Theorem [27]). *For any vectors  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^T$ , a circular convolution in the time domain is a product in the frequency domain, and it always holds that*

$$\mathbf{x} \star \mathbf{y} = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{x}) \circ \mathcal{F}(\mathbf{y})), \quad (7)$$

where  $\mathcal{F}(\cdot)$  and  $\mathcal{F}^{-1}(\cdot)$  denote the discrete Fourier transform (DFT) and the inverse DFT, respectively.  $\mathcal{F}(\mathbf{x}), \mathcal{F}(\mathbf{y}) \in \mathbb{C}^T$  are



the results of DFT on  $\mathbf{x}, \mathbf{y}$  with  $\mathbb{C}$  denoting the set of complex numbers. The symbol  $\circ$  denotes the Hadamard product.

Essentially, Theorem 1 describes the relationship between circular convolution and DFT, showing that the circular convolution can be implemented in the frequency domain. The circulant matrix is advantageous because the required matrix-vector product can usually be done efficiently by leveraging the structure. Since the Laplacian kernel stems from circulant matrices, the temporal regularization in Eq. (5) can be thus reformulated as follows,

$$\mathcal{R}_\tau(\mathbf{x}) = \frac{1}{2} \|\ell \star \mathbf{x}\|_2^2 = \frac{1}{2T} \|\mathcal{F}(\ell) \circ \mathcal{F}(\mathbf{x})\|_2^2. \quad (8)$$

**Remark 3.** We can prove Eq. (8) as follows. Let

$$\begin{cases} \alpha = \ell \star \mathbf{x}, \\ \beta = \mathcal{F}(\ell) \circ \mathcal{F}(\mathbf{x}), \end{cases} \quad (9)$$

then it takes  $\alpha = \mathcal{F}^{-1}(\beta)$  and  $\mathcal{F}(\alpha) = \beta$  (see Theorem 1). Thus, according to the Parseval's theorem [27], we get

$$\|\alpha\|_2^2 = \frac{1}{T} \|\mathcal{F}(\alpha)\|_2^2 = \frac{1}{T} \|\beta\|_2^2, \quad (10)$$

as claimed in Eq. (8).

## 4.2 Univariate Time Series Imputation

### 4.2.1 Problem Definition

Spatiotemporal traffic data modeling is vital to several ITS applications. Typically, traffic flow data by nature involve certain time series characteristics, e.g., global daily/weekly rhythm and local trends. However, such kind of time series are usually incomplete or even sparse due to unpredictable data collection processes. In the univariate case, the imputation problem can be summarized as Problem 1.

**Problem 1** (Univariate Time Series Imputation). For any partially observed time series  $\mathbf{y} \in \mathbb{R}^T$  with observed index set  $\Omega$ , the goal is to impute the missing data  $\mathcal{P}_\Omega^\perp(\mathbf{y})$  from  $\mathcal{P}_\Omega(\mathbf{y})$ . Herein,  $\mathcal{P}_\Omega : \mathbb{R}^T \rightarrow \mathbb{R}^T$  denotes the orthogonal projection supported on  $\Omega$ , while  $\mathcal{P}_\Omega^\perp : \mathbb{R}^T \rightarrow \mathbb{R}^T$  denotes the orthogonal projection supported on the complement of  $\Omega$ .

**Remark 4.** On the vector  $\mathbf{y} \in \mathbb{R}^T$  with observed index set  $\Omega$ , the operator  $\mathcal{P}_\Omega(\cdot)$  can be described as follows,

$$[\mathcal{P}_\Omega(\mathbf{y})]_t = \begin{cases} y_t, & \text{if } t \in \Omega, \\ 0, & \text{otherwise,} \end{cases} \quad (11)$$

where  $t = 1, 2, \dots, T$ .

### 4.2.2 Model Description

Although ConvNNM and CircNNM can reconstruct missing values in time series, both models fail to incorporate global and local consistency appropriately. In this work, we propose the LCR imputation model, in which we utilize circulant matrix nuclear norm to pursue the global trends and use the temporal regularization to characterize the local trends in time series (see Fig. 2 for an illustration). Formally, the LCR model can be formulated as follows,

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathcal{C}(\mathbf{x})\|_* + \gamma \cdot \mathcal{R}_\tau(\mathbf{x}) \\ \text{s.t.} \quad & \mathcal{P}_\Omega(\mathbf{x}) = \mathcal{P}_\Omega(\mathbf{y}), \end{aligned} \quad (12)$$

where  $\|\cdot\|_*$  denotes the nuclear norm of matrix (i.e., the sum of singular values). The vector  $\mathbf{x} \in \mathbb{R}^T$  is the reconstructed time series corresponding to the partially observed time series  $\mathbf{y}$ . In the objective function,  $\gamma$  is the weight parameter.

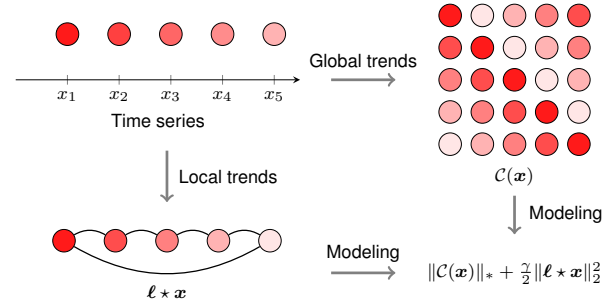


Fig. 2: Illustration of the proposed LCR model.

Since traffic time series data are usually noisy, the strong observation constraint in Eq. (12) should be replaced by  $\|\mathcal{P}_\Omega(\mathbf{z} - \mathbf{y})\|_2 \leq \epsilon$  in which  $\epsilon \geq 0$  is the tolerance. Now, the optimization problem of LCR is given by

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{z}} \quad & \|\mathcal{C}(\mathbf{x})\|_* + \gamma \cdot \mathcal{R}_\tau(\mathbf{x}) \\ \text{s.t.} \quad & \|\mathcal{P}_\Omega(\mathbf{x} - \mathbf{y})\|_2 \leq \epsilon. \end{aligned} \quad (13)$$

Our LCR model stems from ConvNNM [11], [12], and it can be solved by the ADMM framework. To resolve the convex optimization problem of LCR in Eq. (13), we introduce an auxiliary variable  $\mathbf{z}$  to preserve the observation information. Thus, the optimization problem becomes

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{z}} \quad & \|\mathcal{C}(\mathbf{x})\|_* + \gamma \cdot \mathcal{R}_\tau(\mathbf{x}) + \eta \cdot \pi(\mathbf{z}) \\ \text{s.t.} \quad & \mathbf{x} = \mathbf{z}, \end{aligned} \quad (14)$$

where  $\eta$  is a weight parameter. We define  $\pi(\cdot)$  as the reconstructed errors between  $\mathbf{z}$  and  $\mathbf{y}$  in the set  $\Omega$ , which is formally given by  $\pi(\mathbf{z}) = \frac{1}{2} \|\mathcal{P}_\Omega(\mathbf{z} - \mathbf{y})\|_2^2$ . To reinforce both global and local trends in the reconstructed time series  $\mathbf{x}$ , the observation constraint can be related to the noisy version as shown in Eq. (13), thus leading to the denoised and smooth time series in  $\mathbf{x}$ . Accordingly, the augmented Lagrangian function of Eq. (14) can be written as follows,

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \mathbf{z}, \mathbf{w}) = & \|\mathcal{C}(\mathbf{x})\|_* + \gamma \cdot \mathcal{R}_\tau(\mathbf{x}) + \frac{\lambda}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 \\ & + \langle \mathbf{w}, \mathbf{x} - \mathbf{z} \rangle + \eta \cdot \pi(\mathbf{z}), \end{aligned} \quad (15)$$

where  $\mathbf{w} \in \mathbb{R}^T$  is the Lagrange multiplier, and  $\lambda$  is a hyperparameter. The symbol  $\langle \cdot, \cdot \rangle$  denotes the inner product. Note that the constraint  $\mathbf{x} = \mathbf{z}$  in the optimization problem is relaxed by the Lagrange multiplier.

Thus, the ADMM scheme can be summarized as follows,

$$\begin{cases} \mathbf{x} := \arg \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mathbf{z}, \mathbf{w}), \\ \mathbf{z} := \arg \min_{\mathbf{z}} \mathcal{L}(\mathbf{x}, \mathbf{z}, \mathbf{w}), \\ \mathbf{w} := \mathbf{w} + \lambda(\mathbf{x} - \mathbf{z}), \end{cases} \quad (16)$$

which is a two-block ADMM. Since the objective function of Eq. (14) is the sum of two separable convex functions, the convergence of LCR can be proved as in [29].

#### 4.2.3 Estimating the Variable $x$

In particular, with respect to the variable  $x$ , we can rewrite the regularization terms in Eq. (15) as follows,

$$\begin{aligned} f &= \frac{\gamma}{2} \|\ell \star x\|_2^2 + \frac{\lambda}{2} \|x - z + w/\lambda\|_2^2 \\ &= \frac{\gamma}{2T} \|\hat{\ell} \circ \hat{x}\|_2^2 + \frac{\lambda}{2T} \|\hat{x} - \hat{z} + \hat{w}/\lambda\|_2^2, \end{aligned} \quad (17)$$

where  $\hat{\ell} = \mathcal{F}(\ell)$ , and we introduce the variables  $\{\hat{x}, \hat{z}, \hat{w}\} = \{\mathcal{F}(x), \mathcal{F}(z), \mathcal{F}(w)\}$  referring to  $\{x, z, w\}$  in the frequency domain. Notably, temporal regularization  $\mathcal{R}_\tau(x)$  can be converted into a DFT copy (see Eq. (8)), and the Parseval's theorem is also applicable to the remaining term of  $f$ .

**Lemma 1.** For any vector  $x \in \mathbb{R}^T$ , the nuclear norm of the resultant circulant matrix  $\mathcal{C}(x) \in \mathbb{R}^{T \times T}$  is related to the DFT:

$$\|\mathcal{C}(x)\|_* = \|\mathcal{F}(x)\|_1. \quad (18)$$

*Proof.* For any circulant matrix  $\mathcal{C}(x)$  if and only if it is diagonalizable by the unitary matrix, the eigenvalue decomposition [26] can be written as follows,

$$\mathcal{C}(x) = U \text{diag}(\mathcal{F}(x)) U^H, \quad (19)$$

where  $\cdot^H$  denotes the conjugate transpose. Since  $U$  is a unitary matrix, it always holds that

$$\|\mathcal{C}(x)\|_* = \|U \text{diag}(\mathcal{F}(x)) U^H\|_* = \|\text{diag}(\mathcal{F}(x))\|_* = \|\mathcal{F}(x)\|_1,$$

and we can calculate the singular values of  $\mathcal{C}(x)$  from the FFT of  $x$ . Here, FFT is an efficient algorithm for computing the DFT in  $\mathcal{O}(T \log T)$  time.  $\square$

Going back to the ADMM scheme in Eq. (16) and using the property of circulant matrix nuclear norm in Lemma 1, the  $x$ -subproblem can be converted into the optimization over the variable  $\hat{x}$  in the frequency domain. Thus,

$$x := \arg \min_x \|\mathcal{C}(x)\|_* + \frac{\gamma}{2} \|\ell \star x\|_2^2 + \frac{\lambda}{2} \|x - z + w/\lambda\|_2^2, \quad (20)$$

is equivalent to

$$\hat{x} := \arg \min_{\hat{x}} \|\hat{x}\|_1 + \frac{\gamma}{2T} \|\hat{\ell} \circ \hat{x}\|_2^2 + \frac{\lambda}{2T} \|\hat{x} - \hat{z} + \hat{w}/\lambda\|_2^2. \quad (21)$$

On each  $\hat{x}_t$ , the optimization problem is given by

$$\begin{aligned} \hat{x}_t &:= \arg \min_{\hat{x}_t} |\hat{x}_t| + \frac{\gamma}{2T} |\hat{\ell}_t \hat{x}_t|^2 + \frac{\lambda}{2T} |\hat{x}_t - \hat{z}_t + \hat{w}_t/\lambda|^2 \\ &= \arg \min_{\hat{x}_t} |\hat{x}_t| + \frac{\gamma |\hat{\ell}_t|^2 + \lambda}{2T} \left| \hat{x}_t - \frac{\lambda \hat{z}_t - \hat{w}_t}{\gamma |\hat{\ell}_t|^2 + \lambda} \right|^2, \end{aligned} \quad (22)$$

where  $|\hat{\ell}_t \hat{x}_t|^2 = |\hat{\ell}_t|^2 \cdot |\hat{x}_t|^2$ .

The resultant  $\ell_1$ -norm minimization is memory-efficient, easy to compute, and preserves the singular values of circulant matrix that are due to the FFT. The closely related analysis and results are also discussed in [10], [11], [12].

According to Eqs. (21) and (22), we let

$$\hat{h} \triangleq (\lambda \hat{z} - \hat{w}) \oslash (\gamma \hat{\ell}^* \circ \hat{\ell} + \lambda \mathbb{1}_T), \quad (23)$$

where  $\cdot^*$  represents the complex conjugate, and  $\oslash$  denotes the Hadamard division.  $\mathbb{1}_T \in \mathbb{R}^T$  is the vector of ones.

The closed-form solution to  $\hat{x}$  can be found in Lemma 2. As we have the closed-form solution as described in Eq. (28) (i.e., with respect to each  $\hat{x}_t$ ) such that

$$\hat{x}_t := \frac{\hat{h}_t}{|\hat{h}_t|} \cdot \max\{0, |\hat{h}_t| - 1/\delta_t\}, \quad (24)$$

with

$$\begin{cases} \delta_t \triangleq (\gamma |\hat{\ell}_t|^2 + \lambda)/T, \\ \hat{h}_t \triangleq (\lambda \hat{z}_t - \hat{w}_t)/(\gamma |\hat{\ell}_t|^2 + \lambda), \end{cases} \quad (25)$$

we can therefore update the variable  $x$  by

$$x := \mathcal{F}^{-1}(\hat{x}). \quad (26)$$

**Lemma 2.** Following Eqs. (21) and (22), for any  $\ell_1$ -norm minimization problem in complex space such that

$$\min_{\hat{x}} \|\hat{x}\|_1 + \frac{\delta}{2} \|\hat{x} - \hat{h}\|_2^2, \quad (27)$$

with complex-valued vectors  $\hat{x}, \hat{h} \in \mathbb{C}^T$  and weight parameter  $\delta \in \mathbb{R}$ , element-wise, the solution is given by

$$\hat{x}_t := \frac{\hat{h}_t}{|\hat{h}_t|} \cdot \max\{0, |\hat{h}_t| - 1/\delta\}, t = 1, \dots, T. \quad (28)$$

*Proof.* In theory, Lemma 2 invokes the shrinkage operator in [30], [31], [11].  $\square$

#### 4.2.4 Estimating the Variable $z$

In the ADMM scheme (see Eq. (16)), the subproblem with respect to the variable  $z$  can be written as follows,

$$\min_z \frac{\lambda}{2} \|x - z - w/\lambda\|_2^2 + \frac{\eta}{2} \|\mathcal{P}_\Omega(z - y)\|_2^2. \quad (29)$$

Let  $g$  be the objective function, then the partial derivative with respect to  $z$  can be formed by  $\mathcal{P}_\Omega(z)$  and  $\mathcal{P}_\Omega^\perp(z)$ :

$$\begin{cases} \partial g / \partial \mathcal{P}_\Omega(z) = \lambda \mathcal{P}_\Omega(z - x - w/\lambda) + \eta \mathcal{P}_\Omega(z - y), \\ \partial g / \partial \mathcal{P}_\Omega^\perp(z) = \lambda \mathcal{P}_\Omega^\perp(z - x - w/\lambda). \end{cases} \quad (30)$$

As a result,  $\partial g / \partial z = \mathbf{0}$  produces a closed-form solution:

$$\begin{aligned} z &:= \left\{ z \mid \partial g / \partial \mathcal{P}_\Omega(z) + \partial g / \partial \mathcal{P}_\Omega^\perp(z) = \mathbf{0} \right\} \\ &= \frac{1}{\lambda + \eta} \mathcal{P}_\Omega(\lambda x + w + \eta y) + \frac{1}{\lambda} \mathcal{P}_\Omega^\perp(\lambda x + w). \end{aligned} \quad (31)$$

In this case, if  $\eta \rightarrow +\infty$ , then the solution refers to the LCR with strong observation constraint in Eq. (12). In terms of the parameter  $\eta$ , we can preferably set its value to  $\eta = c \cdot \lambda$  with  $c \in \{10^2, 10^3\}$  to preserve the observation information.

#### 4.2.5 Solution Algorithm

As mentioned above, our LCR model reinforces the modeling processes of global low-rank and local temporal trends in time series data. Since we utilize the circulant matrix and circular convolution, it is not hard to show the appealing properties of DFT and lead to an elegant and fast solution algorithm. Algorithm 1 summarizes the implementation of the proposed LCR model.

To analyze the empirical time complexity of LCR (with 50 iterations by default), Fig. 3 shows the running times of LCR on the generated data with different data lengths (i.e., data  $y \in \mathbb{R}^T$  with  $T \in \{2^{10}, 2^{11}, \dots, 2^{20}\}$ ). As shown

### Algorithm 1 Laplacian Convolutional Representation (LCR)

**Input:** Data  $\mathbf{y} \in \mathbb{R}^T$  with observed index set  $\Omega$ , Laplacian kernel size  $\tau \in \mathbb{Z}^+$ , and hyperparameters  $\{\gamma, \lambda, \eta\}$ .

**Output:** Reconstructed vector  $\mathbf{x} \in \mathbb{R}^T$ .

- 1: Initialize  $\{\mathbf{x}_0, \mathbf{z}_0, \mathbf{w}_0\}$ .
- 2: Construct the Laplacian kernel  $\ell$  with  $\tau$  and perform FFT on it to get  $\hat{\ell}$ .
- 3: **for**  $i = 0$  to maximum iteration **do**
- 4:   Perform FFT on  $\{\mathbf{z}_i, \mathbf{w}_i\}$ .
- 5:   Compute  $\hat{\mathbf{h}}$  by Eq. (23).
- 6:   Compute  $\hat{\mathbf{x}}$  by the shrinkage in Eq. (24).
- 7:   Compute  $\mathbf{x}_{i+1}$  by  $\mathbf{x}_{i+1} = \mathcal{F}^{-1}(\hat{\mathbf{x}})$  (see Eq. (26)).
- 8:   Compute  $\mathbf{z}_{i+1}$  by Eq. (31).
- 9:   Compute  $\mathbf{w}_{i+1} = \mathbf{w}_i + \lambda(\mathbf{x}_{i+1} - \mathbf{z}_{i+1})$  (see Eq. (16)).
- 10: **end for**

in Fig. 3(a), we compare LCR with ConvNNM (e.g., kernel size  $\tilde{\tau} = 2^4$  in this case), demonstrating that LCR is more efficient than ConvNNM. Typically, ConvNNM can be converted into a standard nuclear norm minimization with singular value thresholding (i.e., of time complexity  $\mathcal{O}(\tilde{\tau}^2 T)$ ) [6], [5], [32], [11], [12]. The computational cost of ConvNNM would increase with a larger  $\tilde{\tau}$  for the convolution matrix  $\mathcal{C}_{\tilde{\tau}}(\mathbf{y}) \in \mathbb{R}^{T \times \tilde{\tau}}$ . In contrast, both CircNNM and LCR have an efficient solution through FFT in  $\mathcal{O}(T \log T)$  time.

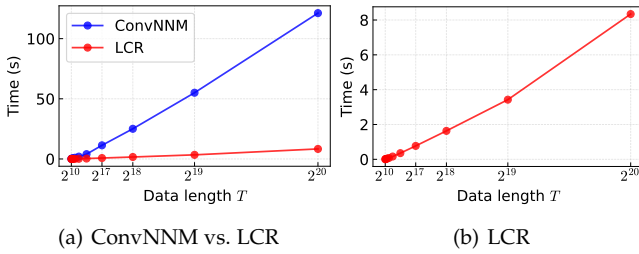


Fig. 3: Empirical time complexity. The model is tested 50 times on each generated data.

## 4.3 Multivariate Time Series Imputation

### 4.3.1 Problem Definition

Considering both the spatial and temporal dimensions in traffic data, we have a multivariate time series imputation task as described in Problem 2. The critical question is how to characterize both spatial and temporal dependencies of traffic time series data in the modeling process.

**Problem 2** (Multivariate Time Series Imputation). *For any partially observed time series  $\mathbf{Y} \in \mathbb{R}^{N \times T}$  with  $N$  variables and  $T$  time steps, if its observed index set is denoted by  $\Omega$ , then the goal is to impute the missing data  $\mathcal{P}_{\Omega}^{\perp}(\mathbf{Y})$  from  $\mathcal{P}_{\Omega}(\mathbf{Y})$ . Herein,  $\mathcal{P}_{\Omega} : \mathbb{R}^{N \times T} \rightarrow \mathbb{R}^{N \times T}$  and  $\mathcal{P}_{\Omega}^{\perp} : \mathbb{R}^{N \times T} \rightarrow \mathbb{R}^{N \times T}$  denote the orthogonal projection supported on  $\Omega$  and the complement of  $\Omega$ , respectively.*

### 4.3.2 Model Description

In this study, we introduce the nuclear norm of the circulant tensor  $\mathcal{C}(\mathbf{X})$  in Definition 2, which follows the tensor nuclear norm proposed in [33], [34].

**Definition 2** (Circulant Tensor Nuclear Norm). *For any matrix  $\mathbf{X} \in \mathbb{R}^{N \times T}$ , the corresponding circulant tensor is  $\mathcal{C}(\mathbf{X}) \in \mathbb{R}^{N \times N \times T \times T}$ , which can be factorized in the Tucker format (i.e., higher-order singular value decomposition (SVD)) [35]:*

$$\mathcal{C}(\mathbf{X}) = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_1 \times_3 \mathbf{U}_2 \times_4 \mathbf{U}_2, \quad (32)$$

where  $\mathcal{S} \in \mathbb{R}^{N \times N \times T \times T}$  is the core tensor (consisting of singular values [33], [34]), while  $\mathbf{U}_1 \in \mathbb{R}^{N \times N}$  and  $\mathbf{U}_2 \in \mathbb{R}^{T \times T}$  are unitary matrices. The notation  $\times_k, \forall k \in \{1, 2, 3, 4\}$  represents the mode- $k$  product between tensor and matrix [35]. The circulant tensor nuclear norm is defined as

$$\|\mathcal{C}(\mathbf{X})\|_* = \sum_{n=1}^N \sum_{t=1}^T s_{n,n,t,t}, \quad (33)$$

where  $s_{n,n,t,t}$  is the  $(n, n, t, t)$ -th entry of the core tensor  $\mathcal{S}$ .

Recall that the temporal regularization in the univariate LCR model works on each time series independently. To jointly characterize the spatial and temporal dependencies for traffic time series, we consider a separable kernel in the LCR model, namely,  $\mathbf{K} \triangleq \ell_s \ell^T \in \mathbb{R}^{N \times T}$  with the spatial kernel  $\ell_s = (1, 0, \dots, 0)^T \in \mathbb{R}^N$  (i.e., the first column of the  $N$ -by- $N$  identity matrix). In the case of spatial modeling,  $\ell_s$  can also be introduced as the Laplacian kernel in Definition 1. The optimization problem of two-dimensional LCR (LCR-2D) can be formulated as follows,

$$\begin{aligned} \min_{\mathbf{X}} \quad & \|\mathcal{C}(\mathbf{X})\|_* + \frac{\gamma}{2} \|\mathbf{K} \star \mathbf{X}\|_F^2 \\ \text{s.t.} \quad & \|\mathcal{P}_{\Omega}(\mathbf{X} - \mathbf{Y})\|_F \leq \epsilon, \end{aligned} \quad (34)$$

where  $\|\cdot\|_F$  denotes the Frobenius norm of a matrix. The two-dimensional circular convolution is described in Definition 3. Although  $\ell_s = (1, 0, \dots, 0)^T \in \mathbb{R}^N$  does not provide any spatial dependencies, the nuclear norm of circulant operator on  $\mathbf{X}$  can achieve implicit spatial modeling. If applicable, one can characterize spatial correlations by using the Laplacian kernel.

**Definition 3** (Two-Dimensional Circular Convolution [36], [25]). *For any matrices  $\mathbf{X} \in \mathbb{R}^{N \times T}$  and  $\mathbf{K} \in \mathbb{R}^{\nu_1 \times \nu_2}$  with  $\nu_1 \leq N, \nu_2 \leq T$ , the circular convolution of two matrices is defined as*

$$\mathbf{Z} = \mathbf{K} \star \mathbf{X} \in \mathbb{R}^{N \times T}, \quad (35)$$

or element-wise,

$$z_{n,t} = \sum_{i=1}^{\nu_1} \sum_{j=1}^{\nu_2} \kappa_{i,j} x_{n-i+1,t-j+1}, \quad (36)$$

where  $n = 1, 2, \dots, N$  and  $t = 1, 2, \dots, T$ .  $\kappa_{i,j}$  is the  $(i, j)$ -th entry of  $\mathbf{K}$ .

**Remark 5.** *In the field of signal processing, the two-dimensional circular convolution also possesses the properties that are associated with two-dimensional DFT. According to the convolution theorem and the Parseval's theorem, we have*

$$\|\mathbf{K} \star \mathbf{X}\|_F^2 = \frac{1}{NT} \|\mathcal{F}(\mathbf{K}) \circ \mathcal{F}(\mathbf{X})\|_F^2, \quad (37)$$

where  $\mathcal{F}(\cdot)$  denotes the two-dimensional DFT. Typically, two-dimensional DFT can be computed by first transforming each column vector (or row vector) and then each row vector (or column vector) of the matrix [25].

### 4.3.3 ADMM Scheme and Solution Algorithm

As mentioned above, LCR-2D in the multivariate case is a convex problem that can be resolved by the ADMM. Following Eq. (16), the ADMM scheme is given by

$$\begin{cases} \mathbf{X} := \arg \min_{\mathbf{X}} \mathcal{L}(\mathbf{X}, \mathbf{Z}, \mathbf{W}), \\ \mathbf{Z} := \arg \min_{\mathbf{Z}} \mathcal{L}(\mathbf{X}, \mathbf{Z}, \mathbf{W}) \\ \quad = \frac{1}{\lambda + \eta} \mathcal{P}_{\Omega}(\lambda \mathbf{X} + \mathbf{W} + \eta \mathbf{Y}) + \frac{1}{\lambda} \mathcal{P}_{\Omega}^{\perp}(\lambda \mathbf{X} + \mathbf{W}), \\ \mathbf{W} := \mathbf{W} + \lambda(\mathbf{X} - \mathbf{Z}), \end{cases} \quad (38)$$

where  $\mathcal{L}(\mathbf{X}, \mathbf{Z}, \mathbf{W})$  is the augmented Lagrangian function:

$$\begin{aligned} \mathcal{L}(\mathbf{X}, \mathbf{Z}, \mathbf{W}) = & \|\mathcal{C}(\mathbf{X})\|_* + \frac{\gamma}{2} \|\mathbf{K} \star \mathbf{X}\|_F^2 \\ & + \frac{\lambda}{2} \|\mathbf{X} - \mathbf{Z}\|_F^2 + \langle \mathbf{W}, \mathbf{X} - \mathbf{Z} \rangle \\ & + \frac{\eta}{2} \|\mathcal{P}_{\Omega}(\mathbf{Z} - \mathbf{Y})\|_F^2. \end{aligned} \quad (39)$$

Although the nuclear norm of circulant tensor in Definition 2 is more complicated than the nuclear norm of circulant matrix (see Lemma 1), two-dimensional DFT allows one to find the solution to  $\mathbf{X}$ , if not mentioning the difficulty of obtaining a unique decomposition of circulant tensor. In the frequency domain, it takes

$$\begin{aligned} \hat{\mathbf{X}} := \arg \min_{\hat{\mathbf{X}}} & \|\hat{\mathbf{X}}\|_1 + \frac{\gamma}{2NT} \|\hat{\mathbf{K}} \circ \hat{\mathbf{X}}\|_F^2 \\ & + \frac{\lambda}{2NT} \|\hat{\mathbf{X}} - \hat{\mathbf{Z}} + \hat{\mathbf{W}}/\lambda\|_F^2, \end{aligned} \quad (40)$$

where  $\{\hat{\mathbf{K}}, \hat{\mathbf{X}}, \hat{\mathbf{Z}}, \hat{\mathbf{W}}\}$  refers to  $\{\mathbf{K}, \mathbf{X}, \mathbf{Z}, \mathbf{W}\}$  in the frequency domain. Without loss of generality, this subproblem for  $\ell_1$ -norm minimization in complex space can also be solved by the shrinkage operator in Eq. (28). Although Definition 2 describes the formula of higher-order SVD that leads to the nuclear norm, the computation of the circulant tensor nuclear norm minimization is actually converted into an  $\ell_1$ -norm minimization with two-dimensional FFT. Algorithm 2 summarizes the whole scheme of LCR-2D.

---

#### Algorithm 2 Two-Dimensional Laplacian Convolutional Representation (LCR-2D)

---

**Input:** Data  $\mathbf{Y} \in \mathbb{R}^{N \times T}$  with observed index set  $\Omega$ , Laplacian kernel size  $\tau \in \mathbb{Z}^+$ , and hyperparameters  $\{\gamma, \lambda, \eta\}$ .

**Output:** Reconstructed matrix  $\mathbf{X} \in \mathbb{R}^{N \times T}$ .

- 1: Initialize  $\{\mathbf{X}_0, \mathbf{Z}_0, \mathbf{W}_0\}$ .
  - 2: Construct the Laplacian kernel  $\ell \in \mathbb{R}^T$  with  $\tau$ .
  - 3: Construct the spatial kernel  $\ell_s = (1, 0, \dots, 0) \in \mathbb{R}^N$  (or  $\ell_s \in \mathbb{R}^N$  with  $\tau_s$ ) and build up a separable kernel  $\mathbf{K} \triangleq \ell_s \ell^T$ .
  - 4: **for**  $i = 0$  to maximum iteration **do**
  - 5:   Perform FFT on  $\{\mathbf{Z}_i, \mathbf{W}_i\}$ .
  - 6:   Compute  $\hat{\mathbf{X}}$  by referring to the shrinkage in Eq. (28).
  - 7:   Compute  $\mathbf{X}_{i+1}$  by  $\mathbf{X}_{i+1} = \mathcal{F}^{-1}(\hat{\mathbf{X}})$ .
  - 8:   Compute  $\mathbf{Z}_{i+1}$  by Eq. (38).
  - 9:   Compute  $\mathbf{W}_{i+1}$  by Eq. (38).
  - 10: **end for**
- 

## 5 UNIVARIATE TRAFFIC TIME SERIES IMPUTATION

This section evaluates the reconstruction of univariate time series from partial observations with LCR. We focus on understanding how well the reconstructed time series preserve the global and local trends in the imputation task. Experiments are conducted on both traffic speed time series (weak periodicity and strong noises) and traffic volume time series (strong periodicity) collected through dual-loop detectors on the highway network of Portland, USA.<sup>1</sup> Specifically, the datasets include: **(Traffic speed)** The speed observations with 15-min time resolution (i.e., 96 expected data samples per day) over three days (i.e., of length 288). **(Traffic volume)** The volume observations have the same time resolution as the traffic speed. In particular, we consider the imputation scenarios on fully observed, 20%, 10%, and 5% observed data, respectively. To generate partially observed data, we randomly mask a certain number of data as missing values. Since both traffic speed and volume time series show cyclical patterns, the start data points and the end data points can be connected by circulant matrix and circular convolution without flipping operation in Remark 1.

In the following experiments, to evaluate the imputation performance, we use the mean absolute percentage error (MAPE) and the root mean square error (RMSE):

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i}, \quad \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2},$$

where  $n$  is the total number of estimated values, and  $y_i$  and  $\hat{y}_i$  are the actual value and its estimation, respectively.

### 5.1 Traffic Speed

Fig. 4(a) demonstrates the reconstructed time series by LCR on the fully observed time series, in which the data noises can be smoothed out due to the existence of temporal regularization and the relaxation of observation constraint  $\|\mathcal{P}_{\Omega}(\mathbf{x} - \mathbf{y})\|_2 \leq \epsilon$  in which  $\mathbf{y}$  and  $\mathbf{x}$  are the partial observations and the reconstructed time series, respectively. Fig. 4(b) and 4(c) show the imputation performance by LCR with partial observations. Of these results, the reconstructed time series can accurately approximate both partial observations and missing values while preserving the trends of the ground truth time series.

Next, we test a more challenging scenario in which we aim to reconstruct 95% missing values from 5% observations (i.e., reconstructing 274 missing values from only 14 data samples). We compare the time series imputation of LCR with the following baseline models: (i) CircNNM [11] (equivalent to LCR without temporal regularization), (ii) ConvNNM [11], and (iii) ConvNNM+ (i.e., ConvNNM with temporal regularization) such that

$$\begin{aligned} \min_{\mathbf{x}} & \|\mathcal{C}_{\tau}(\mathbf{x})\|_* + \frac{\gamma}{2} \|\ell \star \mathbf{x}\|_2^2 \\ \text{s.t.} & \|\mathcal{P}_{\Omega}(\mathbf{x} - \mathbf{y})\|_2 \leq \epsilon. \end{aligned} \quad (41)$$

Essentially, comparing the proposed LCR model with these baseline models allows one to highlight 1) the importance of global and local trends modeling in LCR, and

1. <https://portal.its.pdx.edu/>



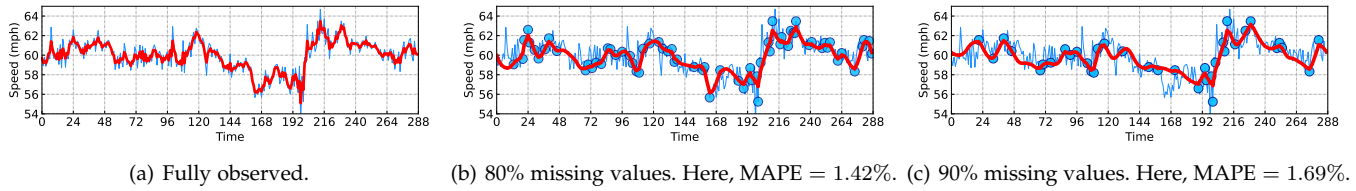


Fig. 4: Univariate traffic time series imputation on the freeway traffic speed time series. The blue curve represents the ground truth time series, while the red curve refers to the reconstructed time series produced by LCR. Here, partial observations are illustrated as blue circles.

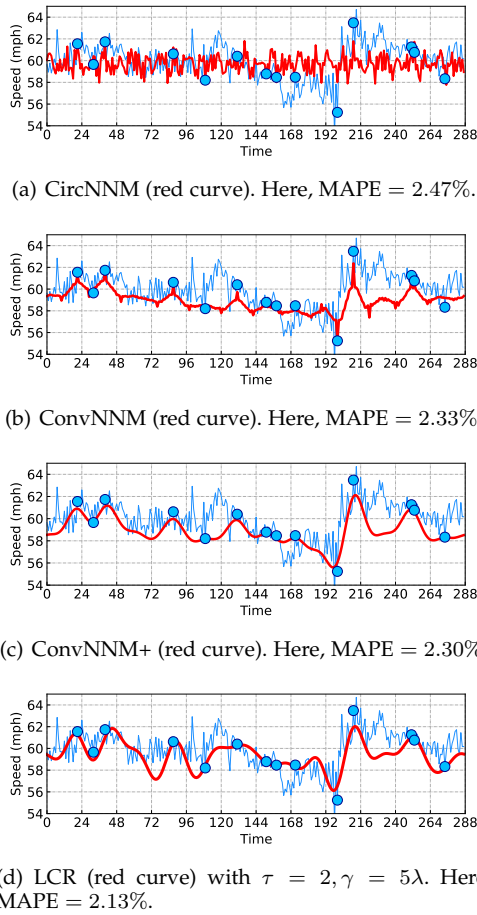


Fig. 5: Univariate traffic time series imputation on the freeway traffic speed time series. In this case, we mask 95% observations as missing values and only have 14 speed observations for training the model.

2) the fast implementation of LCR via the use of FFT. The imputation results on the given time series are shown in Fig. 5, we can summarize the following findings:

- The reconstructed time series by CircNNM shows high fluctuations due to the lack of local trend modeling. As a result, the reconstructed time series fails to reproduce trends of the ground truth at such a high missing rate.
- ConvNNM performs better than CircNNM. The reconstructed time series fits the observed values well, but the trend does not match the ground truth perfectly.

Unlike CircNNM, ConvNNM cannot employ a fast implementation via FFT. As a result, ConvNNM is not well-suited to large problems (also see Fig. 3).

- ConvNNM+ outperforms ConvNNM. This demonstrates the significance of Laplacian kernel for time series modeling. However, both ConvNNM and ConvNNM+ require implementing the singular value thresholding on  $T$ -by- $\tilde{\tau}$  convolution matrices.

As shown in Fig. 5(d), the reconstructed time series by LCR demonstrates consistent global and local trends with the trends of ground truth time series, and our LCR model clearly outperforms the baseline models. By comparing CircNNM with LCR, the imputation results emphasize the importance of temporal regularization  $\mathcal{R}_\tau(x)$ .

## 5.2 Traffic Volume

In Fig. 6, the traffic volume time series is characterized by a strong daily rhythm. The time series shows relatively low traffic volume at night, and the traffic volume reaches a peak during rush hours, implying typical travelers' behavioral rhythms. In this case, the time series possesses strong seasonality and three remarkable peaks over three days.

The task is reconstructing 95% missing values from 5% observations. As can be seen, due to the lack of explicit temporal modeling, both CircNNM and ConvNNM models cannot produce time series as smooth as LCR. Observing the reconstructed time series, it is clear that LCR produces more accurate reconstruction results than both CircNNM and ConvNNM. Yet, in contrast to the traffic speed time series, due to the strong seasonality (i.e., daily rhythm in traffic flow) in the traffic volume time series, both CircNNM and ConvNNM can produce reasonable time series, which seems to be consistent with the results in [12].

## 6 MULTIVARIATE TRAFFIC TIME SERIES IMPUTATION

In this section, we consider some real-world multivariate traffic time series imputation scenarios, including speed field reconstruction of vehicular traffic flow and large-scale traffic speed data imputation. The experiments and evaluation of LCR are expected to demonstrate the efficiency of global and local trends modeling on traffic time series with spatiotemporal settings.



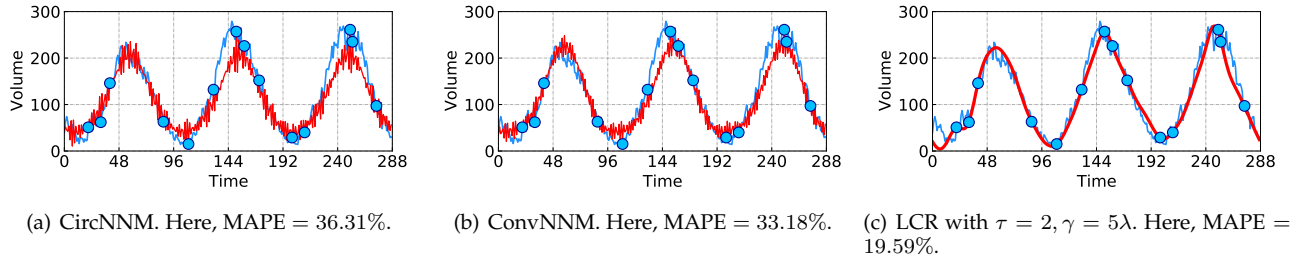


Fig. 6: Univariate time series imputation on the freeway traffic volume time series. In this case, we randomly remove 95% observations as missing values, and we only have 14 volume observations for the reconstruction.

## 6.1 Speed Field Reconstruction

Speed field reconstruction is a critical problem in vehicular traffic flow modeling as the data collection process is often far from ideal [37]. The task is to reconstruct the traffic speed distribution along a road segment during a period using the motion states (i.e., time, location, and speed) of a small portion of vehicles [38], e.g., to estimate the traffic congestion by using only the motion states of taxis.

### 6.1.1 Datasets

Traffic states such as movement speeds measure the status of vehicular traffic flow, which can be extracted from high-resolution vehicle trajectories [37]. In the literature, there are several classical traffic flow datasets, such as HighD<sup>2</sup> and CitySim<sup>3</sup> collected through drones across a fraction of roads [39], [40]. The drones can gather the movement trajectories of vehicles for driving behavior modeling and traffic state analysis. In terms of the HighD dataset, there are 60 video recordings available from several German highway sections, and each covers trajectory data on the 420-meter road segment with multiple lanes. We select the video #46 for experiments and the resultant speed field tensor is of size  $142 \times 595 \times 3$ , i.e., 142 discrete locations (3-meter spatial resolution), 595 time steps (2-second time resolution), and 3 lanes, see Fig. 8(a), 8(d), and 8(g). In terms of the CitySim dataset, it contains video recordings of vehicular trajectory data collected from various road infrastructures, including freeways, expressways, and intersections. We select the trajectory data of one freeway for experiments. The resultant speed field tensor is of size  $126 \times 442 \times 3$ , i.e., 126 discrete locations (5-meter spatial resolution), 442 time steps (2-second time resolution), and 3 lanes, see Fig. 9(a), 9(d), and 9(g). In the experiment, we randomly mask a certain fraction of trajectories (e.g., 30%, 50%, and 70%) and then construct the speed field that shows partial observations.

### 6.1.2 Baseline Models

For comparison, we consider some low-rank matrix/tensor completion models, including LRMC [6], Hankel tensor factorization (HTF [7] in the form of CP factorization), High-accuracy low-rank tensor completion (HaLRTC [41]), and low-rank tensor completion with truncated nuclear norm

(LRTC-TNN [42]). To highlight the advantage of LCR-2D, we also choose the following baseline models:

- LCR<sub>N</sub>: We consider to implement LCR over  $N$  univariate time series problems independently, i.e.,

$$\min_{\mathbf{X}} \sum_{n=1}^N \|\mathcal{C}(\mathbf{x}_n)\|_* + \frac{\gamma}{2} \sum_{n=1}^N \|\ell \star \mathbf{x}_n\|_2^2 \quad (42)$$

s.t.  $\|\mathcal{P}_\Omega(\mathbf{X} - \mathbf{Y})\|_F \leq \epsilon,$

where  $\mathbf{x}_n \in \mathbb{R}^T$ ,  $n = 1, 2, \dots, N$  are the univariate time series of  $\mathbf{X} \in \mathbb{R}^{N \times T}$ .

- CTNNM: Circulant tensor nuclear norm minimization whose objective function is specified as  $\|\mathcal{C}(\mathbf{X})\|_*$  (i.e., the special case of LCR-2D without regularization term).
- Quadratic variation completion (QVC) such that

$$\min_{\mathbf{X}} \frac{\gamma}{2} \text{tr}(\mathbf{X} \tilde{\mathbf{L}} \mathbf{X}^\top) \quad (43)$$

s.t.  $\|\mathcal{P}_\Omega(\mathbf{X} - \mathbf{Y})\|_F \leq \epsilon,$

with the Laplacian matrix  $\tilde{\mathbf{L}}$  referring to Eq. (6), and  $\text{tr}(\cdot)$  denotes the trace of matrix. In the meanwhile, we let the spatial kernel be  $\ell_s = (1, 0, \dots, 0, 0) \in \mathbb{R}^N$  and the temporal kernel be parameterized by  $\tau$  in Definition 1, referring to Laplacian kernelized completion (LKC, i.e., the special case of LCR-2D without circulant tensor nuclear norm).

### 6.1.3 Model Setting

To eliminate correlations between the start data points and the end data points in our model, we introduce a  $2N$ -by- $2T$  block matrix (i.e., with four blocks) which flips the original matrix  $\mathbf{Y}$ , see Fig. 7. This matrix can be regarded as the input into the LCR-2D model. The model results are constructed by averaging the blocks according to the flipping operation.

In this case, we intend to highlight the importance of global/local trend modeling on the speed field. Since each speed field dataset includes lane-level trajectories, we evaluate LCR-2D on each lane-level speed field independently. The default hyperparameter  $\eta$  is set as  $10^2 \lambda$ . On the HighD dataset, we set  $\lambda = 10^{-3} NT$  and  $\gamma = \lambda$ . We validate the kernel size as  $\tau = 1$  on the 30%-trajectory speed field and  $\tau = 2$  on the 50%-/70%-trajectory speed field. On the CitySim dataset, we set  $\lambda = 10^{-4} NT$ ,  $\gamma = \lambda$ , and  $\tau = 3$  after the numerical evaluation.

2. Highway Drone (HighD) dataset is available at <https://www.highd-dataset.com/>.

3. A Drone-Based Vehicle Trajectory Dataset for Safety-Oriented Research and Digital Twins (CitySim) is available at <https://github.com/ozheng1993/UCF-SST-CitySim-Dataset>.

TABLE 1: Reconstruction performance (in MAPE (%)/RMSE) achieved by LCR-2D and baseline models on both speed field datasets. Note that the best results are emphasized in bold fonts, and we underline the second-best results.

Dataset	Rate	LCR-2D	LCR <sub>N</sub>	CTNNM	QVC	LKC	LRMC [6]	HTF [7]	HaLRTC [41]	LRTC-TNN [42]
HighD	30%	<b>3.57/1.41</b>	6.79/4.09	3.91/ <u>1.46</u>	6.79/4.59	6.62/4.60	7.11/4.61	4.38/1.66	4.63/1.78	<u>3.89</u> /1.50
	50%	<b>4.06/1.52</b>	6.88/3.49	4.57/ <u>1.61</u>	7.02/3.97	6.06/3.88	9.37/5.30	4.67/1.63	5.63/2.06	<u>4.53</u> /1.63
	70%	<b>4.73/1.77</b>	7.91/3.32	<u>5.61</u> / <u>1.87</u>	8.12/4.03	6.52/3.87	16.79/8.91	5.76/2.17	7.94/3.01	5.75/2.03
CitySim	30%	<b>8.88/2.71</b>	10.88/4.24	<u>9.23</u> /2.81	10.97/4.30	10.90/4.24	11.27/4.43	9.28/ <u>2.78</u>	10.91/4.31	10.90/4.26
	50%	<b>9.08/2.69</b>	10.56/3.89	9.59/ <u>2.82</u>	10.65/3.96	10.65/3.92	11.42/4.29	<u>9.49</u> /2.83	10.66/3.99	10.49/3.87
	70%	<b>9.07/2.66</b>	10.35/3.67	<u>9.42</u> / <u>2.73</u>	10.33/3.70	10.53/3.71	12.96/4.97	9.93/2.94	11.02/3.92	10.46/3.71

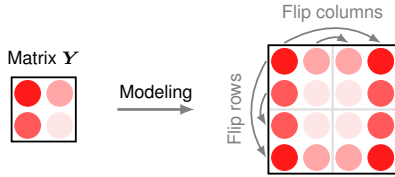


Fig. 7: Constructing the matrix that flips the original matrix  $Y$  along rows and columns simultaneously. This operation can prevent the LCR-2D model from misleading values on the border rows and columns.

#### 6.1.4 Results

Table 1 gives the speed field reconstruction performance of LCR-2D and some baseline models. We can summarize the following findings: (i) As LCR<sub>N</sub> works on the univariate time series independently, spatial correlations of speed fields are totally ignored. The better performance of LCR-2D against LCR<sub>N</sub> highlights the importance of spatial modeling for speed field reconstruction. (ii) Recall that CTNNM is a special case of LCR-2D without Laplacian kernels, we empirically demonstrate that LCR-2D outperforms CTNNM on both datasets, implying that the spatiotemporal local trend modeling is of great significance for producing accurate reconstruction of sparse speed fields. (iii) QVC and LKC cannot produce accurate reconstruction of speed fields as the proposed LCR-2D model, and this implies the importance of circulant tensor nuclear norm in LCR-2D. (iv) LCR-2D performs better than a sequence of matrix/tensor completion models on both datasets. This demonstrates that an appropriate spatiotemporal modeling on the low-rank framework is important for speed field reconstruction.

As mentioned above, both circulant tensor nuclear norm and regularization term with Laplacian kernels in the objective function of LCR-2D are of great significance. Fig. 8 and 9 show the reconstruction of speed fields with sparse inputs (i.e., on 70% masked trajectories). It seems that LCR-2D can achieve satisfactory results for speed field reconstruction because of the spatiotemporal correlation of traffic wave [43] characterized by both global and local trends.

## 6.2 PeMS Traffic Speed Imputation

In what follows, we study the generalization of LCR to high-dimensional data and evaluate the model on a large-scale traffic flow dataset. The data are collected by the California department of transportation through their Performance Measurement System (PeMS) [44]. This dataset contains freeway traffic speed collected from 11,160 traffic

measurement sensors over 4 weeks (the first 4 weeks in the year of 2018) with a 5-minute time resolution (288 time intervals per day) in California, USA.<sup>4</sup> It can be arranged in a matrix of size  $11160 \times 8064$ , and this dataset contains about 90 million observations.

To set up the imputation task, we randomly mask 30%, 50%, 70%, and 90% traffic speed observations as missing values, referred to as 30%, 50%, 70%, and 90% missing rates, respectively. To assess the imputation performance, we use the actual values of the masked missing entries as the ground truth to compute MAPE and RMSE. For comparison, the chosen baseline models are CircNNM [11], LRMC [6], HaLRTC [41], LRTC-TNN [42], and nonstationary temporal matrix factorization (NoTMF [17]). We also consider the comparison with (i) LCR-2D, (ii) LCR<sub>N</sub> (i.e., implementing LCR over  $N$  univariate time series independently), (iii) LCR (i.e., the LCR model via vectorization on the  $N$ -by- $T$  multivariate time series), and (iv) CTNNM. Due to the day-to-day cyclical patterns of this dataset, we do not consider the flipping operation (see Fig. 7) in the numerical evaluation.

As shown in Table 2, LCR-2D, LCR<sub>N</sub>, and LCR achieve very competitive imputation accuracy and outperform baseline models. Among the baseline models in Table 2, CTNNM and CircNNM are the special cases of LCR-2D and LCR, respectively. Both models can be implemented by FFT like our LCR models. Comparing the imputation performance of CTNNM (or CircNNM) and LCR-2D (or LCR) shows the significant improvement of imputation achieved by LCR-2D over CTNNM, mainly due to the existence of the regularization term with Laplacian kernels. Therefore, introducing local trend modeling with Laplacian kernels in our LCR models is of great significance for traffic time series imputation. Despite the aforementioned comparison, our LCR models perform significantly better than some matrix/tensor completion algorithms such as LRMC, HaLRTC, and LRTC-TNN. Here, these matrix/tensor completion algorithms provide well-suited frameworks for reconstructing missing values in data matrix/tensor. However, they involve high time complexity in the singular value thresholding process, e.g.,

$$\mathcal{O}(\min\{N^2T, NT^2\}) \text{ (SVD)} \quad \text{vs.} \quad \mathcal{O}(NT \log(NT)) \text{ (FFT)},$$

making it extremely costly for large-scale problems. In contrast, our LCR models take an  $\ell_1$ -norm minimization in the frequency domain via the use of FFT. NoTMF jointly characterizes global and local trends by a unified and effi-

4. The dataset is available at <https://doi.org/10.5281/zenodo.3939792>.

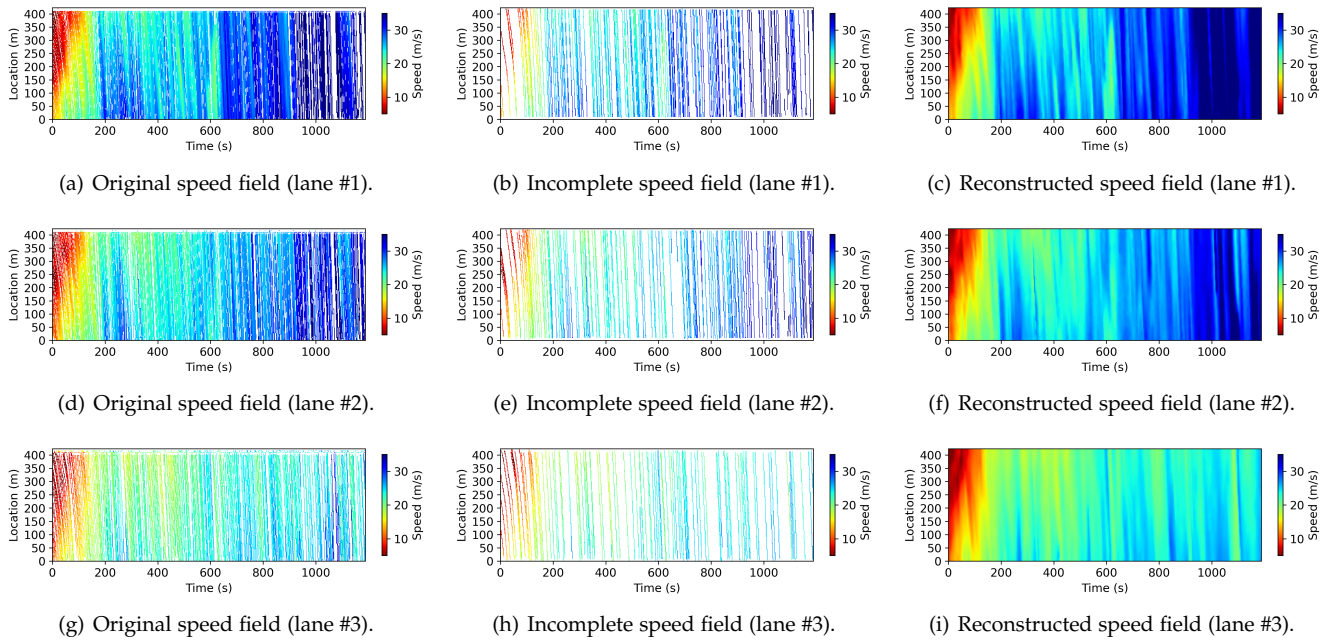


Fig. 8: Speed field reconstruction achieved by LCR-2D on 70% masked trajectories of the HighD data. The speed fields with duration of 1,190 seconds are collected from the road segment of length 426 meters.

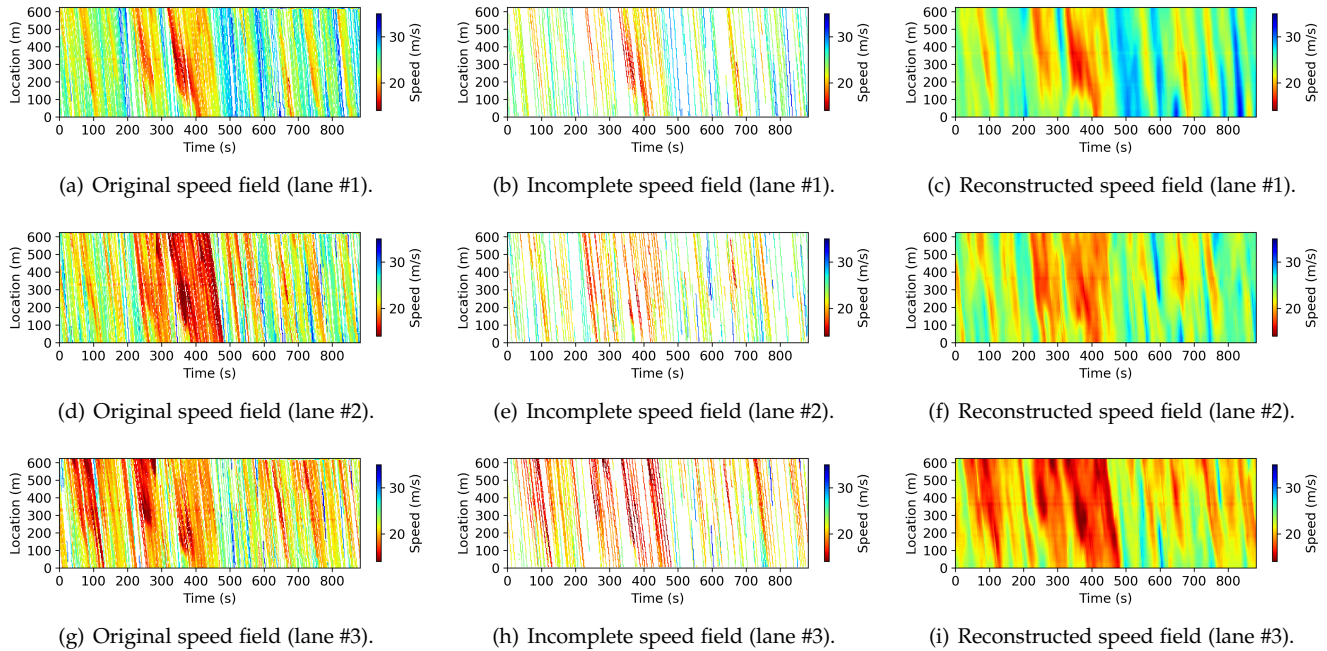


Fig. 9: Speed field reconstruction achieved by LCR-2D on 70% masked trajectories of the CitySim data. The speed fields with duration of 884 seconds are collected from the road segment of length 630 meters.

TABLE 2: Imputation performance (MAPE (%)/RMSE) on the PeMS-4W traffic speed dataset. Note that the best results are emphasized in bold fonts, and we underline the second-best results.

Rate	LCR-2D	LCR <sub>N</sub>	LCR	CTNNM	CircNNM [11]	LRMC [6]	HaLRTC [41]	LRTC-TNN [42]	NoTMF [17]
30%	<u>1.50</u> / <b>1.49</b>	<b>1.48</b> / <u>1.50</u>	<u>1.50</u> / <b>1.49</b>	2.26/1.84	2.26/1.84	2.04/1.80	1.98/1.73	1.68/1.55	2.95/2.65
50%	<b>1.76</b> / <b>1.69</b>	<b>1.73</b> / <u>1.73</u>	<u>1.76</u> / <b>1.69</b>	2.67/2.14	2.69/2.15	2.43/2.12	2.22/1.98	1.93/1.77	3.05/2.73
70%	<b>2.07</b> / <b>2.06</b>	<b>2.07</b> /2.12	<u>2.08</u> / <u>2.07</u>	3.40/2.66	3.43/2.67	3.08/2.66	2.84/2.49	2.33/2.14	3.33/2.97
90%	<b>3.19</b> / <b>3.05</b>	3.24/3.22	<u>3.21</u> / <u>3.06</u>	5.22/3.90	5.34/3.96	6.05/4.43	4.39/3.66	3.40/3.10	5.22/4.71



cient temporal matrix factorization framework, but as can be seen, it is inferior to the LCR models.

## 7 CONCLUSION

In this study, we focus on reconstructing spatiotemporal traffic data from partial observations. To model the local trends in traffic time series, we introduce a Laplacian kernel for temporal regularization in the form of circular convolution. Following that definition, we propose an LCR model that integrates the temporal regularization into a circulant-matrix-based low-rank model for characterizing both global and local trends in traffic time series, bridging the gap between low-rank models and graph Laplacian models. When developing the solution algorithm, we borrow the properties of circulant matrix and circular convolution, and prove that our LCR model has a fast implementation with FFT. Specifically, the nuclear norm minimization with Laplacian kernelized temporal regularization can be converted into an  $\ell_1$ -norm minimization in complex space. Beyond univariate time series imputation, LCR can be easily adapted to multivariate or even multidimensional time series imputation.

In the numerical experiments, we conduct both univariate and multivariate time series imputation tasks on several real-world traffic flow datasets. On the sparse and noisy traffic data, LCR can accurately reconstruct traffic time series with the elimination of data noises and reinforce local time series trends, demonstrating the importance of local trend modeling. On the speed field reconstruction task, the results demonstrate the importance of spatiotemporal modeling with Laplacian kernels. On the large-scale dataset, LCR outperforms the baseline models and demonstrates strong generalization to high-dimensional problems due to the efficient implementation and relatively low time complexity. Despite the great success of temporal modeling in LCR, the key idea lies in connecting the low-rank models and the Laplacian kernelized regularization through FFT, which is also well-suited to some complicated spatiotemporal reconstruction problems. This study provides insight into traffic time series data modeling, nevertheless, the essential idea of LCR also matches the need for time series imputation and forecasting in other domains, while the existing studies (e.g., [11], [12]) already discussed the applications of circulant/convolution matrix-based methods to various time series data.

## ACKNOWLEDGMENT

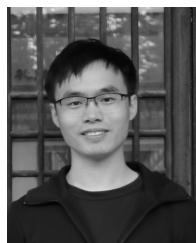
Xinyu Chen would like to thank the Institute for Data Valorisation (IVADO) and the Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT) for providing the PhD Excellence Scholarship to support this study. The work of HanQin Cai is partially supported by NSF DMS 2304489.

## REFERENCES

- [1] X. Chen and L. Sun, "Bayesian temporal factorization for multidimensional time series prediction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 4659–4673, 2021.
- [2] X. Miao, Y. Wu, J. Wang, Y. Gao, X. Mao, and J. Yin, "Generative semi-supervised learning for multivariate time series imputation," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 10, 2021, pp. 8983–8991.
- [3] Y. Wu, D. Zhuang, A. Labbe, and L. Sun, "Inductive graph neural networks for spatiotemporal kriging," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 5, 2021, pp. 4478–4485.
- [4] L. Li, X. Su, Y. Zhang, Y. Lin, and Z. Li, "Trend modeling for traffic time series analysis: An integrated study," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3430–3439, 2015.
- [5] E. J. Candes and Y. Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, 2010.
- [6] J.-F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [7] T. Yokota, B. Erem, S. Guler, S. K. Warfield, and H. Hontani, "Missing slice recovery for tensors using a low-rank model in embedded space," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8251–8259.
- [8] J.-F. Cai, T. Wang, and K. Wei, "Fast and provable algorithms for spectrally sparse signal reconstruction via low-rank hankel matrix completion," *Applied and Computational Harmonic Analysis*, vol. 46, no. 1, pp. 94–121, 2019.
- [9] H. Cai, J.-F. Cai, T. Wang, and G. Yin, "Accelerated structured alternating projections for robust spectrally sparse signal recovery," *IEEE Transactions on Signal Processing*, vol. 69, pp. 809–821, 2021.
- [10] R. Yamamoto, H. Hontani, A. Imakura, and T. Yokota, "Fast algorithm for low-rank tensor completion in delay-embedded space," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 2058–2066.
- [11] G. Liu and W. Zhang, "Recovery of future data via convolution nuclear norm minimization," *IEEE Transactions on Information Theory*, 2022.
- [12] G. Liu, "Time series forecasting via learning convolutionally low-rank models," *IEEE Transactions on Information Theory*, vol. 68, no. 5, pp. 3362–3380, 2022.
- [13] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 8, pp. 1548–1560, 2010.
- [14] X. Mao, K. Qiu, T. Li, and Y. Gu, "Spatio-temporal signal recovery based on low rank and differential smoothness," *IEEE Transactions on Signal Processing*, vol. 66, no. 23, pp. 6281–6296, 2018.
- [15] L. Xiong, X. Chen, T.-K. Huang, J. Schneider, and J. G. Carbonell, "Temporal collaborative filtering with bayesian probabilistic tensor factorization," in *Proceedings of the 2010 SIAM international conference on data mining*. SIAM, 2010, pp. 211–222.
- [16] H.-F. Yu, N. Rao, and I. S. Dhillon, "Temporal regularized matrix factorization for high-dimensional time series prediction," in *Advances in Neural Information Processing Systems*, 2016, pp. 847–855.
- [17] X. Chen, C. Zhang, X.-L. Zhao, N. Saunier, and L. Sun, "Nonstationary temporal matrix factorization for multivariate time series forecasting," *arXiv preprint arXiv:2203.10651*, 2022.
- [18] X. Chen, Z. He, and L. Sun, "A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation," *Transportation Research Part C: Emerging Technologies*, vol. 98, pp. 73–84, 2019.
- [19] F. Sedighin, A. Cichocki, T. Yokota, and Q. Shi, "Matrix and tensor completion in multiway delay embedded space using tensor train, with application to signal reconstruction," *IEEE Signal Processing Letters*, vol. 27, pp. 810–814, 2020.
- [20] Z. Chen and A. Cichocki, "Nonnegative matrix factorization with temporal smoothness and/or spatial decorrelation constraints," *Laboratory for Advanced Brain Signal Processing, RIKEN, Tech. Rep*, vol. 68, 2005.
- [21] Y. Wang, Y. Zhang, X. Piao, H. Liu, and K. Zhang, "Traffic data reconstruction via adaptive spatial-temporal correlations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 4, pp. 1531–1543, 2018.
- [22] X. Chen, Y. Chen, N. Saunier, and L. Sun, "Scalable low-rank tensor learning for spatiotemporal traffic data imputation," *Transportation research part C: emerging technologies*, vol. 129, p. 103226, 2021.
- [23] X. Chen, M. Lei, N. Saunier, and L. Sun, "Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation," *IEEE Transactions on Intelligent Transportation Systems*, 2021.
- [24] N. Rao, H.-F. Yu, P. K. Ravikumar, and I. S. Dhillon, "Collaborative filtering with graph information: Consistency and scalable methods," *Advances in neural information processing systems*, vol. 28, 2015.



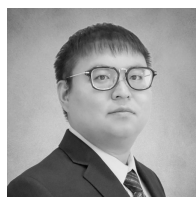
- [25] P. C. Hansen, J. G. Nagy, and D. P. O'leary, *Deblurring images: matrices, spectra, and filtering*. SIAM, 2006.
- [26] J. Wright and Y. Ma, *High-dimensional data analysis with low-dimensional models: Principles, computation, and applications*. Cambridge University Press, 2022.
- [27] S. L. Brunton and J. N. Kutz, *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2022.
- [28] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs: Graph fourier transform," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 6167–6170.
- [29] C. Chen, B. He, Y. Ye, and X. Yuan, "The direct extension of admm for multi-block convex minimization problems is not necessarily convergent," *Mathematical Programming*, vol. 155, no. 1-2, pp. 57–79, 2016.
- [30] J. Yang, W. Yin, Y. Zhang, and Y. Wang, "A fast algorithm for edge-preserving variational multichannel image restoration," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 569–592, 2009.
- [31] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 171–184, 2012.
- [32] E. Candes and B. Recht, "Exact matrix completion via convex optimization," *Communications of the ACM*, vol. 55, no. 6, pp. 111–119, 2012.
- [33] O. Semerci, N. Hao, M. E. Kilmer, and E. L. Miller, "Tensor-based formulation and nuclear norm regularization for multienergy computed tomography," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1678–1693, 2014.
- [34] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 4, pp. 925–938, 2019.
- [35] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.
- [36] E. O. Brigham, *The fast Fourier transform and its applications*. Prentice-Hall, Inc., 1988.
- [37] M. Treiber and A. Kesting, "Traffic flow dynamics," *Traffic Flow Dynamics: Data, Models and Simulation*, Springer-Verlag Berlin Heidelberg, pp. 983–1000, 2013.
- [38] C. De Fabritiis, R. Ragona, and G. Valenti, "Traffic estimation and prediction based on real time floating car data," in *2008 11th international IEEE conference on intelligent transportation systems*. IEEE, 2008, pp. 197–203.
- [39] R. Krajewski, J. Bock, L. Kloecker, and L. Eckstein, "The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems," in *2018 21st international conference on intelligent transportation systems (ITSC)*. IEEE, 2018, pp. 2118–2125.
- [40] O. Zheng, M. Abdel-Aty, L. Yue, A. Abdelraouf, Z. Wang, and N. Mahmoud, "CitySim: A drone-based vehicle trajectory dataset for safety-oriented research and digital twins," *Transportation Research Record: Journal of the Transportation Research Board*, jul 2023.
- [41] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 208–220, 2013.
- [42] X. Chen, J. Yang, and L. Sun, "A nonconvex low-rank tensor completion model for spatiotemporal traffic data imputation," *arXiv preprint arXiv:2003.10271*, 2020.
- [43] X. Wang, Y. Wu, D. Zhuang, and L. Sun, "Low-rank hankel tensor completion for traffic speed estimation," *arXiv preprint arXiv:2105.11335*, 2021.
- [44] C. Chen, K. Petty, A. Skabardonis, P. Varaiya, and Z. Jia, "Freeway performance measurement system: mining loop detector data," *Transportation Research Record*, vol. 1748, no. 1, pp. 96–102, 2001.



**Xinyu Chen** received his Ph.D. degree from the University of Montreal, Montreal, QC, Canada. He is now a Postdoctoral Associate at Massachusetts Institute of Technology, Cambridge, MA, United States. His current research centers on machine learning, spatiotemporal data modeling, intelligent transportation systems, and urban science.



**Zhanhong Cheng** received his Ph.D. degree from McGill University. He received his B.S. and M.S. degrees from Harbin Institute of Technology, Harbin, China. He is now a Postdoc researcher in the Department of Civil Engineering at McGill University, Montreal, QC, Canada. His research interests include public transportation, travel behavior modeling, spatiotemporal forecasting, and machine learning in transportation.



**HanQin Cai** received the PhD degree in applied mathematics and computational sciences from the University of Iowa. He is currently the Paul N. Somerville Endowed assistant professor with the Department of Statistics and Data Science and the Department of Computer Science, University of Central Florida. He is also the director of Data Science Lab. His research interests include machine learning, data science, mathematical optimization, and applied harmonic analysis.



**Nicolas Saunier** received an engineering degree and a Doctorate (Ph.D.) in computer science from Telecom ParisTech, Paris, France, respectively in 2001 and 2005. He is currently a Full Professor with the Civil, Geological and Mining Engineering Department at Polytechnique Montreal, Montreal, QC, Canada. His research interests include intelligent transportation, road safety, and data science for transportation.



**Lijun Sun** (Senior Member, IEEE) received the B.S. degree in civil engineering from Tsinghua University, Beijing, China, in 2011, and the Ph.D. degree in civil engineering (transportation) from the National the University of Singapore in 2015. He is currently an Associate Professor and William Dawson Scholar in the Department of Civil Engineering, McGill University, Montreal, QC, Canada. His research centers on intelligent transportation systems, traffic control and management, spatiotemporal modeling, Bayesian statistics, and agent-based simulation.

Bayesian statistics, and agent-based simulation.