Low-Rank Autoregressive Tensor Completion for Spatiotemporal Traffic Data Imputation

Xinyu Chen, Mengying Lei, Nicolas Saunier, and Lijun Sun, *Member, IEEE*

Abstract—**Spatiotemporal traffic time series (e.g., traffic volume/speed) collected from sensing systems are often incomplete with considerable corruption and large amounts of missing values, preventing users from harnessing the full power of the data. Missing data imputation has been a long-standing research topic and critical application for real-world intelligent transportation systems. A widely applied imputation method is low-rank matrix/tensor completion; however, the low-rank assumption only preserves the global structure while ignores the strong local consistency in spatiotemporal data. In this paper, we propose a low-rank autoregressive tensor completion (LATC) framework by introducing *temporal variation* as a new regularization term into the completion of a third-order (sensor × time of day × day) tensor. The third-order tensor structure allows us to better capture the global consistency of traffic data, such as the inherent seasonality and day-to-day similarity. To achieve local consistency, we design the temporal variation by imposing an autoregressive model for each time series with coefficients as learnable parameters. Different from previous spatial and temporal regularization schemes, the minimization of temporal variation can better characterize temporal generative mechanisms beyond local smoothness, allowing us to deal with more challenging scenarios such as "blackout" missing. To solve the optimization problem in LATC, we introduce an alternating minimization scheme that estimates the low-rank tensor and autoregressive coefficients iteratively. We conduct extensive numerical experiments on several real-world traffic data sets, and our results demonstrate the effectiveness of LATC in diverse missing scenarios.**

Index Terms—**Spatiotemporal traffic data, missing data imputation, low-rank tensor completion, truncated nuclear norm, autoregressive time series model.**

I. INTRODUCTION

SPATIOTEMPORAL traffic data collected from various sensing systems (e.g. loop detectors and floating cars) serve as the foundation to a wide range of applications and decision-making processes in intelligent transportation systems. The emerging "big" data is often large-scale, high-dimensional, and incomplete, posing new challenges to modeling spatiotemporal traffic data [1], [2]. Missing data imputation is one of the most important research questions in spatiotemporal data analysis, since accurate and reliable imputation can help various downstream applications such as traffic forecasting, traffic control, and traffic management. In intelligent transportation systems, the issue of missing and corrupted data may arise from many reasons, including complicated sensor malfunctioning, communication failure, and unsatisfied sensor coverage. Therefore, developing effective missing data imputation approaches is important for getting true signals from partially observed data and improving data quality.

The key to missing data imputation is to efficiently characterize and leverage the complex dependencies and correlations across both spatial and temporal dimensions [3]. Different from point-referenced systems, traffic state data (e.g., speed and flow) is actually an individual time series with a fixed time interval. This allows one to represent spatiotemporal traffic state data in the form of matrix (e.g., sensor × time) or tensor (e.g., sensor × time of day × day) [4], and low-rank matrix/tensor completion becomes a natural solution to solve the imputation problem. Over the past decade, extensive effort has been made on developing low-rank models through principle component analysis, matrix/tensor factorization (with predefined rank) and nuclear norm minimization (see e.g., [4]–[7]). However, the default low-rank structure (e.g., nuclear norm) purely relies on the algebraic property of the data, which is invariant to permutation in the spatial and temporal dimensions. In other words, with the low-rank assumption alone, we essentially overlook the strong "local" spatial and temporal consistency in the data. For instance, we expect traffic flow data collected in a short period to be similar and adjacent sensors to show similar patterns. To this end, some recent studies have tried to encode such "local" consistency by introducing total/quadratic variation and graph regularization as a "smoothness" prior into low-rank factorization models [2], [3], [8]–[10] and imposing time series dynamics on the temporal latent factor in the factorization framework [1], [11], [12]. However, these studies essentially adopt a bilinear or multi-linear factorization model, which requires a predefined rank as a hyperparameter.

In this paper, we propose a low-rank autoregressive tensor completion (LATC) framework to impute missing values in spatiotemporal traffic data. For each completed time series,

we define temporal variation as the accumulated sum of autoregressive errors. To characterize the low-rankness property, we use truncated nuclear norm as an effective tool to avoid the rank determination problem in factorization models. The final objective function of LATC consists of two components, i.e., the truncated nuclear of the completed tensor and the temporal variation defined on the unfolded time series matrix. The combination allows us to effectively characterize both global patterns and local consistency in spatiotemporal traffic data. The overall contribution of this work is threefold:

1) We integrate the autoregressive time series process into a low-rank tensor completion model to capture both global and local trends in spatiotemporal traffic data. By minimizing the truncated nuclear norm of the third-order (sensor×time of day×day) tensor, we can better characterize day-to-day similarity, which is a unique property of traffic time series data [13].

2) We develop an alternating minimization algorithm to update tensor and coefficient matrix separately. The tensor is updated via Alternating Direction Method of Multipliers (ADMM), and the coefficient matrix is updated by least squares with closed-form solution.

3) We conduct extensive numerical experiments on four traffic data sets. Imputation results show the superiority and advantage of LATC over some recent state-of-the-art models.

The remainder of this paper is organized as follows. We introduce related work and notations in Section II and Section III, respectively. Section IV introduces in detail the proposed LATC model. In Section V, we conduct extensive experiments on some traffic data sets and make comparison with some baseline models. Finally, we summarize the study and provide future directions in Section VI.

II. RELATED WORK

There are two types of low-rank models to solve the spatiotemporal missing data imputation problem.

Temporal matrix factorization. Factorization models approximate the complete spatiotemporal matrix/tensor using bilinear or multi-linear factorization models with a predefined rank. To encode temporal consistency, recent studies have introduced local smoothness and time series dynamics to regularize the temporal factor (see e.g., [1], [2], [8], [11], [12]). The introduction of generative mechanism (e.g., autoregressive model) not only offers better interpolation/imputation accuracy, but also enables the factorization models to perform forecasting. However, a major limitation of these models is that they often require careful tuning and selection of the rank.

Tensor representation. Another approach is to fold a time series matrix into a third-order tensor (sensor × time of day × day) by introducing an additional "day" dimension (e.g., [6], [14], [15]). This is a particular case for traffic data given the clear day-to-day similarity, but many real-world time series data resulted from human behavior/activities (e.g., energy/electricity consumption) also exhibit similar patterns. It is expected that the third-order tensor representation captures more information, given that the multivariate time series

matrix is in fact one of the unfoldings of the third-order tensor. As a result, the tensor structure not only preserves the dependencies among sensors but also provides an alternative to capture both local and global temporal patterns (e.g., traffic speed data at 9:00 am on Monday might be similar to that of 9:00 am on Tuesday). These tensor-based models have shown superior performance over matrix-based models in missing data imputation tasks.

Overall, low-rank matrix and tensor models aim at capturing both temporal dynamics and spatial consistency for spatiotemporal traffic data. Despite the wide use of low-rank matrix and tensor imputation approaches, there are also some classical imputation approaches, including time series autoregression-based methods [16], k-nearest neighbor-based methods [17], [18], Gaussian process-based methods [19]. However, compared to low-rank matrix/tensor models, these models are limited for characterizing the spatial dependencies, time series similarity, and temporal dynamics.

III. NOTATIONS

Throughout this work, we use boldface uppercase letters to denote matrices, e.g., $X \in \mathbb{R}^{M \times N}$, boldface lowercase letters to denote vectors, e.g., $x \in \mathbb{R}^M$, and lowercase letters to denote scalars, e.g., x. Given a matrix $X \in \mathbb{R}^{M \times N}$, we denote the (m, n)th entry in X by $x_{m,n}$, and use $x_{m,[t+1:]} \in \mathbb{R}^{(N-t)}$ to denote the sub-vector that consists of the last $N - t$ entries of $x_m \in \mathbb{R}^N$. The Frobenius norm of X is defined as $\|X\|_F = \sqrt{\sum_{m,n} x_{m,n}^2}$, and the ℓ_2-norm of x is defined as $\|x\|_2 = \sqrt{\sum_m x_m^2}$. We denote a third-order tensor by $\mathcal{X} \in \mathbb{R}^{M \times I \times J}$ and the kth-mode ($k = 1, 2, 3$) unfolding of $\mathcal{X}$ by $\mathcal{X}_{(k)}$ [20]. Correspondingly, the folding operator $\text{fold}_k(\cdot)$ converts a matrix to a third-order tensor in the kth-mode. Thus, we have $\text{fold}_k(\mathcal{X}_{(k)}) = \mathcal{X}$ for any tensor $\mathcal{X}$. For $\mathcal{X} \in \mathbb{R}^{M \times I \times J}$, its Frobenius norm is defined as $\|\mathcal{X}\|_F = \sqrt{\sum_{m,i,j} x_{m,i,j}^2}$ and its inner product with another tensor is given by $\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{m,i,j} x_{m,i,j} y_{m,i,j}$ where $\mathcal{Y}$ and $\mathcal{X}$ are of the same size.

IV. METHODOLOGY

A. Tensorization for Global Consistency

We denote the true spatiotemporal traffic data collected from M sensors over J days by Y, whose columns correspond to time points and rows correspond to sensors:

$$Y = \begin{bmatrix} | & | & & | \\ y_1 & y_2 & \cdots & y_{IJ} \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{M \times (IJ)}, \quad (1)$$

where I is the number of time points per day. In the form of matrix, the data has IJ columns in total. The partially observed matrix can be written as $\mathcal{P}_\Omega(Y)$ with observed entries supported on the observed index set Ω:

$$[\mathcal{P}_\Omega(Y)]_{m,n} = \begin{cases} y_{m,n}, & \text{if } (m, n) \in \Omega, \\ 0, & \text{otherwise,} \end{cases}$$

where $m = 1, \ldots, M$ and $n = 1, \ldots, IJ$.

We next introduce the forward tensorization operator $\mathcal{Q}(\cdot)$ that converts the multivariate time series matrix into a third-order tensor. Temporal dimension of traffic time series is divided into two dimensions, i.e., time of day and day. Formally, a third-order tensor can be generated by the forward tensorization operator as $\boldsymbol{\mathcal{X}} = \mathcal{Q}(Y) \in \mathbb{R}^{M \times I \times J}$. Conversely, the resulted tensor can also be converted into the original matrix by $Y = \mathcal{Q}^{-1}(\boldsymbol{\mathcal{X}}) \in \mathbb{R}^{M \times (IJ)}$ where $\mathcal{Q}^{-1}(\cdot)$ denotes the inverse operator of $\mathcal{Q}(\cdot)$.

The tensorization step transforms matrix-based imputation problem to a low-rank tensor completion problem. Global consistency can be achieved by minimizing tensor rank. In practice, tensor rank is often approximated using sum of nuclear norms $\|\boldsymbol{\mathcal{X}}\|_*$ [21] or truncated nuclear norms $\|\boldsymbol{\mathcal{X}}\|_{r,*}$ [6], where r is a truncation parameter (see section IV-C). Our motivation for doing so is that the spatiotemporal traffic data can be characterized by both long-term global trends and short-term local trends. The long-term trends refer to certain periodic, seasonal, and cyclical patterns. Traffic flow data over 24 hours on a typical weekday often shows a systematic "M" shape resulted from travelers' behavioral rhythms, with two peaks during morning and evening rush hours [22]. The pattern also exists at the weekly level with substantial differences from weekdays to weekends. The short-term trends capture certain temporary volatility/perturbation that deviates from the global patterns (e.g., due to incident or special event). The short-term trends seem to be more "random", but they are common and ubiquitous in reality. LATC leverages both global and local patterns by using matrix and tensor representations simultaneously.

B. Temporal Variation for Local Consistency

We define temporal variation of a time series matrix Z given a coefficient matrix $A \in \mathbb{R}^{M \times d}$ and a time lag set $\mathcal{H} = \{h_1, \ldots, h_d\}$ as

$$\|Z\|_{A,\mathcal{H}} = \sum_{m,t}\left(z_{m,t} - \sum_i a_{m,i} z_{m,t-h_i}\right)^2. \tag{2}$$

As can be seen, $\|Z\|_{A,\mathcal{H}}$ quantifies the total squared error when fitting each individual time series z_m with an autoregressive model with coefficient a_m. Given an estimated A, minimizing the temporal variation will encourage the time series data Z to show stronger temporal consistency. In other words, the multivariate time series matrix Z will be better explained by a series of autoregressive models parameterized by A. It should be noted that both Z and A are variables in the proposed temporal variation term.

C. Low-Rank Autoregressive Tensor Completion (LATC)

To ensure both global consistency and local consistency, we propose LATC as the following optimization model

$$\min_{\boldsymbol{\mathcal{X}},Z,A} \quad \|\boldsymbol{\mathcal{X}}\|_{r,*} + \frac{\lambda}{2}\|Z\|_{A,\mathcal{H}}$$
$$\text{s.t.} \quad \begin{cases} \boldsymbol{\mathcal{X}} = \mathcal{Q}(Z), \\ \mathcal{P}_\Omega(Z) = \mathcal{P}_\Omega(Y), \end{cases} \tag{3}$$

where $Y \in \mathbb{R}^{M \times (IJ)}$ is the partially observed time series matrix. $r \in \mathbb{N}_+$ is the truncation which satisfies $r < \min\{M, I, J\}$.

The formulation of LATC ensures both global consistency and local consistency by combining truncated nuclear norm minimization with temporal variation minimization. The weight parameter λ in the objective function controls the trade-off between truncated nuclear norm and temporal variation. Fig. 1 shows that Y can be reconstructed with both low-rank properties and time series dynamics because the constraint in (3), i.e., $\boldsymbol{\mathcal{X}} = \mathcal{Q}(Z)$, is closely related to the partially observed matrix Y.

Most nuclear norm-based tensor completion models employ the Alternating Direction Method of Multipliers (ADMM) algorithm to solve the optimization problem. However, due to the introduction of autoregression coefficient matrix, we can no longer apply the default ADMM algorithm to solve the optimization problem (3). Here we consider applying an alternating minimization scheme by separating the original optimization into two subproblems. Starting with some given initial values $(\boldsymbol{\mathcal{X}}^0, Z^0, A^0)$, we can update $\{(\boldsymbol{\mathcal{X}}^\ell, Z^\ell, A^\ell)\}_{\ell \in \mathbb{N}}$ by solving the two subproblems in an iterative manner. In the implementation, we first fix A^ℓ and solve the following problem to update the variables $\boldsymbol{\mathcal{X}}^{\ell+1}$ and $Z^{\ell+1}$:

$$\boldsymbol{\mathcal{X}}^{\ell+1}, Z^{\ell+1} := \arg\min_{\boldsymbol{\mathcal{X}},Z} \|\boldsymbol{\mathcal{X}}\|_{r,*} + \frac{\lambda}{2}\|Z\|_{A^\ell,\mathcal{H}}$$
$$\text{s.t.} \quad \begin{cases} \boldsymbol{\mathcal{X}} = \mathcal{Q}(Z), \\ \mathcal{P}_\Omega(Z) = \mathcal{P}_\Omega(Y). \end{cases} \tag{4}$$

where ℓ denotes the count of iteration in the alternating minimization scheme. Then, we fix $Z^{\ell+1}$ and solve the following least squares problem to estimate the coefficient matrix $A^{\ell+1}$:

$$A^{\ell+1} := \arg\min_A \|Z^{\ell+1}\|_{A,\mathcal{H}}. \tag{5}$$

When A^ℓ is fixed, the subproblem in Eq. (4) becomes a general low-rank tensor completion problem, it can be solved using ADMM in a similar way as in [21] and [23]. The augmented Lagrangian function of the optimization in Eq. (4) can be written as

$$\mathcal{L}(\boldsymbol{\mathcal{X}}, Z, A^\ell, \boldsymbol{\mathcal{T}}) = \|\boldsymbol{\mathcal{X}}\|_{r,*} + \frac{\lambda}{2}\|Z\|_{A^\ell,\mathcal{H}}$$
$$+ \frac{\rho}{2}\|\boldsymbol{\mathcal{X}} - \mathcal{Q}(Z)\|_F^2$$
$$+ \langle \boldsymbol{\mathcal{X}} - \mathcal{Q}(Z), \boldsymbol{\mathcal{T}} \rangle, \tag{6}$$

where ρ is the weight parameter of the additional Frobenius norm penalty and $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{M \times I \times J}$ is the dual variable. In particular, we keep $\mathcal{P}_\Omega(Z) = \mathcal{P}_\Omega(Y)$ as a fixed constraint to maintain observation consistency. According to the augmented Lagrangian function, ADMM can transform the problem in Eq. (4) into the following subproblems in an iterative manner:

$$\boldsymbol{\mathcal{X}}^{\ell+1,\nu+1} := \arg\min_{\boldsymbol{\mathcal{X}}} \mathcal{L}(\boldsymbol{\mathcal{X}}, Z^{\ell+1,\nu}, A^\ell, \boldsymbol{\mathcal{T}}^{\ell+1,\nu}), \tag{7}$$

$$Z^{\ell+1,\nu+1} := \arg\min_Z \mathcal{L}(\boldsymbol{\mathcal{X}}^{\ell+1,\nu+1}, Z, A^\ell, \boldsymbol{\mathcal{T}}^{\ell+1,\nu}), \tag{8}$$

$$\boldsymbol{\mathcal{T}}^{\ell+1,\nu+1} := \boldsymbol{\mathcal{T}}^{\ell+1,\nu} + \rho(\boldsymbol{\mathcal{X}}^{\ell+1,\nu+1} - \mathcal{Q}(Z^{\ell+1,\nu+1})), \tag{9}$$

where ν denotes the count of iteration in the ADMM. In the following, we discuss the detailed solutions to Eqs. (7) and (8).
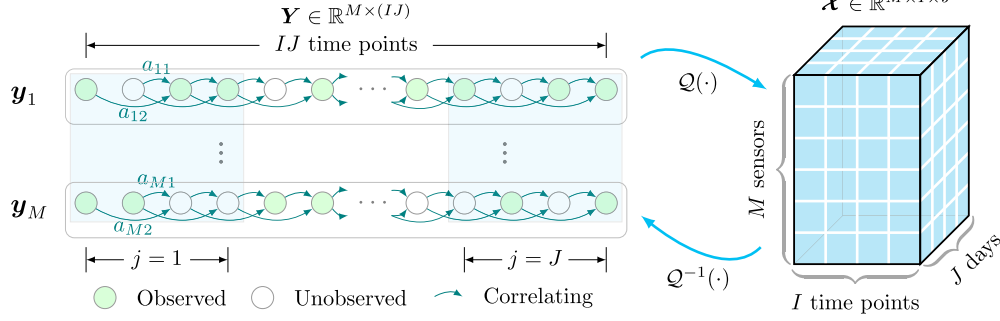
Fig. 1.   Illustration of the proposed LATC framework for spatiotemporal traffic data imputation with time lags $\mathcal{H} = \{1, 2\}$. Each time series $\boldsymbol{y}_m, \forall m \in \{1, 2, \ldots, M\}$ is modeled by the autoregressive coefficients $\{a_{m1}, a_{m2}\}$.

1) Update Variable $\boldsymbol{\mathcal{X}}$: The optimization over $\boldsymbol{\mathcal{X}}$ is a truncated nuclear norm minimization problem. Truncated nuclear norm of any given tensor is the weighted sum of truncated nuclear norm on the unfolding matrices of the tensor, which takes the form:

$$\|\boldsymbol{\mathcal{X}}\|_{r,*} = \sum_{k=1}^{3} \alpha_k \|\boldsymbol{\mathcal{X}}_{(k)}\|_{r,*} \qquad (10)$$

for tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{M \times I \times J}$ with $\sum_{k=1}^{3} \alpha_k = 1$. For the minimization of truncated nuclear norm on tensor, the above formula is not in its appropriate form because unfolding a tensor in different modes cannot guarantee the dependencies of variables [21]. Therefore, we introduce $\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{X}}_2, \boldsymbol{\mathcal{X}}_3$ and they correspond to the unfoldings of $\boldsymbol{\mathcal{X}}$. Accordingly, it is possible to obtain the closed-form solution for each $\boldsymbol{\mathcal{X}}_k$:

$$
\begin{aligned}
\boldsymbol{\mathcal{X}}_k :=& \arg\min_{\boldsymbol{\mathcal{X}}} \; \alpha_k \|\boldsymbol{\mathcal{X}}_{(k)}\|_{r,*} + \frac{\rho}{2} \left\| \mathcal{Q}^{-1}(\boldsymbol{\mathcal{X}}) - \boldsymbol{Z}^{\ell+1,v} \right\|_F^2 \\
&+ \langle \mathcal{Q}^{-1}(\boldsymbol{\mathcal{X}}) - \boldsymbol{Z}^{\ell+1,v}, \mathcal{Q}^{-1}(\boldsymbol{\mathcal{T}}^{\ell+1,v}) \rangle \\
=& \arg\min_{\boldsymbol{\mathcal{X}}} \; \alpha_k \|\boldsymbol{\mathcal{X}}_{(k)}\|_{r,*} \\
&+ \frac{\rho}{2} \left\| \boldsymbol{\mathcal{X}} - \left(\mathcal{Q}(\boldsymbol{Z}^{\ell+1,v}) - \boldsymbol{\mathcal{T}}^{\ell+1,v}/\rho \right) \right\|_F^2 \\
=& \; \text{fold}_k \left(\mathcal{D}_{r,\alpha_k/\rho} \left(\mathcal{Q}(\boldsymbol{Z}^{\ell+1,v})_{(k)} - \boldsymbol{\mathcal{T}}_{(k)}^{\ell+1,v}/\rho \right) \right), \quad (11)
\end{aligned}
$$

where $\mathcal{D}.(\cdot)$ denotes the generalized singular value thresholding that associated with truncated nuclear norm minimization as shown in Lemma 1.

Lemma 1: For any $\alpha, \rho > 0$, $\boldsymbol{Z} \in \mathbb{R}^{m \times n}$, and $r \in \mathbb{N}_+$ where $r < \min\{m, n\}$, an optimal solution to the truncated nuclear norm minimization problem

$$\min_{X} \; \alpha \|X\|_{r,*} + \frac{\rho}{2} \|X - Z\|_F^2, \qquad (12)$$

is given by the generalized singular value thresholding [24]–[26]:

$$\hat{X} = \mathcal{D}_{r,\alpha/\rho}(\boldsymbol{Z}) = \boldsymbol{U} \, \text{diag}\left([\boldsymbol{\sigma} - \mathbb{1}_r \cdot \alpha/\rho]_+ \right) \boldsymbol{V}^\top, \quad (13)$$

where $\boldsymbol{U} \, \text{diag}(\boldsymbol{\sigma}) \boldsymbol{V}^\top$ is the SVD of $\boldsymbol{Z}$. $[\cdot]_+$ denotes the positive truncation at 0 which satisfies $[\boldsymbol{\sigma} - \alpha/\rho]_+ = \max\{\boldsymbol{\sigma} - \alpha/\rho, 0\}$. $\mathbb{1}_r \in \{0, 1\}^{\min\{m,n\}}$ is a binary indicator vector whose first r entries are 0 and other entries are 1.

Gathering the results of $\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{X}}_2, \boldsymbol{\mathcal{X}}_3$ in Eq. (11), we can update the variable $\boldsymbol{\mathcal{X}}$ by

$$\boldsymbol{\mathcal{X}}^{\ell+1,v+1} := \sum_{k=1}^{3} \alpha_k \boldsymbol{\mathcal{X}}_k. \qquad (14)$$

2) Update Variable $\boldsymbol{Z}$: Given that $\boldsymbol{\mathcal{X}} = \mathcal{Q}(\boldsymbol{Z})$, we can rewrite Eq. (8) with respect to $\boldsymbol{Z}$ as follows,

$$
\begin{aligned}
\boldsymbol{Z}^{\ell+1,v+1} :=& \arg\min_{\boldsymbol{Z}} \; \frac{\lambda}{2} \|\boldsymbol{Z}\|_{A^\ell, \mathcal{H}} + \frac{\rho}{2} \left\| \boldsymbol{\mathcal{X}}^{\ell+1,v+1} - \mathcal{Q}(\boldsymbol{Z}) \right\|_F^2 \\
&- \langle \mathcal{Q}(\boldsymbol{Z}), \boldsymbol{\mathcal{T}}^{\ell+1,v} \rangle \\
=& \arg\min_{\boldsymbol{Z}} \; \frac{\lambda}{2} \|\boldsymbol{Z}\|_{A^\ell, \mathcal{H}} \\
&+ \frac{\rho}{2} \left\| \boldsymbol{Z} - \mathcal{Q}^{-1}(\boldsymbol{\mathcal{X}}^{\ell+1,v+1} + \boldsymbol{\mathcal{T}}^{\ell+1,v}/\rho) \right\|_F^2 .
\end{aligned}
$$
$$\qquad (15)$$

We use the following Lemma 2 to solve this optimization problem.

Lemma 2: For any multivariate time series $\boldsymbol{Z} \in \mathbb{R}^{M \times T}$ which consists of M time series over T consecutive time points, the autoregressive process for any (m, t)th element of $\boldsymbol{Z}$ takes

$$z_{m,t} \approx \sum_{i=1}^{d} a_{m,i} z_{m,t-h_i}, \qquad (16)$$

with autoregressive coefficient $\boldsymbol{A} \in \mathbb{R}^{M \times d}$ and time lag set $\mathcal{H} = \{h_1, h_2, \ldots, h_d\}$. This autoregressive process also takes the following general formula:

$$\boldsymbol{\Psi}_0 \boldsymbol{Z}^\top \approx \sum_{i=1}^{d} \boldsymbol{\Psi}_i (\boldsymbol{a}_i^\top \odot \boldsymbol{Z}^\top) = \boldsymbol{\Psi}(\boldsymbol{A}^\top \odot \boldsymbol{Z}^\top), \qquad (17)$$

and for each time series $\boldsymbol{z}_m \in \mathbb{R}^T$, $\forall m$, we have

$$\boldsymbol{\Psi}_0 \boldsymbol{z}_m \approx \sum_{i=1}^{d} a_{m,i} \boldsymbol{\Psi}_i \boldsymbol{z}_m, \qquad (18)$$

where $\odot$ denotes the Khatri-Rao product, and

$$
\begin{aligned}
\boldsymbol{\Psi}_0 &= \begin{bmatrix} \boldsymbol{0}_{(T-h_d) \times h_d} & \boldsymbol{I}_{T-h_d} \end{bmatrix} \in \mathbb{R}^{(T-h_d) \times T}, \\
\boldsymbol{\Psi}_i &= \begin{bmatrix} \boldsymbol{0}_{(T-h_d) \times (h_d - h_i)} & \boldsymbol{I}_{T-h_d} & \boldsymbol{0}_{(T-h_d) \times h_i} \end{bmatrix} \\
&\quad \in \mathbb{R}^{(T-h_d) \times T}, i = 1, 2, \ldots, d, \\
\boldsymbol{\Psi} &= \begin{bmatrix} \boldsymbol{\Psi}_1 & \boldsymbol{\Psi}_2 & \cdots & \boldsymbol{\Psi}_d \end{bmatrix} \in \mathbb{R}^{(T-h_d) \times (dT)},
\end{aligned}
$$

are matrices defined based on time lag set $\mathcal{H}$. Without loss of generality, $\mathbf{0}_{m \times n}$ denotes the m-by-n matrix of zeros, and $\boldsymbol{I}_m$ denotes the m-by-m identity matrix.

According to Lemma 2, there are two options for updating $\boldsymbol{Z}$ when $\boldsymbol{A}$ and $\mathcal{H}$ are known. The first is to minimize the errors in the form of matrix as described in Eq. (17), and the second is to minimize the errors in the form of vector as described in Eq. (18). The first solution involves complicated operations and possibly high computational cost (see Theorem 1 in Appendix A for details). We follow the second approach which takes the vector form for optimizing $\boldsymbol{Z}$. This yields a closed-form solution in Lemma 3.

Lemma 3: Suppose $\boldsymbol{\Psi}_0, \boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_d \in \mathbb{R}^{(T-h_d) \times T}$ and autoregressive coefficient $\boldsymbol{A} \in \mathbb{R}^{M \times d}$ are known as defined in Lemma 2, then for any $m \in \{1, 2, \ldots, M\}$, an optimal solution to the problem

$$z_m := \arg\min_{z} \frac{1}{2} \left\| \boldsymbol{\Psi}_0 z - \sum_{i=1}^{d} a_{m,i} \boldsymbol{\Psi}_i z \right\|_2^2 + \frac{\alpha}{2} \|z - x_m\|_2^2, \tag{19}$$

is given by

$$z_m := \alpha (\boldsymbol{B}_m^\top \boldsymbol{B}_m + \alpha \boldsymbol{I}_T)^{-1} x_m, \tag{20}$$

where $\boldsymbol{B}_m = \boldsymbol{\Psi}_0 - \sum_{i=1}^{d} a_{m,i} \boldsymbol{\Psi}_i$.

Remark. Lemma 3 in fact provides a least squares solution for z_m. It is also helpful to define $\boldsymbol{B}_m, m = 1, 2, \ldots, M$ as sparse matrices and interpret z_m as the solution of the following linear equation:

$$(\boldsymbol{B}_m^\top \boldsymbol{B}_m + \alpha \boldsymbol{I}_T) z_m = \alpha x_m. \tag{21}$$

This can help avoid the expensive inverse operation on the T-by-T matrix because T is a possibly large value.

According to Lemma 3, for any $m \in \{1, 2, \ldots, M\}$, the closed-form solution to Eq. (15) is given by

$$z_m^{\ell+1,v+1} := \frac{\rho}{\lambda} \left(\boldsymbol{B}_m^\top \boldsymbol{B}_m + \frac{\rho}{\lambda} \boldsymbol{I}_T \right)^{-1} \cdot \mathcal{Q}_m^{-1}(\mathcal{X}^{\ell+1,v+1} + \mathcal{T}^{\ell+1,v}/\rho), \tag{22}$$

where $\boldsymbol{B}_m = \boldsymbol{\Psi}_0 - \sum_{i=1}^{d} a_{m,i}^\ell \boldsymbol{\Psi}_i$ in which $\boldsymbol{\Psi}_0, \boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_d$ follow the same definition as in Lemma 2.

3) Update Variable $\boldsymbol{A}$: As mentioned above, $\boldsymbol{A} \in \mathbb{R}^{M \times d}$ is the coefficient matrix in the defined temporal variation term. To estimate $\boldsymbol{A}$, we solve the following problem derived from Eq. (5):

$$\boldsymbol{A}^{\ell+1} := \arg\min_{\boldsymbol{A}} \sum_{m,t} (z_{m,t}^{\ell+1,K} - \sum_i a_{m,i} z_{m,t-h_i}^{\ell+1,K})^2$$
$$= \arg\min_{\boldsymbol{A}} \sum_m \left\| z_{m,[h_d+1:]}^{\ell+1,K} - \boldsymbol{V}_m \boldsymbol{a}_m \right\|_2^2, \tag{23}$$

where $\boldsymbol{V}_m = (\boldsymbol{v}_{h_d+1}, \cdots, \boldsymbol{v}_{IJ})^\top \in \mathbb{R}^{(IJ-h_d) \times d}$ and $\boldsymbol{v}_t = (z_{m,t-h_1}^{\ell+1,K}, \cdots, z_{m,t-h_d}^{\ell+1,K})^\top \in \mathbb{R}^d, t = h_d+1, \ldots, IJ$ are formed by $\boldsymbol{Z}^{\ell+1,K}$. Here, K is the maximum iteration in the ADMM. Obviously, this optimization has a closed-form solution, which is given by

$$\boldsymbol{a}_m^{\ell+1} := \boldsymbol{V}_m^\dagger z_{m,[h_d+1:]}^{\ell+1,K}, \forall m, \tag{24}$$

where $\cdot^\dagger$ denotes the Moore-Penrose pseudo-inverse.

Algorithm 1 Imputer$(\boldsymbol{Y}, \mathcal{H}, \rho, \lambda, r)$

Initialize $\mathcal{T}^{0,0}$ as zeros and $\boldsymbol{A}^0$ as small random values.
Set $\mathcal{P}_\Omega(\boldsymbol{Z}^{0,0}) = \mathcal{P}_\Omega(\boldsymbol{Y})$, $\alpha_1 = \alpha_2 = \alpha_3 = \frac{1}{3}$, $K = 3$, and $\ell = 0$.
while *not converged* **do**
  for $v = 0$ **to** $K - 1$ **do**
    $\rho = \min\{1.05 \times \rho, \rho_{\max}\}$;
    for $k = 1$ **to** 3 **do**
      Compute $\mathcal{X}_k$ by Eq. (11);
    Update $\mathcal{X}^{\ell+1,v+1}$ by Eq. (14);
    for $m = 1$ **to** M **do**
      Update $z_m^{\ell+1,v+1}$ by Eq. (22);
    Update $\mathcal{T}^{\ell+1,v+1}$ by Eq. (9);
    Transform observation information by letting
    $\mathcal{P}_\Omega(\boldsymbol{Z}^{\ell+1,v+1}) = \mathcal{P}_\Omega(\boldsymbol{Y})$;
  for $m = 1$ **to** M **do**
    Update $\boldsymbol{a}_m^{\ell+1}$ by Eq. (24);
  $\ell := \ell + 1$;
return recovered matrix $\hat{\boldsymbol{X}}$.

Algorithm 1 shows the overall algorithm for solving LATC. The algorithm has three parameters ρ, λ and r. Parameter ρ controls the ADMM and the singular value thresholding. Parameter λ is the trade-off between truncated nuclear norm and temporal variation, which can be typically set to $\lambda = c \cdot \rho$. Thus, $c = 1$ implies that these two norms have the same importance in the objective. The recovered matrix is computed by $\hat{\boldsymbol{X}}^\ell = \mathcal{Q}^{-1}(\mathcal{X}^{\ell,K})$ at each outer iteration. The algorithm returns the converged $\hat{\boldsymbol{X}}$ as the final result, if the convergence criteria is met.

V. EXPERIMENTS

In this section, we evaluate the proposed LATC model on several real-world traffic data sets with different missing patterns.

A. Traffic Data Sets

We use the following four spatiotemporal traffic data sets for our benchmark experiment.

- **(G)**: Guangzhou urban traffic speed data set.[1] This data set contains traffic speed collected from 214 road segments over two months (from August 1 to September 30, 2016) with a 10-minute resolution (i.e., 144 time intervals per day) in Guangzhou, China. The prepared data is of size 214×8784 in the form of multivariate time series matrix (or tensor of size $214 \times 144 \times 61$).
- **(H)**: Hangzhou metro passenger flow data set.[2] This data set provides incoming passenger flow of 80 metro stations over 25 days (from January 1 to January 25, 2019) with a 10-minute resolution in Hangzhou, China. We discard the interval 0:00 a.m. 6:00 a.m. with no services, and

---

[1] https://doi.org/10.5281/zenodo.1205229
[2] https://tianchi.aliyun.com/competition/entrance/231708/information

only consider the remaining 108 time intervals of a day. The prepared data is of size 80×2700 in the form of multivariate time series matrix (or tensor of size $80 \times 108 \times 25$).

- **(S)**: Seattle freeway traffic speed data set.[3] This data set contains freeway traffic speed from 323 loop detectors with a 5-minute resolution (i.e., 288 time intervals per day) over the first four weeks of January, 2015 in Seattle, USA. The prepared data is of size 323×8064 in the form of multivariate time series matrix (or tensor of size $323 \times 288 \times 28$).
- **(P)**: Portland highway traffic volume data set.[4] This data set is collected from highways in the Portland-Vancouver Metropolitan region, which contains traffic volume from 1156 loop detectors with a 15-minute resolution (i.e., 96 time intervals per day) in January, 2021. The prepared data is of size 1156×2976 in the form of multivariate time series matrix (or tensor of size $1156 \times 96 \times 31$).

Note that the adapted data sets and Python codes for our experiments are available on Github.[5]

B. Missing Data Generation

To evaluate the performance of LATC for missing traffic data imputation thoroughly, we take into account three missing data patterns as shown in Fig. 2, i.e., random missing (RM), non-random missing (NM), and blackout missing (BM). RM and NM data are generated by referring to our prior work [4]. According to the mechanism of RM and NM data, we mask certain amount of observations as missing values (e.g., 30%, 70%, 90%), and the remaining partial observations are input data for learning a well-behaved model. BM pattern is different from RM and NM patterns, which masks observations of all spatial sensors/locations as missing values with certain window length. BM is a challenging scenario with complete column-wise missing. We set the missing rate in the following experiments to 30%.

To assess the imputation performance, we use the actual values of the masked missing entries as the ground truth to compute MAPE and RMSE:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100,$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}, \tag{25}$$

where y_i and $\hat{y}_i$ are actual values and imputed values, respectively.

C. Baseline Models

For comparison, we take into account the following baseline:

- Low-Rank Autoregressive Matrix Completion (LAMC). This is a matrix-form variant of the LATC model.

(a) Random missing (RM).

(b) Non-random missing (NM).
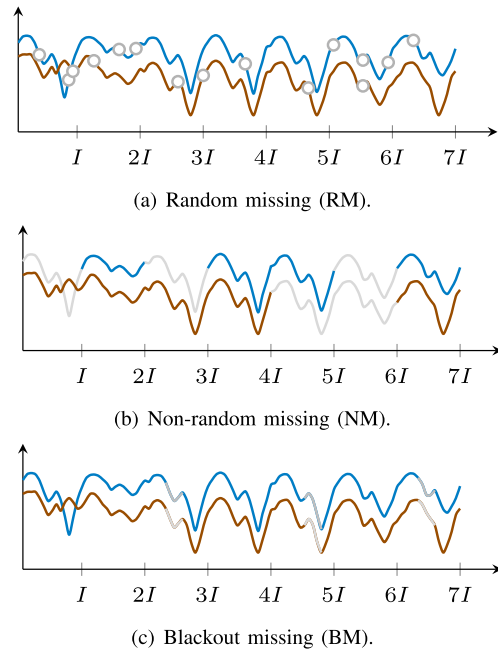

(c) Blackout missing (BM).

Fig. 2.   Illustration of three missing data patterns for spatiotemporal traffic data (e.g., traffic speed). Each time series represent the collected data from a given sensor. In these graphics, two curves correspond to two different time series. (a) Data are missing at random. Small circles indicate the missing values. (b) Data are missing continuously during a few time periods. Segments in gray indicate missing values. (c) No sensors are available (i.e., blackout) over a certain time window.

- Low-Rank Tensor Completion with Truncation Nuclear Norm minimization (LRTC-TNN, [6]). This is a low-rank completion model in which truncated nuclear norm minimization can help maintain the most important low-rank patterns. Since the truncation in LRTC-TNN is a defined as a rate parameter, we adapt LRTC-TNN to use integer truncation in order to make it consistent with LATC.
- Bayesian Temporal Matrix Factorization (BTMF, [1]). This is a fully Bayesian temporal factorization framework which builds the correlation of temporal dynamics on latent factors by vector autoregressive process. Due to the temporal modeling, it outperforms the standard matrix factorization in the missing data imputation tasks [1].
- Smooth PARAFAC Tensor Completion (SPC, [10]). This is a tensor decomposition based completion model with total variation smoothness constraints.

D. Results

There are several parameters in LATC, including learning rate ρ, weight parameter λ, truncation r, and time lag set $\mathcal{H}$. The most important parameters are the coefficient $c = \lambda/\rho$ and the truncation r. For other parameters including ρ and time lag set $\mathcal{H}$, we conduct preliminary test for choosing them. ρ is chosen from $\{1 \times 10^{-5}, 1 \times 10^{-4}\}$ for all data sets. To assess the sensitivity of the model over c and r, we develop the following setting for our imputation experiments:

- Time lag set is set as $\{1, 2, \ldots, 6\}$ for (G), (H), and (S) data, and $\{1, 2, 3, 4\}$ for (P) data;
- $\lambda = c \cdot \rho$ where $c \in \{\frac{1}{10}, \frac{1}{5}, 1, 5, 10\}$;

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

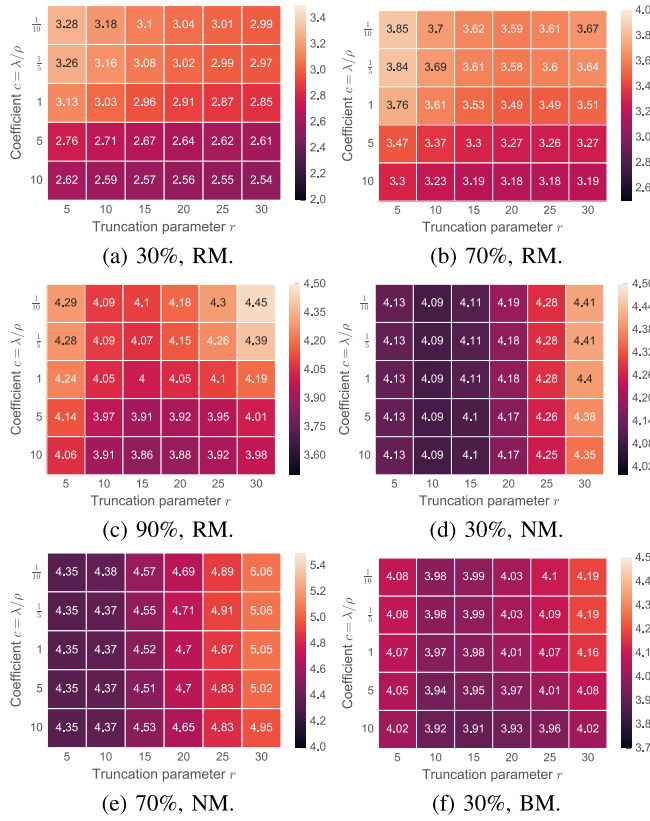CHEN *et al.*: LATC FOR SPATIOTEMPORAL TRAFFIC DATA IMPUTATION

7

Fig. 3. RMSEs of LATC imputation on Guangzhou urban traffic speed data where $\rho = 1 \times 10^{-4}$ for RM data and $\rho = 1 \times 10^{-5}$ for NM/BM data. The smallest RMSE is achieved by: (a) $c = 10, r = 30$; (b) $c = 10, r = 20, 25$; (c) $c = 10, r = 15$; (d) $r = 10$; (e) $r = 5$; (f) $c = 10, r = 15$.
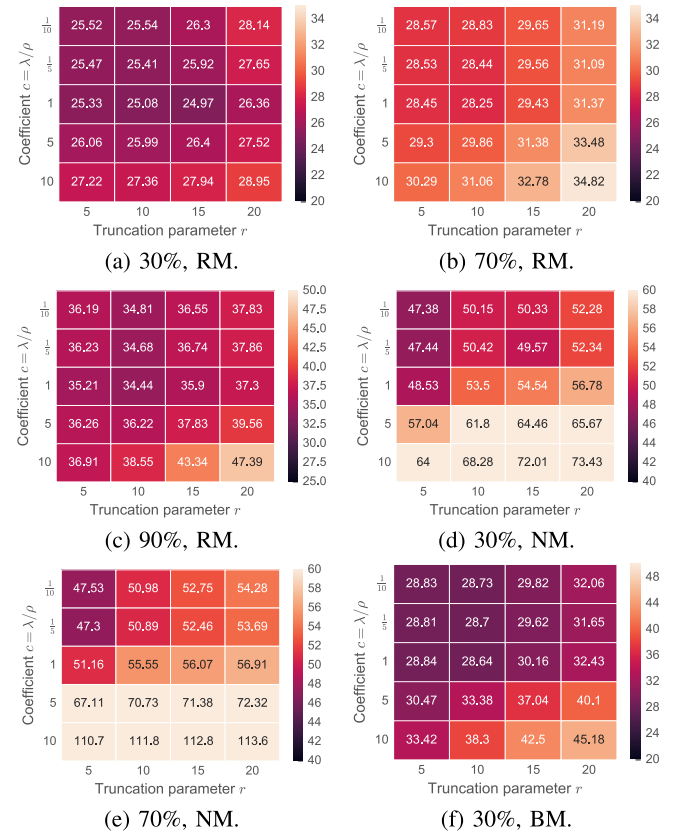

Fig. 4. RMSEs of LATC imputation on Hangzhou metro passenger flow data ($\rho = 1 \times 10^{-5}$). The smallest RMSE is achieved by: (a) $c = 1, r = 15$; (b-c) $c = 1, r = 10$; (d) $c = \frac{1}{10}, r = 5$; (e) $c = \frac{1}{5}, r = 5$; (f) $c = 1, r = 10$.

- $r \in \{5, 10, 15, 20, 25, 30\}$ and $r < \min\{M, I, J\}$.

Fig. 3 shows the heatmaps of imputation RMSE values achieved by LATC model on Guangzhou urban traffic speed data. It demonstrates that: 1) for RM and BM data, when $c = 10$, LATC model achieves the best imputation performance and the truncation r has little impact on the final results; 2) for NM data, the coefficient c is less important than the truncation r. LATC model achieves the best performance when the truncation is a relatively small value (e.g., 5, 10). These results verifies the importance of temporal variation minimization for RM and BM imputation.

Fig. 4 shows similar heatmaps for Hangzhou metro passenger flow data. It can be seen that: 1) for RM and BM data, when $c = 1$, LATC model achieves the best imputation performance; 2) for NM data, LATC model achieves the best performance with small coefficient c and truncation r (e.g., 5).

By testing the LATC model in the same way, it can indicate the importance of temporal variation on other two data sets. On Seattle freeway traffic speed data, we observe that the coefficient c has little impact on the final imputation for the RM and NM data. However, there show the positive influence of temporal variation in LATC for BM data. On Portland highway traffic volume data, a relatively large coefficient c (e.g., 5 and 10) can make the model less sensitive to the various truncation values for RM and BM data.

As mentioned above, despite the use of truncated nuclear norm built on tensor, the results also show the advantage of temporal variation built on the multivariate time series matrix. Due to the temporal modeling, temporal variation can improve the imputation performance for missing traffic data imputation. Table I shows the overall imputation performance of LATC and baseline models on the four selected traffic data sets with various missing scenarios. Of these results, NM and BM data seem to be more difficult to reconstruct with all these imputation models than RM data. In most cases, LATC outperforms other baseline models. Comparing LATC with LAMC shows the advantage of tensor structure, i.e., LATC with tensor structure performs better than LAMC with matrix structure. Comparing LATC with LRTC-TNN shows the advantage of temporal variation, i.e., temporal modeling with autoregressive process has positive influence for improving the imputation performance. For volume data sets (H) and (P), the relative errors are quite high because some volume values are close to 0 or relatively small and estimating these values would accumulate relatively large relative errors.

Figs. 5, 6, and 7 show some imputation examples with different missing scenarios that achieved by LATC. In these examples, we can see explicit temporal dependencies underlying traffic time series data. For all missing scenarios, LATC can achieve accurate imputation and learn the true signals from observations even with severe missing data (e.g., NM/BM data). In Fig. 5, it shows that the time series signal of passenger

TABLE I

PERFORMANCE COMPARISON (IN MAPE/RMSE) OF LATC AND BASELINE MODELS FOR RM, NM, AND BM DATA IMPUTATION

Data	Missing	LATC	LAMC	LRTC-TNN	BTMF	SPC
(G)	30%, RM	**5.71/2.54**	9.51/4.04	6.99/3.00	7.54/3.27	7.37/5.06
	70%, RM	**7.22/3.18**	10.40/4.37	8.38/3.59	8.75/3.73	8.91/4.44
	90%, RM	**9.11/3.86**	11.65/4.79	9.55/4.20	10.02/4.21	10.60/4.85
	30%, NM	9.63/4.09	10.11/4.23	9.61/**4.07**	10.32/4.33	**9.13**/5.29
	70%, NM	10.37/4.35	11.15/4.60	**10.36/4.34**	11.36/4.85	11.15/5.17
	30%, BM-6	**9.23/3.91**	12.15/5.17	9.45/3.97	12.43/7.04	11.14/5.13
(H)	30%, RM	19.12/24.97	22.65/42.94	**18.87/24.90**	22.37/28.66	19.82/26.21
	70%, RM	20.25/28.25	25.30/51.26	**20.07/28.13**	25.65/32.23	21.02/31.91
	90%, RM	24.32/**34.44**	32.30/66.13	**23.46**/35.84	31.51/46.24	24.97/49.68
	30%, NM	**19.93/47.38**	22.93/67.08	19.94/50.12	25.61/77.00	27.46/68.56
	70%, NM	24.30/47.30	29.23/63.95	**23.88/45.06**	34.50/70.11	46.86/98.81
	30%, BM-6	21.93/28.64	30.78/66.03	**21.40/27.83**	52.15/57.61	22.49/37.53
(S)	30%, RM	**4.90/3.16**	5.98/3.73	4.99/3.20	5.91/3.72	5.92/3.62
	70%, RM	**5.96/3.71**	8.02/4.70	6.10/3.77	6.47/3.98	7.38/4.30
	90%, RM	**7.47/4.51**	10.56/5.91	8.08/4.80	8.17/4.81	9.75/5.31
	30%, NM	7.11/4.33	6.99/4.25	**6.85/4.21**	9.26/5.36	8.87/4.99
	70%, NM	9.46/5.42	9.75/6.16	**9.23/5.35**	10.47/6.15	11.32/5.92
	30%, BM-12	**9.44/5.36**	27.05/13.66	9.52/5.41	14.33/13.60	11.30/5.84
(P)	30%, RM	17.46/**15.89**	17.93/16.03	**17.27**/16.08	18.22/19.14	21.29/56.73
	70%, RM	**19.56/18.70**	21.26/19.37	19.99/18.73	19.96/22.21	24.35/43.32
	90%, RM	23.47/22.74	25.64/23.75	**22.90/22.68**	23.90/25.71	28.45/39.65
	30%, NM	**18.90/18.84**	19.93/19.69	19.59/18.91	19.55/20.38	26.96/60.33
	70%, NM	24.67/31.74	25.75/28.25	30.26/60.85	**23.86/26.74**	33.42/47.34
	30%, BM-4	**24.04/23.52**	29.21/27.60	31.74/74.42	27.85/25.68	31.01/60.33

Best results are highlighted in bold fonts. The number next to the BM denotes the window length.
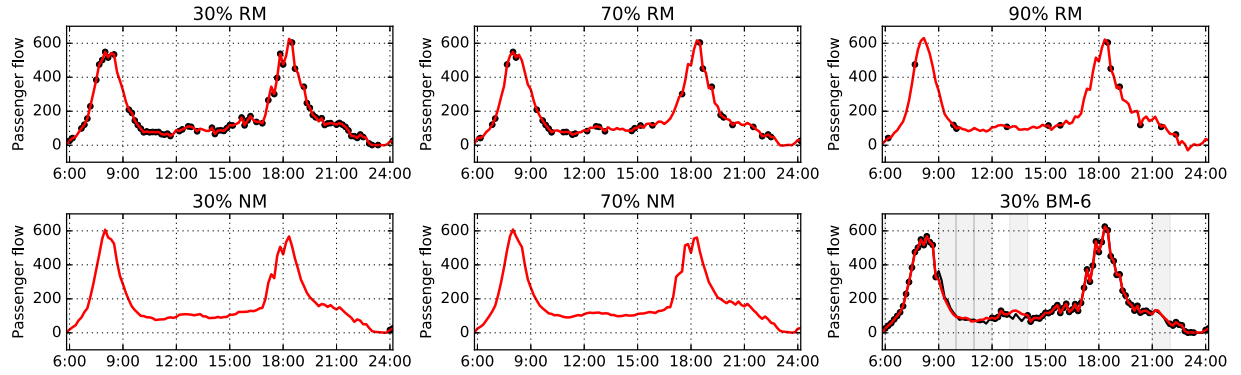

Fig. 5. Imputed values by LATC for Hangzhou metro passenger flow data. This example corresponds to metro station #3 and the 4th day of the data set. Black dots/curves indicate the partially observed data, gray rectangles indicate blackout missing, while red curves indicate the imputed values.
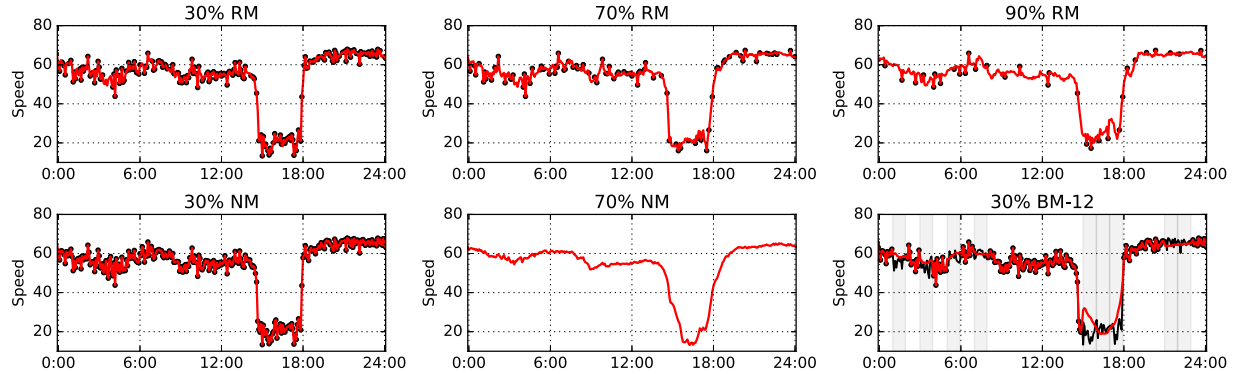

Fig. 6. Imputed values by LATC for Seattle freeway traffic speed data. This example corresponds to detector #3 and the 7th day of the data set.

flow is not complex. By referring to Table I, we can see that LRTC-TNN without temporal variation outperforms the proposed LATC model on Hangzhou metro passenger flow data, and this demonstrates that not all multivariate time series
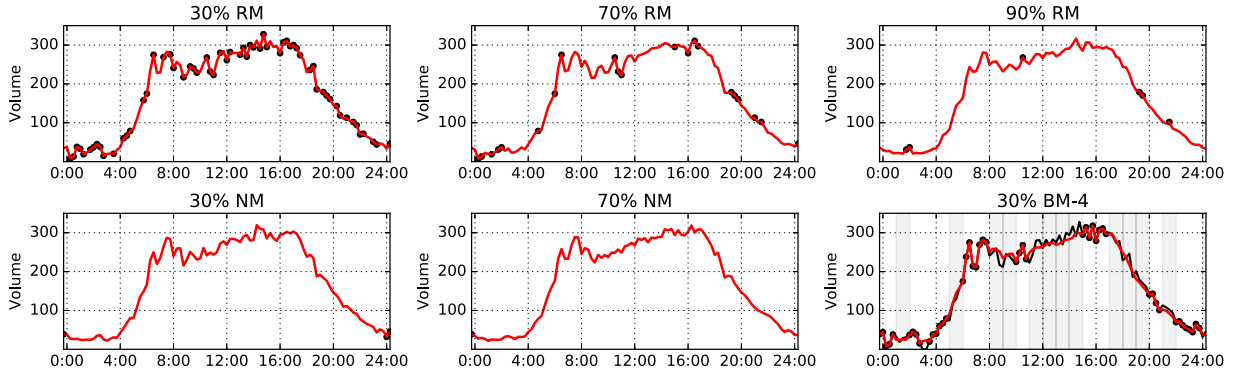
Fig. 7. Imputed values by LATC for Portland traffic volume data. This example corresponds to detector #3 and the 8th day of the data set.

imputation cases require temporal modeling, for some cases that the signal does not show strong temporal dependencies, purely low-rank model can also provide accurate imputation.

VI. CONCLUSION

Spatiotemporal traffic data imputation is of great significance in data-driven intelligent transportation systems. In this study, we leverage some fundamental features such as low-rankness properties and temporal dynamics into the modeling of spatiotemporal traffic data in the presence of missing values. The proposed LATC model preserves both global low-rank structure (i.e., by minimizing the truncated nuclear norm) and local temporal structure (i.e., by minimizing the temporal variation) in multivariate traffic time series data. We perform numerical experiments on some real-world traffic data sets and the results show the advantages of LATC over other low-rank models through the use of temporal variation.

There are several directions to advance this research. First, for simplicity, the proposed temporal variation ignores the correlation structure among different time series and assumes each time series follows an independent univariate autoregressive model. A potential direction is to introduce a multivariate scheme (e.g., vector autoregressive process) to better characterize the correlation structure. The main challenge here is how to encode such dependency when the number of time series M becomes large. Second, the current modeling framework assumes that there exist no noise/outliers in the observed data, which may undermines the underlying low-rank assumption of the data. To address this issue, one can extend the LATC based on Noisy Low-Rank Tensor Completion [26] by introducing an unknown error term. Third, in addition to the imputation capability of LATC, LATC can also be applied to spatiotemporal traffic forecasting in the presence of missing values. The forecasting can be obtained by considering future observations as missing values following [27], [28].

APPENDIX A
SUPPLEMENTARY THEOREM

Theorem 1: Suppose $\mathbf{\Phi}_0 \in \mathbb{R}^{(T-h_d)\times T}$, $\mathbf{\Phi} \in \mathbb{R}^{(T-h_d)\times(dT)}$, and autoregressive coefficient $\mathbf{A} \in \mathbb{R}^{M\times d}$ as defined in

Lemma 2, then an optimal solution to the problem

$$\min_{\mathbf{Z}} \ \frac{1}{2}\left\|\mathbf{\Phi}_0\mathbf{Z}^\top - \mathbf{\Phi}(\mathbf{A}^\top \odot \mathbf{Z}^\top)\right\|_F^2 + \frac{\alpha}{2}\|\mathbf{Z}-\mathbf{X}\|_F^2,$$

is given by

$$\mathrm{vec}(\mathbf{Z}^\top) := \alpha[(\mathbf{B}-\mathbf{C})^\top(\mathbf{B}-\mathbf{C})+\alpha\mathbf{I}_{MT}]^{-1}\cdot\mathrm{vec}(\mathbf{X}^\top),$$

where $\mathbf{B}=(\mathbf{I}_M\otimes\mathbf{\Phi}_0)$ and $\mathbf{C}=(\mathbf{I}_M\otimes\mathbf{\Phi})[(\mathbf{I}_M\odot\mathbf{A}^\top)\otimes\mathbf{I}_T]$. $\otimes$ denotes the Kronecker product.

Proof: In this case, we can use vectorization:

$$\mathrm{vec}(\mathbf{\Phi}_0\mathbf{Z}^\top) = (\mathbf{I}_M\otimes\mathbf{\Phi}_0)\cdot\mathrm{vec}(\mathbf{Z}^\top),$$
$$\mathrm{vec}(\mathbf{\Phi}(\mathbf{A}^\top\odot\mathbf{Z}^\top)) = (\mathbf{I}_M\otimes\mathbf{\Phi})\cdot\mathrm{vec}(\mathbf{A}^\top\odot\mathbf{Z}^\top)$$
$$= (\mathbf{I}_M\otimes\mathbf{\Phi})[(\mathbf{I}_M\odot\mathbf{A}^\top)\otimes\mathbf{I}_T]\cdot\mathrm{vec}(\mathbf{Z}^\top),$$

where $\mathrm{vec}(\cdot)$ denotes the vectorization operator for any given matrix. Denote by f the objective of problem (26):

$$f = \frac{1}{2}\|(\mathbf{B}-\mathbf{C})\cdot\mathrm{vec}(\mathbf{Z}^\top)\|_2^2 + \frac{\alpha}{2}\|\mathrm{vec}(\mathbf{Z}^\top)-\mathrm{vec}(\mathbf{X}^\top)\|_2^2.$$

By letting

$$\frac{df}{d\,\mathrm{vec}(\mathbf{Z}^\top)} = (\mathbf{B}-\mathbf{C})^\top(\mathbf{B}-\mathbf{C})\,\mathrm{vec}(\mathbf{Z}^\top)$$
$$+\alpha[\mathrm{vec}(\mathbf{Z}^\top)-\mathrm{vec}(\mathbf{X}^\top)]=\mathbf{0},$$

we have

$$\mathrm{vec}(\mathbf{Z}^\top) = \alpha[(\mathbf{B}-\mathbf{C})^\top(\mathbf{B}-\mathbf{C})+\alpha\mathbf{I}_{MT}]^{-1}\cdot\mathrm{vec}(\mathbf{X}^\top).$$

$\square$

REFERENCES

[1] X. Chen and L. Sun, "Bayesian temporal factorization for multidimensional time series prediction," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Mar. 17, 2021, doi: 10.1109/TPAMI.2021.3066551.

[2] X. Chen, Y. Chen, N. Saunier, and L. Sun, "Scalable low-rank tensor learning for spatiotemporal traffic data imputation," *Transp. Res. C, Emerg. Technol.*, vol. 129, Aug. 2021, Art. no. 103226.

[3] M. T. Bahadori, Q. R. Yu, and Y. Liu, "Fast multivariate spatio-temporal analysis via low rank tensor learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3491–3499.

[4] X. Chen, Z. He, and L. Sun, "A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation," *Transp. Res. C, Emerg. Technol.*, vol. 98, pp. 73–84, Jan. 2019.

[5] L. Li, J. McCann, N. S. Pollard, and C. Faloutsos, "DynaMMo: Mining and summarization of coevolving sequences with missing values," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2009, pp. 507–516.

[6] X. Chen, J. Yang, and L. Sun, "A nonconvex low-rank tensor completion model for spatiotemporal traffic data imputation," *Transp. Res. C, Emerg. Technol.*, vol. 117, Aug. 2020, Art. no. 102673.

[7] A. B. Said and A. Erradi, "Spatiotemporal tensor completion for improved urban traffic imputation," *IEEE Trans. Intell. Transp. Syst.*, early access, Mar. 11, 2021, doi: 10.1109/TITS.2021.3062999.

[8] L. Xiong, X. Chen, T.-K. Huang, J. Schneider, and J. G. Carbonell, "Temporal collaborative filtering with Bayesian probabilistic tensor factorization," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2010, pp. 211–222.

[9] N. Rao, H.-F. Yu, P. Ravikumar, and I. S. Dhillon, "Collaborative filtering with graph information: Consistency and scalable methods," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 2107–2115.

[10] T. Yokota, Q. Zhao, and A. Cichocki, "Smooth PARAFAC decomposition for tensor completion," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5423–5436, Oct. 2016.

[11] H.-F. Yu, N. Rao, and I. S. Dhillon, "Temporal regularized matrix factorization for high-dimensional time series prediction," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 847–855.

[12] R. Sen, H.-F. Yu, and I. S. Dhillon, "Think globally, act locally: A deep neural network approach to high-dimensional time series forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 4838–4847.

[13] L. Li, X. Su, Y. Zhang, Y. Lin, and Z. Li, "Trend modeling for traffic time series analysis: An integrated study," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3430–3439, Dec. 2015.

[14] H. Tan, Y. Wu, B. Shen, P. J. Jin, and B. Ran, "Short-term traffic prediction based on dynamic tensor completion," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 8, pp. 2123–2133, Aug. 2016.

[15] Z. Li, N. D. Sergin, H. Yan, C. Zhang, and F. Tsung, "Tensor completion for weakly-dependent data on graph for metro passenger flow prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 4, pp. 4804–4810.

[16] J. L. Schafer, *Analysis of Incomplete Multivariate Data*. Boca Raton, FL, USA: CRC Press, Aug. 1997.

[17] J. Chen and J. Shao, "Nearest neighbor imputation for survey data," *J. Off. Statist.*, vol. 16, no. 2, pp. 113–132, 2000.

[18] S. Tak, S. Woo, and H. Yeo, "Data-driven imputation method for traffic data in sectional units of road links," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 6, pp. 1762–1771, Jun. 2016.

[19] F. Rodrigues, K. Henrickson, and F. C. Pereira, "Multi-output Gaussian processes for crowdsourced traffic data imputation," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 594–603, Feb. 2019.

[20] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[21] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.

[22] G. Lai, W.-C. Chang, Y. Yang, and H. Liu, "Modeling long- and short-term temporal patterns with deep neural networks," in *Proc. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jun. 2018, pp. 95–104.

[23] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2117–2130, Sep. 2013.

[24] Y. Zhang and Z. Lu, "Penalty decomposition methods for rank minimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 46–54.

[25] K. Chen, H. Dong, and K.-S. Chan, "Reduced rank regression via adaptive nuclear norm penalization," *Biometrika*, vol. 100, no. 4, pp. 901–920, 2013.

[26] C. Lu, C. Zhu, C. Xu, S. Yan, and Z. Lin, "Generalized singular value thresholding," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 1805–1811.

[27] J. Gillard and K. Usevich, "Structured low-rank matrix completion for forecasting in time series analysis," *Int. J. Forecasting*, vol. 34, no. 4, pp. 582–597, Oct. 2018.

[28] X. Chen and L. Sun, "Low-rank autoregressive tensor completion for multivariate time series forecasting," 2020, *arXiv:2006.10436*. [Online]. Available: http://arxiv.org/abs/2006.10436
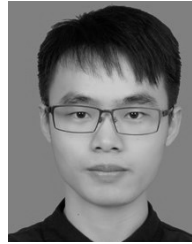
Xinyu Chen received the B.S. degree in traffic engineering from Guangzhou University, Guangzhou, China, in 2016, and the M.S. degree in transportation information engineering and control from Sun Yat-Sen University, Guangzhou, in 2019. He is currently pursuing the Ph.D. degree with the Civil, Geological and Mining Engineering Department, Polytechnique Montreal, Montreal, QC, Canada. His current research centers on machine learning, spatiotemporal data modeling, and intelligent transportation systems.
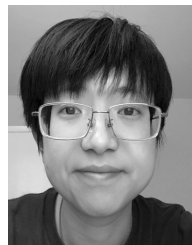
Mengying Lei received the B.S. degree in automation from Huazhong Agricultural University in 2016 and the M.S. degree from the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China, in 2019. She is currently pursuing the Ph.D. degree with the Department of Civil Engineering, McGill University, Montreal, QC, Canada. Her research currently focuses on spatiotemporal data modeling and intelligent transportation systems.

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

Lijun Sun (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 University of Singapore in 2015. He is currently an Assistant Professor with the Department of Civil Engineering, McGill University, Montreal, QC, Canada. His research centers on intelligent transportation systems, machine learning, spatiotemporal modeling, travel behavior, and agent-based simulation.