

Low-Rank Autoregressive Tucker Decomposition for Traffic Data Imputation

1st Jiaxin Lu

Shenzhen Key Laboratory of Safety and Security for Next Generation of Industrial Internet
Department of Statistics and Data Science
Southern University of Science and Technology
Shenzhen, China
12332881@mail.sustech.edu.cn

2nd Wenwu Gong

Department of Statistics and Data Science
Southern University of Science and Technology
Shenzhen, China
12031299@mail.sustech.edu.cn

3rd Lili Yang

Department of Statistics and Data Science
Southern University of Science and Technology
Shenzhen, China
yangll@sustech.edu.cn

Abstract—Real traffic data is often missing due to diverse interference. However, uncompleted inputs will weaken the abilities of intelligent transportation systems. Therefore, it is of great interest that suitable imputation methods be designed. This paper proposes a low-rank autoregressive Tucker decomposition (LATD) method by exploring the spatiotemporal correlations embedded in high-dimensional traffic data. The low-rank factor matrices and core tensor introduced by the Tucker decomposition allow us to better characterize the long-term trends of the traffic data. We incorporate an autoregressive model to extract the short-term patterns involved. Besides implementing differences between neighboring elements to promote smoothness, this regularization is also well interpretable for characterizing the spatiotemporal correlations. To solve the LATD model, we design a proximal alternating linear minimization algorithm to update each variable iteratively. Numerical experiments on two real traffic datasets indicate that our proposed model outperforms other imputation methods in achieving higher accuracy.

Index Terms—Traffic data imputation, low rankness, spatiotemporal correlations, Tucker Decomposition, autoregressive model.

I. INTRODUCTION

With significant advancements in sensor technology, data measurement instruments offer better spatial coverage and time precision. For instance, traffic data collected across the city network records daily vehicle flow and speeds on different roads. Those extensive traffic datasets offer new possibilities for intelligent transportation systems, enabling more reliable applications such as route planning and commute time management. However, real-world traffic data often suffer from missing problems due to conversion errors and environmental obstructions [1]. Therefore, imputing the incomplete traffic data is essential for improving data quality.

Traffic data from nearby sensors often display similar regularity due to their proximity, with recurring short-term patterns over adjacent time points [2]. Understanding these complex relationships across spatial and temporal dimensions is crucial [3]. Recently, attention has been paid to the tensor-

based imputation methods, particularly due to the periodicity and seasonality reflected in traffic data. Low rankness is regarded as a key assumption for capturing those long-term trends. Meanwhile, traffic data imputation focuses on exploring spatiotemporal correlations. Strategies like incorporating total/quadratic variance [4], [5] and graph regularization [6] as ‘smooth’ priors have been proposed. However, those conventional ‘smooth’ priors are difficult to explore this potential connection. Fortunately, the autoregressive model has excellent interpretability in characterizing the dependence between the time series of the road segments. Besides, the Tucker decomposition effectively incorporates spatiotemporal prior terms while extracting low rankness from the different modes of the tensor. Therefore, the low-rank Tucker model serves as an efficient tool to make a good link between long-term trends and short-term patterns within traffic data [7], [8].

This paper proposes a novel Low-Rank Autoregressive Tucker Decomposition (LATD) model for imputing traffic data. The key contributions are outlined as follows:

- We utilize the Tucker model’s factor matrix and core tensor to represent long-term trends in the traffic data, then combine the autoregressive model to emphasize short-term patterns. The imputation capability is enhanced by describing the spatiotemporal correlations.
- We propose a proximal alternating linear minimization algorithm to solve the nonconvex-nonsmooth optimization problem and demonstrate the convergence in numerical.
- Extensive experiments are conducted on real-world traffic datasets, which show that LATD achieves superior accuracy compared to several tensor-based imputation models.

II. RELATED WORK

Researchers have proposed tensor imputation methods leveraging low rankness to analyze high-dimensional structures and impute missing traffic data. Those tensor-based methods efficiently identify global traffic situations to recover missing

entries from observed data [9]. Xie et al. [10] introduced a sequential tensor imputation algorithm utilizing a third-order traffic tensor. Ran et al. [11] employed a fourth-way traffic flow tensor to capture the overall conditions, resulting in enhanced performance. Similarly, Wang et al. [12] used the Hankel operator to capture traffic information in a data-driven way. However, the performance of low-rank tensor imputation techniques needs to improve under high-level missing rates, as they overlook cross-dimensional effects in traffic data [13].

Considering the spatiotemporal correlations of traffic data, previous studies have shown that combining low-rank assumption with ‘smooth’ priors outperforms other imputation methods [14]. They can be categorized into two types.

On the one hand, introducing an auxiliary matrix to depict spatiotemporal correlations in a low-rank tensor decomposition (LRTD) model has gained widespread adoption. Such as the factor graph embedding [8], Toeplitz matrix [15], and Laplacian kernel [16]. Compared with other LRTD methods, the Tucker model offers a unique advantage in capturing the spatiotemporal correlations within a subspace [17].

On the other hand, the difference operator of the unfolding matrix is used to encode spatiotemporal correlations. Pan et al. [18] presented a unified low-rank and sparse enhanced Tucker decomposition model. The inherent similarity of the traffic data is characterized by merging the structured matrix acting between the tensor modes. Chen et al. [19] developed a scalable tensor learning model. It uses the unitary transform matrix to preserve the daily dependency features between the traffic data. In another work, Chen et al. [20] used truncated nuclear norm minimization to ensure the global consistency of the traffic data. With the help of the autoregressive model, the temporal variation term was designed to explore the similarity between contiguous elements.

It is worth noting that these methods only stay on the surface by leveraging the spatiotemporal correlations merely by designing a regularization term. They fail to thoroughly explain why such models are constructed in this manner without integrating the special properties of traffic data.

III. PRELIMINARIES

We present the symbols in Table I and briefly review some basic concepts. Moreover, we will analyze some of the unique properties of the traffic data and the specific process of establishing the LATD model.

A. Notations

We can represent the Tucker decomposition $\mathcal{X} = \mathcal{G} \times_1 \mathbf{U}_1 \cdots \times_N \mathbf{U}_N = \mathcal{G} \times_{n=1}^N \mathbf{U}_n$ by $\mathbf{X}_{(n)} = \mathbf{U}_n \mathbf{G}_{(n)} \mathbf{V}_n^T$, $\mathbf{V}_n = (\mathbf{U}_N \otimes \cdots \otimes \mathbf{U}_{n+1} \otimes \mathbf{U}_{n-1} \otimes \cdots \otimes \mathbf{U}_1)$. It is not difficult to verify that $\text{vec}(\mathcal{X}) = (\mathbf{U}_N \otimes \cdots \otimes \mathbf{U}_n \otimes \cdots \otimes \mathbf{U}_1) \text{vec}(\mathcal{G}) = \otimes_{n=1}^N \mathbf{U}_n \text{vec}(\mathcal{G})$.

For a given tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and observed index set Ω , we define $\mathcal{P}_\Omega(\mathcal{X})$ as a projector that keeps the given nonzero values and maps other values to zero,

$$\mathcal{P}_\Omega(\mathcal{X}) := \begin{cases} x_{i_1, i_2, \dots, i_n}, & \text{if } (i_1, i_2, \dots, i_n) \in \Omega \\ 0, & \text{otherwise.} \end{cases}$$

TABLE I
NOTATIONS

Symbol	Description
$\mathbf{x}, \mathbf{X}, \mathcal{X}$	A vector, matrix, and tensor, respectively.
$\mathbf{X}_{(n)}$	Mode- n unfolding of tensor \mathcal{X} .
$\mathbf{x}_{m, [t+1:]}$	A sub-vector containing the last $N - t$ entries of \mathbf{x}_m .
$x_{m, n}$	(m, n) th entry in \mathbf{X} .
$\Omega, \bar{\Omega}$	Observed index set and its complement.
\times_n	Mode- n product.
\otimes	Kronecker product.
$\ \cdot\ _F$	Frobenius norm.
$\ \cdot\ _1$	ℓ_1 norm.
$\ \cdot\ _2$	ℓ_2 norm.
$\ \cdot\ _*$	Nuclear norm.
$\langle \cdot, \cdot \rangle$	Inner product.
$\mathcal{S}_\eta(\mathbf{x})$	Shrinkage operator with η in component-wise.
$\mathcal{D}_\eta(\mathbf{X})$	Operator yielding SVD shrinkage with η .

For convenience, we introduce the tensorization operator $\mathcal{Q}(\cdot)$ to implement the folding operation for matrices. Specifically, a third-order tensor can be generated by $\mathcal{X} = \mathcal{Q}(\mathbf{Y}) \in \mathbb{R}^{M \times I \times J}$. Conversely, the resulting tensor can be transformed into the original matrix by $\mathbf{Y} = \mathcal{Q}^{-1}(\mathcal{X}) \in \mathbb{R}^{M \times (IJ)}$, where $\mathcal{Q}^{-1}(\cdot)$ denotes the inverse operator.

B. Model Rationality

Traffic data can be represented in the form of “sensor \times time of day \times day”. Alongside global low rankness, piece-wise ‘smoothing’ often exist across different dimensions. This is attributed to dependencies between neighboring data, prompting the introduction of spatiotemporal prior. By performing difference operations on the unfolding matrix, we unveil potential interactions between the modes of the original tensor, thereby enriching information for traffic data imputation.

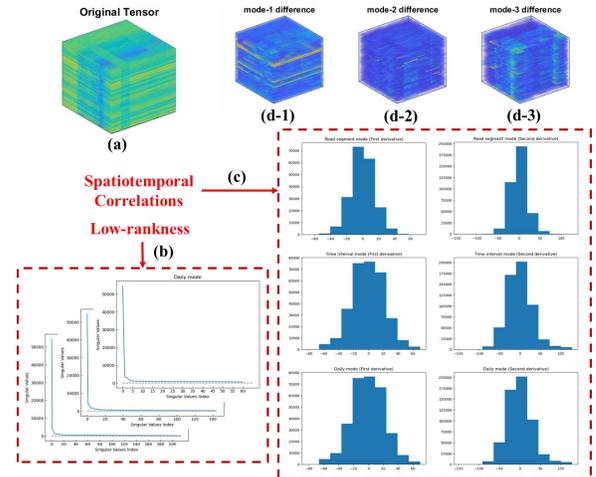


Fig. 1. Statistical analysis of the traffic dataset.

When analyzing traffic data, it’s essential to consider its long-term trends and short-term patterns. Fig. 1-(b) indicates a sharp drop of singular values at the beginning, signifying highly concentrated data energy. In Fig. 1-(c), elements are

distributed around zero, suggesting minimal variances between neighboring data points. Fig. 1-(a) reveals that the entries are intricately mixed, with subtle discrepancies due to the interference of surrounding elements. However, after the first modal difference operation, the boundaries among each data cluster become clearer. In Fig. 1-(d-1), distinct vehicle speed ranges carried on each road segment are easily discernible. Conversely, applying the second and third modal difference operations to the original data fails to achieve a similar processing effect. As shown in Fig. 1-(d-2) and Fig. 1-(d-3), there is no apparent hierarchy of the data recorded by sensors in every scene. This implies that the potential spatiotemporal features that traffic dataset possesses are not currently refined, so some spatiotemporal priors need to be considered.

Based on our analysis, we use Tucker decomposition to capture the low-rank prior and emphasize spatiotemporal correlations through an autoregressive model. This construction improves the model performance by describing the dependency between dimensions. In the following sections, we will detail the design intent of the LATD model.

C. Low Rankness based on Tucker Decomposition

The Tucker decomposition is a powerful tool for revealing low-rank priors in multidimensional traffic data. It efficiently captures global low rankness while preserving the tensor structure without needing a predefined rank. We utilize the nuclear norm of the factor matrices and the ℓ_1 norm of the core tensor for relaxation.

$$\begin{aligned} \min_{\mathcal{G}, \{\mathbf{U}_n\}} (1 - \alpha) \prod_{n=1}^N \|\mathbf{U}_n\|_* + \alpha \|\mathcal{G}\|_1 \\ \text{s.t.}, \mathcal{X} = \mathcal{G} \times_{n=1}^N \mathbf{U}_n, 0 < \alpha < 1. \end{aligned}$$

Specifically, when tensor \mathcal{X} exhibits low-rank, its core tensor \mathcal{G} is expected to be sparse or $\{\mathbf{U}_n\}$ forms a collection of low-rank matrices. Since it is more difficult to solve the product function term that represents the block size of the core tensor, we use the adaptive weighted factor matrix nuclear norm summation term instead.

D. Spatiotemporal Correlations

‘Smoothness’ in the tensor structure is reflected through its unfolding matrix along the specific modes. The autoregressive model, constructed using prior traffic information, maintains spatiotemporal correlations effectively. In this paper, we depict the spatiotemporal dependence of the traffic dataset by computing the difference between neighboring elements and taking the ℓ_2 norm to preserve it. Given a set of time lags $\mathcal{H} = \{h_1, \dots, h_d\}$, the spatiotemporal regularization based on the autoregressive model can be expressed as

$$\|\mathbf{Z}\|_{\mathbf{A}, \mathcal{H}} = \sum_{m,t} (z_{m,t} - \sum_i a_{m,i} z_{m,t-h_i})^2, \mathbf{Z} = \mathcal{Q}^{-1}(\mathcal{X}).$$

In practice, the tensor is unfolded to separate temporal and spatial information. This guarantees that each column of the resulting matrix represents a complete time series recorded for a specific road segment. The corresponding spatiotemporal

relevance of the traffic dataset is then measured through the suitable autoregressive coefficient matrix. It is worth noting that \mathbf{A} is also a variable that needs to be updated.

IV. METHODOLOGY

A. Proposed Model

The proposed LATD model is defined as

$$\begin{aligned} \min_{\mathcal{G}, \{\mathbf{U}_n\}, \mathcal{X}, \mathbf{Z}, \mathbf{A}} (1 - \alpha) \sum_{n=1}^N \omega_n \|\mathbf{U}_n\|_* + \alpha \|\mathcal{G}\|_1 + \frac{\gamma}{2} \|\mathbf{Z}\|_{\mathbf{A}, \mathcal{H}} \\ \text{s.t.}, \mathcal{X} = \mathcal{G} \times_{n=1}^N \mathbf{U}_n, \mathcal{X} = \mathcal{Q}(\mathbf{Z}), \\ \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{Y}) \end{aligned} \quad (1)$$

where $0 < \alpha < 1$, $\omega_n = \prod_{i=1, i \neq n}^3 \frac{1}{R_i}$, $R_i = \sum \sigma(\mathbf{U}_i)$, and \mathcal{Y} is the partially observed tensor. The parameter γ controls the trade-off between the low-rank term and spatiotemporal regularization.

B. Proposed Algorithm

To solve the above problem, we first penalize the two constraints in (1), where β and ρ are the weight parameters of the introduced penalty.

$$\begin{aligned} \min_{\mathcal{G}, \{\mathbf{U}_n\}, \mathcal{X}, \mathbf{Z}, \mathbf{A}} (1 - \alpha) \sum_{n=1}^N \omega_n \|\mathbf{U}_n\|_* + \alpha \|\mathcal{G}\|_1 + \frac{\gamma}{2} \|\mathbf{Z}\|_{\mathbf{A}, \mathcal{H}} \\ + \frac{\beta}{2} \|\mathcal{X} - \mathcal{G} \times_{n=1}^N \mathbf{U}_n\|_F^2 + \frac{\rho}{2} \|\mathcal{X} - \mathcal{Q}(\mathbf{Z})\|_F^2 \\ \text{s.t.}, \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{Y}) \end{aligned} \quad (2)$$

Since (2) is a nonconvex-nonsmooth problem, we use the proximal alternating linear minimization method. Specifically, its solution process will be decomposed into the following subproblems.

Optimization of \mathcal{G} . We use the proximal gradient method to update the core tensor \mathcal{G} based on a vectorized form:

$$\begin{aligned} \hat{\mathcal{G}} &= \operatorname{argmin}_{\mathcal{G}} \alpha \|\mathcal{G}\|_1 + \frac{\beta}{2} \|\operatorname{vec}(\mathcal{X}) - (\otimes_{n=1}^N \mathbf{U}_n) \operatorname{vec}(\mathcal{G})\|_F^2 \\ &\approx \operatorname{argmin}_{\mathcal{G}} \alpha \|\mathcal{G}\|_1 + \langle \mathcal{G} - \tilde{\mathcal{G}}, \nabla_{\mathcal{G}} f(\tilde{\mathcal{G}}) \rangle + \frac{L_{\mathcal{G}}}{2} \|\mathcal{G} - \tilde{\mathcal{G}}\|_F^2 \\ &= \mathcal{S}_{\frac{\alpha}{L_{\mathcal{G}}}} \left(\tilde{\mathcal{G}} - \frac{1}{L_{\mathcal{G}}} \nabla_{\mathcal{G}} f(\tilde{\mathcal{G}}) \right). \end{aligned} \quad (3)$$

The corresponding gradient and Lipschitz constant are

$$\begin{aligned} \nabla_{\mathcal{G}} f(\mathcal{G}) &= \beta (\mathcal{G} \times_{n=1}^N \mathbf{U}_n^T \mathbf{U}_n - \mathcal{X} \times_{n=1}^N \mathbf{U}_n^T), \\ L_{\mathcal{G}} &= \beta \prod_{n=1}^N \|\mathbf{U}_n^T \mathbf{U}_n\|_2. \end{aligned}$$

$\tilde{\mathcal{G}}$ is updated by

$$\tilde{\mathcal{G}}^k = \mathcal{G}^k + \eta_k (\mathcal{G}^k - \mathcal{G}^{k-1}), \text{ for } k \geq 1$$

with the updated step size η_k

$$\eta_k = \frac{t^{k-1} - 1}{t^k}, t^k = \frac{1 + \sqrt{4(t^{k-1})^2 + 1}}{2}, \text{ for } k \geq 1, t^0 = 1.$$

Optimization of \mathbf{U}_n , $n = 1, \dots, N$. We perform mode- n unfolding of the given objective tensor to get the hidden factor matrix \mathbf{U}_n subproblem:

$$\begin{aligned} \hat{\mathbf{U}}_n &= \underset{\mathbf{U}_n}{\operatorname{argmin}} (1 - \alpha)\omega_n \|\mathbf{U}_n\|_* + \frac{\beta}{2} \|\mathbf{X}_{(n)} - \mathbf{U}_n \mathbf{G}_{(n)} \mathbf{V}_n^T\|_F^2 \\ &\approx \underset{\mathbf{U}_n}{\operatorname{argmin}} (1 - \alpha)\omega_n \|\mathbf{U}_n\|_* + \langle \mathbf{U}_n - \tilde{\mathbf{U}}_n, \nabla_{\mathbf{U}_n} f(\tilde{\mathbf{U}}_n) \rangle \\ &\quad + \frac{L_{\mathbf{U}_n}}{2} \|\mathbf{U}_n - \tilde{\mathbf{U}}_n\|_F^2 \\ &= \mathcal{D}_{\frac{(1-\alpha)\omega_n}{L_{\mathbf{U}_n}}} \left(\tilde{\mathbf{U}}_n - \frac{1}{L_{\mathbf{U}_n}} \nabla_{\mathbf{U}_n} f(\tilde{\mathbf{U}}_n) \right), \end{aligned} \quad (4)$$

where

$$\mathbf{V}_n = (\mathbf{U}_N \otimes \dots \otimes \mathbf{U}_{n+1} \otimes \mathbf{U}_{n-1} \otimes \dots \otimes \mathbf{U}_1).$$

The corresponding gradient and Lipschitz constant are

$$\nabla_{\mathbf{U}_n} f(\mathbf{U}_n) = \beta \left(\mathbf{U}_n \mathbf{G}_{(n)} \mathbf{V}_n^T \mathbf{V}_n \mathbf{G}_{(n)}^T - \mathbf{X}_{(n)} \mathbf{V}_n \mathbf{G}_{(n)}^T \right),$$

$$L_{\mathbf{U}_n} = \beta \left\| \mathbf{G}_{(n)} \mathbf{V}_n^T \mathbf{V}_n \mathbf{G}_{(n)}^T \right\|_2.$$

$\tilde{\mathbf{U}}_n$ is updated by

$$\tilde{\mathbf{U}}_n^k = \mathbf{U}_n^k + \eta_k (\mathbf{U}_n^k - \mathbf{U}_n^{k-1}), \text{ for } k \geq 1.$$

Optimization of \mathbf{Z} . We can rewrite the original problem with respect to \mathbf{Z} as (5):

$$\begin{aligned} \hat{\mathbf{Z}} &= \underset{\mathbf{Z}}{\operatorname{argmin}} \frac{\gamma}{2} \|\mathbf{Z}\|_{\mathbf{A}, \mathcal{H}} + \frac{\rho}{2} \|\mathbf{Z} - \mathcal{Q}^{-1}(\mathcal{X})\|_2^2 \\ &= \underset{\mathbf{Z}}{\operatorname{argmin}} \sum_m \left[\frac{\gamma}{2} \|\Psi_0 \mathbf{z}_m - \sum_i a_{m,i} \Psi_i \mathbf{z}_m\|_2^2 \right. \\ &\quad \left. + \frac{\rho}{2} \|\mathbf{z}_m - \mathcal{Q}_m^{-1}(\mathcal{X})\|_2^2 \right]. \end{aligned} \quad (5)$$

The closed-form solution is given by

$$\hat{\mathbf{z}}_m = \frac{\rho}{\gamma} \left(\mathbf{B}_m^T \mathbf{B}_m + \frac{\rho}{\gamma} \mathbf{I} \right)^{-1} \cdot \mathcal{Q}_m^{-1}(\mathcal{X}), \quad m = 1, \dots, M, \quad (6)$$

where

$$\mathbf{B}_m = \Psi_0 - \sum_i a_{m,i} \Psi_i, \quad m = 1, \dots, M,$$

and

$$\begin{aligned} \Psi_0 &= [\mathbf{0}_{(T-h_d) \times h_d} \quad \mathbf{I}_{T-h_d}], \\ \Psi_i &= [\mathbf{0}_{(T-h_d) \times (h_d-h_i)} \quad \mathbf{I}_{T-h_d} \quad \mathbf{0}_{(T-h_d) \times h_i}], \quad i = 1, 2, \dots, d, \\ \mathbf{I} &= [\Psi_1 \quad \Psi_2 \quad \dots \quad \Psi_d] \end{aligned}$$

are matrices defined based on time lag set $\mathcal{H} = \{h_1, \dots, h_d\}$.

Optimization of \mathcal{X} . The optimal solution \mathcal{X} follows

$$\begin{aligned} \hat{\mathcal{X}}_{\Omega} &= \underset{\mathcal{X}}{\operatorname{argmin}} \frac{\beta}{2} \|\mathcal{X} - \mathcal{G} \times_{n=1}^3 \mathbf{U}_n\|_F^2 + \frac{\rho}{2} \|\mathcal{X} - \mathcal{Q}(\mathbf{Z})\|_F^2 \\ &= \left[\frac{1}{\beta + \rho} (\beta \mathcal{G} \times_{n=1}^3 \mathbf{U}_n + \rho \mathcal{Q}(\mathbf{Z})) \right]_{\Omega}, \quad \hat{\mathcal{X}}_{\Omega} = \mathcal{Y}_{\Omega}. \end{aligned} \quad (7)$$

Optimization of \mathbf{A} . we solve the following (8):

$$\begin{aligned} \hat{\mathbf{A}} &= \underset{\mathbf{A}}{\operatorname{argmin}} \sum_{m,t} (z_{m,t} - \sum_i a_{m,i} z_{m,t-h_i})^2 \\ &= \underset{\mathbf{A}}{\operatorname{argmin}} \sum_m \|z_{m,[h_d+1:]} - \mathbf{V}_m \mathbf{a}_m\|_2^2, \end{aligned} \quad (8)$$

where

$$\mathbf{V}_m = (\mathbf{v}_{h_d+1}, \dots, \mathbf{v}_T)^T, \quad m = 1, \dots, M,$$

$$\mathbf{v}_t = (z_{m,t-h_1}, \dots, z_{m,t-h_d})^T, \quad t = h_d + 1, \dots, T.$$

It has a closed-form solution, which is given by

$$\hat{\mathbf{a}}_m = \mathbf{V}_m^{\dagger} z_{m,[h_d+1:]}, \quad m = 1, \dots, M, \quad (9)$$

where \dagger denotes the Moore-Penrose pseudo-inverse.

The proposed algorithm for LATD can be summarized in Algorithm 1. If the convergence condition is satisfied, the algorithm returns $\hat{\mathcal{X}}$ as the final result.

Algorithm 1: PALM-based LATD

Input : Missing tensor \mathcal{Y} , observed index set Ω .
Output: Recovered tensor $\hat{\mathcal{X}}$.
1 Initialize: $\mathcal{G}^0, \{\mathbf{U}_n^0\}$ ($1 \leq n \leq N$), $0 < \alpha < 1$,
 $\beta = \rho = 1, \gamma = \min\{\beta, \rho\}, s = 1$;
2 repeat
3 for $k = 0$ **to** $K - 1$ **do**
4 Update $\mathcal{G}^{s+1,k+1}$ **using** (3).
5 for $n = 1$ **to** N **do**
6 Update $\mathbf{U}_n^{s+1,k+1}$ **using** (4).
7 end
8 for $m = 1$ **to** M **do**
9 Update $\mathbf{z}_m^{s+1,k+1}$ **using** (6).
10 end
11 Update $\mathcal{X}^{s+1,k+1}$ **using** (7).
12 end
13 for $m = 1$ **to** M **do**
14 Update \mathbf{a}_m^{s+1} **using** (9).
15 end
16 s=s+1;
17 until $\|\mathcal{X}^{s+1,K} - \mathcal{X}^{s,K}\|_F \|\mathcal{X}^{s,K}\|_F^{-1} < \epsilon$;

V. EXPERIMENTS

This section presents the numerical experiments conducted on two real traffic datasets. The results demonstrate that the LATD model outperforms several state-of-the-art methods.

A. Traffic Datasets

- **Guangzhou urban traffic speed dataset.**¹ This dataset contains the 10-minute resolution of traffic speed collected from 214 road segments in Guangzhou, China, over one week, and is $214 \times 144 \times 7$ in size.

¹<https://doi.org/10.5281/zenodo.1205229>

- **Hangzhou metro passenger flow dataset.**² This dataset provides a 10-minute resolution of inbound passenger flow at 80 subway stations in Hangzhou, China, over one week. We exclude the time interval (6 hours) in which there is no service at a fixed time of the day, and the retained portion is $80 \times 108 \times 7$ in size.

B. Experimental Settings

Missing scenario. For verifying the effectiveness of the LATD model, we primarily focus on the random missing scenario where the missing data are uniformly distributed. Following this mechanism, we mask the index set Ω and use the remaining observations as model input.

Evaluation indicators. To ensure a standardized criterion for comparing imputation performance, we utilize mean absolute percentage error (MAPE) and normalized mean absolute error (NMAE) as evaluation indicators:

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

$$\text{NMAE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{\sum_{i=1}^n |y_i|},$$

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

Baseline models. For comparison, some state-of-the-art imputation methods for traffic data have been selected: LATC, LSTC, STRTD, LRSETD, stTT, and STH-LRTC, to show the superiority of our model. All the contrast methods integrate low-rank term and spatiotemporal regularization, consistent with the basic structure of the LATD model. Their main properties are listed in Table II.

TABLE II
SOME EXISTING TRAFFIC DATA IMPUTATION METHODS

Methods	Types of the prior			Structures
	Low rankness	Spatial	Temporal	
LATD	✓		✓	3rd tensor
LATC [20]	✓		✓	3rd tensor
LSTC [19]	✓		✓	3rd tensor
STRTD [8]	✓	✓	✓	3rd tensor
LRSETD [18]	✓	✓	✓	3rd tensor
stTT [15]	✓	✓	✓	3rd tensor
STH-LRTC [12]	✓	✓	✓	4th tensor

Parameters selection. In LATD model, five parameters need to be tuned. They are α , β related to the low-rank term, γ , ρ controlling the spatiotemporal regularization, and the set of time lag \mathcal{H} . From practice, the ratio of the two pairs of parameters α to β and ρ to γ plays a key role in the model. We conducted preliminary tests on the selection of the above parameters and finally decided on the following choices: $\{\alpha = 0.8, \beta = 1.6, \rho = 20^{-15}, \gamma = 2\}$ for the first dataset and $\{\alpha = 0.1, \beta = 0.001, \rho = 10^{-16}, \gamma = 10^{-5}\}$ for the second. In both experiments \mathcal{H} was set to $\{1, 2, \dots, 6\}$. For better model comparison, the termination condition of all

²<https://tianchi.aliyun.com/competition/entrance/231708/information>
We only select data from the first week of the original dataset.

experiments was set to $\text{tol} = 10^{-5}$, and the maximum number of iterations was 300. The parameters of the baseline were chosen by optimal assignment.

C. Results

Table III shows the overall imputation performance of LATD and baseline models on two specific traffic datasets, where the best results are bolded. For the G and H datasets, we plot the MAPE values and NMAE values for different missing rates in Fig. 2. It should be noted that we have omitted some points with unusual values from the figure.

The results show that LATD outperforms other models except for the case of extremely high deficiency in the dataset G. Even when the missing rate is 97%, LATD exhibits suboptimal imputation performance. This suggests combining short-term and long-term trends would benefit traffic data imputation.

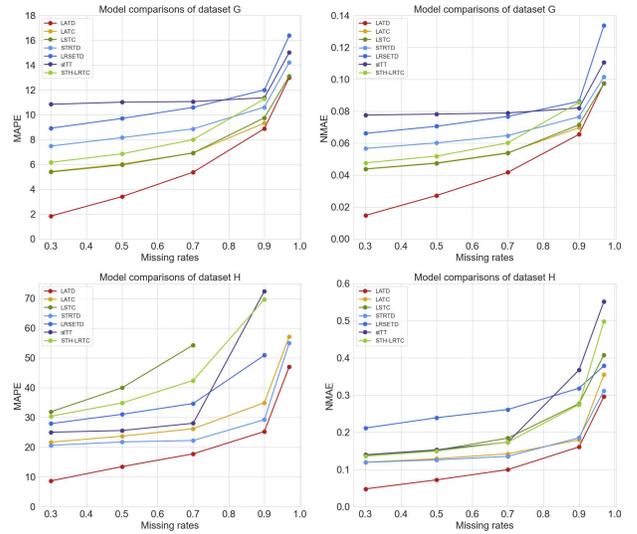


Fig. 2. Performance comparison in terms of MAPE values (left) and NMAE values (right) with different missing rates for G (upper) and H (lower) datasets.

VI. CONCLUSION

The proposed LATD model leverages the long-term trends and short-term patterns in traffic data to impute missing entries. By the Tucker decomposition, LATD captures the inherent low rankness utilizing weighted factor matrix nuclear norm and core tensor sparsity. LATD integrates an autoregressive model to explore spatiotemporal correlations by differencing neighboring elements. Experimental results on real traffic datasets validate the superiority of LATD, showing higher accuracy compared to other state-of-the-art methods.

There are still some limitations to our model. For simplicity, it is assumed that each component obeys an independent autoregressive model and ignores the mutual influences that may be implied between them. In the future, we can continue to construct new models based on tensor representations by introducing time variation into the classical idea of spatial imputation method.

TABLE III
PERFORMANCE COMPARISON OF LATD AND BASELINE MODELS

Dataset	Missing Rate	LATD	LATC	LSTC	STRTD	LRSETD	stTT	STH-LRTC
(G)-MAPE	30%	1.8251	5.4187	5.3787	7.4668	8.9087	10.8340	6.1564
	50%	3.3913	6.0129	5.9502	8.1463	9.6923	11.0070	6.8428
	70%	5.3624	6.9227	6.9144	8.8558	10.5840	11.0540	7.9906
	90%	8.8714	9.3191	9.7448	10.6005	11.9931	11.3639	11.2653
	97%	12.9962	36.2762	13.1013	14.2247	16.3794	15.0146	20.8861
(G)-NMAE	30%	0.0146	0.0438	0.0437	0.0566	0.0660	0.0775	0.0475
	50%	0.0271	0.0475	0.0473	0.0600	0.0705	0.0781	0.0518
	70%	0.0416	0.0539	0.0536	0.0647	0.0767	0.0788	0.0601
	90%	0.0654	0.0696	0.0714	0.0764	0.0860	0.0820	0.0855
	97%	0.0975	0.3644	0.0973	0.1014	0.1336	0.1105	0.1621
(H)-MAPE	30%	8.6480	21.6605	31.8677	20.5789	27.8822	24.9918	30.3576
	50%	13.4345	23.6602	39.9123	21.7194	31.0080	25.5648	34.8166
	70%	17.7279	26.1652	54.2691	22.2156	34.6403	28.0127	42.3733
	90%	25.2299	34.8887	103.1330	29.2018	50.8457	72.3720	69.6340
	97%	47.0237	57.1542	131.9420	54.9850	90.0668	192.2020	130.9310
(H)-NMAE	30%	0.0476	0.1193	0.1386	0.1189	0.2113	0.1394	0.1354
	50%	0.0719	0.1287	0.1515	0.1256	0.2388	0.1526	0.1491
	70%	0.0994	0.1420	0.1842	0.1350	0.2610	0.1730	0.1732
	90%	0.1604	0.1793	0.2768	0.1850	0.3182	0.3672	0.2741
	97%	0.2955	0.3547	0.4075	0.3115	0.3783	0.5514	0.4980

ACKNOWLEDGMENT

This work is supported by the Shenzhen Science and Technology Program (Grant No. ZDSYS20210623092007023) and the Educational Commission of Guangdong Province (Grant No. 2021ZDZX1069).

REFERENCES

- [1] X. Chen, Z. He, and J. Wang, "Spatial-temporal traffic speed patterns discovery and incomplete data recovery via SVD-combined tensor decomposition," *Transp. Res. Pt. C-Emerg. Technol.*, vol. 86, pp. 59–77, Jan. 2018.
- [2] M. Roughan, Y. Zhang, W. Willinger, and L. Qiu, "Spatio-temporal compressive sensing and internet traffic matrices (extended version)," *IEEE-ACM Trans. Netw.*, vol. 20, no. 3, pp. 662–676, Jun. 2012.
- [3] M. T. Bahadori, Q. R. Yu, and Y. Liu, "Fast multivariate spatio-temporal analysis via low rank tensor learning," in *Adv. Neural Inf. Process. Syst.*, vol. 27, Montreal, Canada, Jan. 2014, pp. 3491–3499.
- [4] 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.
- [5] H. Wang, J. Peng, W. Qin, J. Wang, and D. Meng, "Guaranteed tensor recovery fused low-rankness and smoothness," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 9, pp. 10990–11007, Sep. 2023.
- [6] N. Rao, H.-F. Yu, P. K. Ravikumar, and I. S. Dhillon, "Collaborative filtering with graph information: Consistency and scalable methods," in *Adv. Neural Inf. Process. Syst.*, vol. 28, Montreal, Canada, Jan. 2015, pp. 2107–2115.
- [7] X. Li, M. K. Ng, G. Cong, Y. Ye, and Q. Wu, "MR-NTD: Manifold regularization nonnegative Tucker decomposition for tensor data dimension reduction and representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 8, pp. 1787–1800, Aug. 2017.
- [8] W. Gong, Z. Huang, and L. Yang, "Spatiotemporal regularized Tucker decomposition for traffic data imputation," arXiv:2305.06563, 2023, to be published.
- [9] P. Sure, C. P. Srinivasan, and C. N. Babu, "Spatio-temporal constraint-based low rank matrix completion approaches for road traffic networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 13452–13462, Aug. 2022.
- [10] K. Xie, L. Wang, X. Wang, G. Xie, J. Wen, and G. Zhang, "Accurate recovery of internet traffic data: A tensor completion approach," in *Proc. IEEE INFOCOM*, San Francisco, United states, Jul. 2016.
- [11] B. Ran, H. Tan, Y. Wu, and P. J. Jin, "Tensor based missing traffic data completion with spatial-temporal correlation," *Physica A*, vol. 446, pp. 54–63, Mar. 2016.
- [12] X. Wang, Y. Wu, D. Zhuang, and L. Sun, "Low-rank Hankel tensor completion for traffic speed estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 5, pp. 4862–4871, May 2023.
- [13] X. Chen, Z. He, Y. Chen, Y. Lu, and J. Wang, "Missing traffic data imputation and pattern discovery with a Bayesian augmented tensor factorization model," *Transp. Res. Pt. C-Emerg. Technol.*, vol. 104, pp. 66–77, Jul. 2019.
- [14] A. B. Said and A. Erradi, "Spatiotemporal tensor completion for improved urban traffic imputation," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6836–6849, Jul. 2022.
- [15] Z. Zhang, C. Ling, H. He, and L. Qi, "A tensor train approach for internet traffic data completion," *Ann. Oper. Res.*, Jun. 2021.
- [16] X. Chen, Z. Cheng, N. Saunier, and L. Sun, "Laplacian convolutional representation for traffic time series imputation," arXiv:2212.01529, 2022, to be published.
- [17] W. Gong, Z. Huang, and L. Yang, "LSPTD: Low-rank and spatiotemporal priors enhanced Tucker decomposition for internet traffic data imputation," in *Proc. IEEE Conf Intell Transport Syst Proc ITSC*, Bilbao, Spain, 2023, pp. 460–465.
- [18] C. Pan, C. Ling, H. He, L. Qi, and Y. Xu, "A low-rank and sparse enhanced Tucker decomposition approach for tensor completion," *Appl. Math. Comput.*, vol. 465, Mar. 2024.
- [19] X. Chen, Y. Chen, N. Saunier, and L. Sun, "Scalable low-rank tensor learning for spatiotemporal traffic data imputation," *Transp. Res. Pt. C-Emerg. Technol.*, vol. 129, Aug. 2021.
- [20] X. Chen, M. Lei, N. Saunier, and L. Sun, "Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 12301–12310, Aug. 2022.