LSPTD: Low-rank and spatiotemporal priors enhanced Tucker decomposition for internet traffic data imputation

Wenwu Gong[1] and Zhejun Huang [2] and Lili Yang [3]

***Abstract*— Low-rank tensor methods and their relaxation forms have performed excellently in tensor completion problems, including internet traffic data imputation. However, most are based on the unfolding matrix's nuclear norm, which inevitably destroys the traffic tensor structure and significantly suffers from computation burden. Also, few consider the intrinsic spatiotemporal features, especially for the underlying spatial similarity. This paper proposes a novel low-rank and spatiotemporal priors enhanced Tucker decomposition (called LSPTD) for internet traffic data imputation. LSPTD model exploits the spatial similarity using factor graph embedding and characterizes the temporal correlation using the Toeplitz matrix. Two easily implementable algorithms and the closed-form updating rules are designed to solve the LSPTD model. Numerical experiments on the Abilene and GÉANT datasets demonstrate that our proposed model is superior to the other imputation methods in terms of NMAE and computation time.**

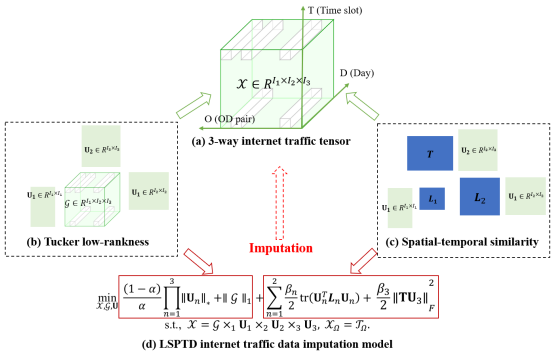

Fig. 1. Visual display for the proposed LSPTD model.

I. INTRODUCTION

Internet traffic data (ITD) records subjects' movement from an origin to a destination, involving pair, time, and day modes, which contain useful information for intelligent transportation systems. Unfortunately, the missing data problem frequently occurs due to communication malfunctions, transmission distortions, or adverse weather conditions [1]. As many traffic network engineering tasks require complete traffic information or are susceptible to missing data [2], internet traffic data imputation (ITDI), which refers to estimating the missing value from partial measurements, is still a typical and challenging problem.

ITD is often organized by matrices or tensors, which accurately store spatial and temporal information. Roughan et al. [3] first introduced a decomposition-based model using the low-rank matrix factorization to impute the missing traffic data. Wang et al. [4] then proposed an adaptive spatial-temporal constraint model to improve imputation performance by leveraging traffic network data's low rankness and spatial-temporal correlation. However, the matrix-based methods ignore the multidimensional nature, which cannot sufficiently capture the spatial-temporal features and are unsuitable for high-level missing scenarios.

To overcome the shortcomings of the matrix-based methods, Xie et al. [2] used the third-order traffic tensor and proposed a sequential tensor completion algorithm. Furthermore, Tan et al. [5] organized a fourth-way traffic flow tensor to cover the spatial-temporal correlations and achieved better performance. Therefore, keeping the tensor structure of the raw data is more appropriate for ITDI. In this paper, to minimize computation complexity, the raw ITD is structured as a third-order tensor with origin-destination (OD) pairs, time slots, and days as the three axes (refer to Fig. 1).

Exploiting the spatial and temporal relations among traffic data is the core idea for ITDI. On the one hand, Chen et al. [6] introduced an autoregressive term into the low-rank tensor completion framework to capture the temporal correlations and achieved good performance in both low- and high-level missing scenarios. On the other hand, decomposition-based, i.e., representing the tensor structure under learned lower dimensional subspace, imposes spatial-temporal constraints directly on ITD [7]–[9] to improve model performance. To our knowledge, the Tucker-based models perform more robustness and precisely, enhancing the learning capabilities of the traffic data imputation (TDI) [10]. So, we introduce the Tucker-based model for ITDI.

There are three major open questions yet to be addressed for ITDI. Firstly, the global low rankness of ITD has not been well defined. Considering that the Tucker rank relaxation is not unique, how to construct an effective low-rank Tucker is a crucial issue. Secondly, the spatial-temporal correlation should not be neglected. However, the existing models are often based on purely temporal regularization terms. Finally, the Tucker-based relaxed model is a nonconvex optimization problem, an efficient optimization algorithm is thus needed to find the stationary points and attain optimal solutions.

To address the aforementioned shortcomings, a novel Tucker-based low-rank with spatiotemporal prior method,

[1] W. Gong (ORCID: 0000-0002-8019-0582) is a Ph.D. student at the Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, 518055, China.

[2] Zhejun Huang with the Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, 518055, China.

[3] Lili Yang with the Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, 518055, China.

called **L**ow-rank **S**patiotemporal **P**riors enhanced **T**ucker **D**ecomposition (LSPTD) model, is proposed in this paper (Fig 1 gives a visual display for the LSPTD model). The LSPTD model encodes the global low rankness in a Tucker manner. It exploits the spatial-temporal correlation using factor graph embedding (GE) prior [11] and Toeplitz matrix constraints to reduce the imputation error. The LSPTD method outperforms the existing imputation methods on real-world datasets regarding accuracy and computation time. The main contributions of this paper are summarized as follows.

1) Motivated by [12] and [13], we use the weighted factor matrix rank and core tensor sparsity to encode the global low Tucker rank. Furthermore, the factor matrix weight is self-adaptive, and a tradeoff parameter is used to balance the low-rank and sparsity.
2) To further characterize the spatial-temporal correlations, we consider the factor GE prior and the Toeplitz matrix as the smoothness constraints to capture such similarities, which improves the model's ability and reduces the imputation error.
3) Two optimization algorithms are specifically designed and our algorithms have closed-form update rules.

TABLE I
NOTATIONS

$\mathcal{X}, \mathbf{U}, \alpha$	A tensor, matrix and real value, respectively.
$\Omega, \bar{\Omega}$	Observed index set and its complement.
$\mathcal{S}_\eta(x)$	Shrinkage operator with η in component-wise.
$\mathcal{D}_\eta(\mathbf{U})$	Singular Value Decomposition (SVD) shrinkage of matrix $\mathbf{U}$.
$\mathcal{X}_\Omega$	Observed entries supported on the observed index.
$\times_n$	Mode-n product.
$\otimes$	Kronecker product.
tr	Trace operator.
$\|\cdot\|_F$	Frobenius norm.
$\|\cdot\|_*$	Nuclear norm.
$\mathbf{X}_{(n)}$	Mode-n unfolding of tensor $\mathcal{X}$.

II. NOTATIONS AND RELATED WORKS

A. Notations

We give related concepts of Tucker decomposition as follows and present all notations used in this paper in Table I (please refer to [14] for more details).

Tensors are multidimensional arrays, the higher-order generalization of vectors and matrices. Given a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, it can be decomposed into a core tensor $\mathcal{G} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ multiplying a matrix $\mathbf{U}_n \in \mathbb{R}^{I_n \times I_n}$ along each mode, i.e., $\mathcal{X} = \mathcal{G} \times_1 \mathbf{U}_1 \cdots \times_N \mathbf{U}_N = \mathcal{G} \times_{n=1}^{N} \mathbf{U}_n$. We call that Tucker decomposition, which represents the underlying structure of the tensor.

Based on the matrix Kronecker product $\otimes$, the Tucker decomposition can be written in the matrix form:

$$\mathbf{X}_{(n)} = \mathbf{U}_n \mathbf{G}_{(n)} \mathbf{V}_n^{\mathrm{T}}$$

where $\mathbf{V}_n = (\mathbf{U}_N \otimes \cdots \otimes \mathbf{U}_{n+1} \otimes \mathbf{U}_{n-1} \otimes \cdots \otimes \mathbf{U}_1)$ and the superscript 'T' represent matrix transpose. It is not difficult to verify that $\mathrm{vec}(\mathcal{X}) = (\mathbf{U}_\mathrm{N} \otimes \cdots \otimes \mathbf{U}_\mathrm{n} \otimes \cdots \otimes \mathbf{U}_1)\,\mathrm{vec}(\mathcal{G}) = \otimes_{\mathrm{n=N}}^{1} \mathbf{U}_\mathrm{n} \mathrm{vec}(\mathcal{G})$.

TABLE II
SOME EXISTING TUCKER-BASED TRAFFIC DATA IMPUTATION METHODS

TC methods	Types of the prior		
	Low-rank	Spatial	Temporal
LSPTD	✓	✓	✓
ManiRTD [10]	✓	✓	✓
LR-SETD [8]	✓		✓
LATC [6]	✓		✓
STD [21]	✓		
IFHST [22]	✓		
Tucker-Wopt [19]	✓		

B. Related works

In general, internet traffic tensor data has spatial similarity and temporal correlation features; many studies have proven that the low-rank assumption combined with the spatiotemporal prior performs better for ITDI [4], [15], [16].

Low-rank property, which depicts the inherent global correlations in ITD, is an essential assumption in the TDI problem. Tan et al. [17] first modeled the traffic data as a third-order tensor and applied the low-rank tensor nuclear norm minimization method to estimate the traffic flow. However, the nuclear norm minimization in the unfolding matrix fails to exploit the tensor structure [18] and cannot impose spatial-temporal constraints directly on data matrices [4]. To overcome these shortcomings, Tan et al. [19] applied Tucker decomposition to TDI. Chen et al. [1] proposed a Bayesian augmented CANDECOMP/PARAFAC (CP) decomposition model for traffic data analysis and combined domain knowledge to enhance the imputation performance. Furthermore, Zhang et al. [9] introduced the tensor train (TT)-based models to improve the recovery performance. Hence, the low-rank decomposition model is more appropriate for ITDI.

Given the spatiotemporal features in ITD noted in [3], regularization terms were considered in many papers. Rose et al. [20] proposed a unified low-rank tensor learning framework considering spatial similarity by constructing a Laplacian regularizer for spatial-temporal data analysis. Chen et al. [21] exploited the property of the low rankness and l_2 norm regularization of factor matrices in the Tucker model to estimate the missing traffic speed. Besides, Wu et al. [7] proposed a modified CP decomposition framework to capture spatial-temporal features in the ITDI problem.

The joint priors of low Tucker rank and spatial-temporal were also researched. Pan et al. [8] proposed a sparse enhanced Tucker model in the tensor completion problem. Gong et al. [10] reshaped a 3rd/4th order Hankel traffic tensor and proposed an innovative manifold regularized Tucker decomposition model for traffic data imputation. However, the spatiotemporal characteristics of ITD have yet to be fully explored. The main attributes of our proposed LSPTD model and that of several Tucker-based popular ITDI methods are listed in Tab. II.

III. PROPOSED MODEL AND ALGORITHMS

We propose a Tucker-based model and consider low-rank and spatiotemporal priors for ITDI. Firstly, we use

the weighted factor matrix rank and core tensor sparsity to encode the global low rankness; Then, we consider the factor GE prior and the Toeplitz matrix to characterize the spatial-temporal correlation, which improves the model's ability to deal with internet traffic data. Finally, two optimization algorithms based on the Gauss-Seidel scheme [23] are designed to solve the proposed LSPTD model.

A. Low-rank enhanced Tucker decomposition

Considering that Tucker is an effective and efficient model to express the global low rankness, we use the Tucker decomposition to characterize the global low-rank prior.

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

***Remark 1*: The low-rank enhanced Tucker decomposition differs from the KBR-based [12] and SBCD-based [13] model with the factor matrix constraint. The reasonableness of our construction is summarised as follows: minimizing factor matrix rank in Tucker decomposition is equivalence to minimizing the unfolding matrix rank; the low rankness of factor matrix implies that many elements of the core tensor equal zero; it is therefore rational to impose sparsity on the core tensor.**

B. Spatiotemporal prior enhanced Tucker decomposition

Since internet traffic tensor data often have spatial and temporal features [10], we use the factor GE prior and the Toeplitz matrix as smoothness constraints to characterize the spatiotemporal correlations and further improve the model performance. The proposed LSPTD model is defined as

$$\begin{aligned}&\min_{\mathcal{G},\mathbf{U}_n}\ \frac{(1-\alpha)}{\alpha}\prod_{n=1}^{N}\|\mathbf{U}_n\|_*+\|\mathcal{G}\|_1\\&\qquad+\sum_{n=1}^{N-1}\frac{\beta_n}{2}\operatorname{tr}\left(\mathbf{U}_n^{\mathrm{T}}\mathbf{L}_n\mathbf{U}_n\right)+\frac{\beta_N}{2}\|\mathbf{T}\mathbf{U}_N\|_F^2\\&\text{s.t. }\ \mathcal{X}=\mathcal{G}\times_{n=1}^{N}\mathbf{U}_n,\\&\qquad 0<\alpha<1,\quad \beta_n\geq 0,\ n=1,2,\cdots,N.\end{aligned}\tag{1}$$

To assess the effectiveness of the spatial and temporal priors, we present the comparative results obtained from our numerical experiments (refer to Fig. 4).

***Remark 2*: the product function of the LSPTD model is nonconvex, which is hard to solve. Also, the product only represents the block size of the core tensor, we thus use the weighted factor matrix nuclear norm summation term in our algorithm design.**

C. Algorithms

We introduce the ITDI problem (2) based on the LSPTD model (1) with setting $N=3$ since our numerical experiments focus on third-order tensors.

$$\begin{aligned}&\min_{\mathcal{G},\{\mathbf{U}_n\}}\ \lambda\sum_{n=1}^{3}\omega_n\|\mathbf{U}_n\|_*+\|\mathcal{G}\|_1\\&\qquad+\sum_{n=1}^{2}\frac{\beta_n}{2}\operatorname{tr}\left(\mathbf{U}_n^{\mathrm{T}}\mathbf{L}_n\mathbf{U}_n\right)+\frac{\beta_3}{2}\|\mathbf{T}\mathbf{U}_3\|_F^2,\\&\text{s.t. }\mathcal{X}=\mathcal{G}\times_{n=1}^{3}\mathbf{U}_n,\quad \mathcal{X}_\Omega=\mathcal{T}_\Omega,\end{aligned}\tag{2}$$

where

$$\lambda=\frac{1-\alpha}{\alpha},\ 0<\alpha<1,\ \omega_n=\prod_{i=1,i\neq n}^{3}\frac{1}{R_i},\ R_i=\sum\sigma\left(\mathbf{U}_i\right).$$

Model (2) is a nonconvex minimization problem (mode-n is a bilinear function) containing many local minima; it is much harder to find the optimal solution of (2). However, every subproblem can be simplified to a convex optimization problem based on the Gauss-Seidel scheme, and we propose two algorithms for ITDI. (Note that our algorithms can be extended to higher-order tensor cases, we derive the algorithmic details by setting $N=3$.)

Algorithm 1: Proximal Alternating Linearized Minimization. We penalize (2) into the (3) and apply the proximal alternating linearized parallel multi-block minimization method (PLAM) to solve the LSPTD model.

$$\begin{aligned}&\min_{\mathcal{G},\{\mathbf{U}_n\},\mathcal{X}}\ \lambda\sum_{n=1}^{3}\omega_n\|\mathbf{U}_n\|_*+\|\mathcal{G}\|_1+\frac{\mu}{2}\left\|\mathcal{X}-\mathcal{G}\times_{n=1}^{3}\mathbf{U}_n\right\|_{\mathrm{F}}^2\\&\qquad+\sum_{n=1}^{2}\frac{\beta_n}{2}\operatorname{tr}\left(\mathbf{U}_n^{\mathrm{T}}\mathbf{L}_n\mathbf{U}_n\right)+\frac{\beta_3}{2}\|\mathbf{T}\mathbf{U}_3\|_F^2,\\&\text{s.t. }\mathcal{X}_\Omega=\mathcal{T}_\Omega,\quad \mathcal{X}_{\bar{\Omega}}=\{\mathcal{G}\times_{n=1}^{3}\mathbf{U}_n\}_{\bar{\Omega}}.\end{aligned}\tag{3}$$

- Optimization of $\mathcal{G}$. We have the LASSO problem

$$\hat{\mathcal{G}}=\ \operatorname*{argmin}_{\mathcal{G}}\ \frac{\mu L_{\mathcal{G}}}{2}\left\|\mathcal{G}-(\tilde{\mathcal{G}}-\frac{1}{L_{\mathcal{G}}}\nabla_{\mathcal{G}}f\left(\tilde{\mathcal{G}}\right))\right\|_{\mathrm{F}}^2+\|\mathcal{G}\|_1\tag{4}$$

  where $\tilde{\mathcal{G}}$ denotes the extrapolated point and the $\nabla_{\mathcal{G}}f(\mathcal{G})$ can be calculated by (5), and the Lipschitz constant is given by $L_{\mathcal{G}}=\left\|\otimes_{n=3}^{1}\mathbf{U}_n^{\mathrm{T}}\mathbf{U}_n\right\|_2=\prod_{n=1}^{3}\left\|\mathbf{U}_n^{\mathrm{T}}\mathbf{U}_n\right\|_2$.

$$\nabla_{\mathcal{G}}f(\mathcal{G})=\ \mathcal{G}\times_{n=1}^{3}\mathbf{U}_n^{\mathrm{T}}\mathbf{U}_n-\mathcal{X}\times_{n=1}^{3}\mathbf{U}_n^{\mathrm{T}}\tag{5}$$

  Then, the close form of the Tucker core tensor updating rule (6) is given by

$$\hat{\mathcal{G}}=\mathcal{S}_{\frac{1}{\mu L_{\mathcal{G}}}}(\tilde{\mathcal{G}}-\frac{1}{L_{\mathcal{G}}}\nabla_{\mathcal{G}}f\left(\tilde{\mathcal{G}}\right)),\tag{6}$$

  and $\tilde{\mathcal{G}}$ is updated by (7) with the updated step size η_k for $k\geq 1$

$$\tilde{\mathcal{G}}^k=\mathcal{G}^k+\eta_k\left(\mathcal{G}^k-\mathcal{G}^{k-1}\right),\tag{7}$$

$$\eta_k=\tfrac{t^{k-1}-1}{t^k},\ t^k=\tfrac{1+\sqrt{4(t^{k-1})^2+1}}{2},\ t^0=1.\tag{8}$$

- Optimization of $\mathbf{U}_n$. By performing mode-n unfolding (we detail the GE prior, and the gradient of Toeplitz constraint has the $\left(\beta_n\mathbf{U}_n\mathbf{T}\mathbf{T}^{\mathrm{T}}\right)$ form), with $\mathbf{U}_j,\ j\leq n$

and other parameters fixed, we have the matrix with nuclear norm optimization

$$\hat{\mathbf{U}}_n = \underset{\mathbf{U}_n}{\operatorname{argmin}} \frac{L_{\mathbf{U}_n}}{2} \left\| \mathbf{U}_n - (\tilde{\mathbf{U}}_n - \frac{1}{L_{\mathbf{U}_n}} \nabla_{\mathbf{U}_n} f\left(\tilde{\mathbf{U}}_n\right)) \right\|_{\mathrm{F}}^2 + \frac{\lambda \omega_n}{\mu} \left\| \mathbf{U}_n \right\|_*. \tag{9}$$

We calculate the gradient and its Lipschitz constant as

$$\nabla_{\mathbf{U}_n} f(\mathbf{U}_n) = \mathbf{U}_n \mathbf{G}_{(n)} \mathbf{V}_n^{\mathrm{T}} \mathbf{V} \mathbf{G}_{(n)}^{\mathrm{T}} - \mathbf{X}_{(n)} \mathbf{V} \mathbf{G}_{(n)}^{\mathrm{T}} + \beta_n \mathbf{L}_n \mathbf{U}_n$$

$$L_{\mathbf{U}_n} = \left\| \mathbf{G}_{(n)} \mathbf{V}_n^{\mathrm{T}} \mathbf{V} \mathbf{G}_{(n)}^{\mathrm{T}} \right\|_2 + \left\| \beta_n \mathbf{L}_n \right\|_2$$

Then, the closed-form solution of the nuclear norm proximal operator is

$$\hat{\mathbf{U}}_n = \mathcal{D}_{\frac{\lambda \omega_n}{\mu L_{\mathbf{U}_n}}} \left(\tilde{\mathbf{U}}_n - \frac{1}{L_{\mathbf{U}_n}} \nabla_{\mathbf{U}_n} f\left(\tilde{\mathbf{U}}_n\right) \right), \tag{10}$$

and $\tilde{\mathbf{U}}_n$ is updated using (11) with η_k

$$\tilde{\mathbf{U}}_n^k = \mathbf{U}_n^k + \eta_k \left(\mathbf{U}_n^k - \mathbf{U}_n^{k-1} \right), \text{ for } k \geq 1 \tag{11}$$

- The optimal solution $\mathcal{X}$ satisfies (12)

$$\hat{\mathcal{X}}_\Omega = \mathcal{T}_\Omega, \ \hat{\mathcal{X}}_{\bar{\Omega}} = \left(\hat{\mathcal{G}} \times_{n=1}^3 \hat{\mathbf{U}}_n \right)_{\bar{\Omega}} \tag{12}$$

The proposed PLAM-based algorithm for ITDI can be summarized in Algorithm 1.

Algorithm 1 PLAM-based LSPTD

Input: Missing traffic tensor $\mathcal{T}$, observed entries Ω.
Output: Imputed traffic tensor $\hat{\mathcal{X}}$.
Initialize $\mathcal{G}^0, \{\mathbf{U}_n^0\}$ $(1 \leq n \leq 3)$ randomly, $0 < \alpha < 1$, $\mu = 1$, and define $\mathcal{Z}^0$ as null tensor;
$\mathcal{X}_\Omega = \mathcal{T}_\Omega$, $\mathcal{X}_{\bar{\Omega}} = \mathcal{Z}_{\bar{\Omega}}^0$;
while $k < K$ **do**
  Optimize $\mathcal{G}$ by Eq. (6);
  for $n = 1$ to 3 **do**
    Optimize $\mathbf{U}_n$ using Eq. (10);
  end for
  Update Tucker decomposition $\mathcal{Z}^k$ using $\{\mathbf{U}_n^k\}$ and $\mathcal{G}^k$;
  if $\mathbb{F}\left(\mathcal{G}^k, \mathbf{U}_{j \leq n}, \mathbf{U}_{j > n}\right)$ is decreasing **then**
    Re-update $\mathcal{G}^k$ and $\mathbf{U}_n^k$ respectively;
  else
    Re-update $\mathcal{G}^k$ and $\mathbf{U}_n^k$ respectively with $\tilde{\mathcal{G}}^k = \mathcal{G}^{k-1}$ and $\tilde{\mathbf{U}}_n^k = \mathbf{U}_n^{k-1}$;
  end if
  until $\left\| \mathcal{X}^{k+1} - \mathcal{X}^k \right\|_F \left\| \mathcal{X}^k \right\|_F^{-1} < 1e^{-3}$ are satisfied.
end while
return $\hat{\mathcal{X}}_\Omega = \mathcal{T}_\Omega, \ \hat{\mathcal{X}}_{\bar{\Omega}} = \left(\hat{\mathcal{G}} \times_{n=1}^3 \hat{\mathbf{U}}_n \right)_{\bar{\Omega}}$.

Remark 3: Following [13], we can prove that the sequences generated by Algorithm 1 have the stationary points. Here, we skip their convergence analysis for the conciseness of the paper.

Algorithm 2: Inexact Augmented Lagrange Multiplier Framework. To solve for the variables in (2), we define the augmented Lagrange function as follows

$$\begin{aligned} \min_{\mathcal{G}, \mathbf{U}_n, \mathcal{X}} \ & \lambda \sum_{n=1}^3 \omega_n \left\| \mathbf{U}_n \right\|_* + \left\| \mathcal{G} \right\|_1 \\ & + \sum_{n=1}^2 \frac{\beta_n}{2} \operatorname{tr}\left(\mathbf{U}_n^{\mathrm{T}} \mathbf{L}_n \mathbf{U}_n \right) + \frac{\beta_3}{2} \left\| \mathbf{T} \mathbf{U}_3 \right\|_F^2 \\ & + \frac{\mu}{2} \left\| \mathcal{X} - \mathcal{G} \times_{n=1}^3 \mathbf{U}_n \right\|_{\mathrm{F}}^2 + \left\langle \mathcal{Y}, \mathcal{X} - \mathcal{G} \times_{n=1}^3 \mathbf{U}_n \right\rangle. \end{aligned} \tag{13}$$

where μ is a positive scalar that is adaptively changing, and $\mathcal{Y}$ is the Lagrange multiplier. Now we can solve the problem under the ADMM framework and apply the linearization technique to solve the separable optimization problem.

- Optimization of $\mathcal{G}$.

$$\begin{aligned} \hat{\mathcal{G}} = \ & \mathcal{S}_{\frac{1}{\mu L_{\mathcal{G}}}} (\tilde{\mathcal{G}} - \frac{1}{L_{\mathcal{G}}} \nabla_{\mathcal{G}} f\left(\tilde{\mathcal{G}}\right)), \ L_{\mathcal{G}} = \prod_{n=1}^3 \left\| \mathbf{U}_n^{\mathrm{T}} \mathbf{U}_n \right\|_2 \\ \nabla_{\mathcal{G}} f(\mathcal{G}) = & \ \mathcal{G} \times_{n=1}^3 \mathbf{U}_n^{\mathrm{T}} \mathbf{U}_n - \left(\mathcal{X} + \frac{\mathcal{Y}}{\mu} \right) \times_{n=1}^3 \mathbf{U}_n^{\mathrm{T}}. \end{aligned} \tag{14}$$

- Optimization of $\mathbf{U}_n$ with $\mathbf{U}_j, j \neq n$ and other parameters fixed.

$$\begin{aligned} \hat{\mathbf{U}}_n = & \ \mathcal{D}_{\frac{\lambda \omega_n}{L_{\mathbf{U}_n}}} \left(\tilde{\mathbf{U}}_n - \frac{1}{L_{\mathbf{U}_n}} \nabla_{\mathbf{U}_n} f\left(\tilde{\mathbf{U}}_n\right) \right). \\ \nabla_{\mathbf{U}_n} f(\mathbf{U}_n) = & \ \mu \mathbf{U}_n \mathbf{G}_{(n)} \mathbf{V}_n^{\mathrm{T}} \mathbf{V} \mathbf{G}_{(n)}^{\mathrm{T}} + \beta_n \mathbf{L}_n \mathbf{U}_n \\ & - \left(\mu \mathbf{X}_{(n)} + \mathbf{Y}_{(n)} \right) \mathbf{V} \mathbf{G}_{(n)}^{\mathrm{T}}, \\ L_{\mathbf{U}_n} = & \left\| \mu \mathbf{G}_{(n)} \mathbf{V}_n^{\mathrm{T}} \mathbf{V} \mathbf{G}_{(n)}^{\mathrm{T}} \right\|_2 + \left\| \beta_n \mathbf{L}_n \right\|_2 \end{aligned} \tag{15}$$

- Optimization of $\mathcal{X}$.

$$\begin{aligned} & \min_{\mathcal{X}} \left\| \mathcal{X} - \left(\mathcal{G} \times_{n=1}^3 \mathbf{U}_n - \frac{\mathcal{Y}}{\mu} \right) \right\|_{\mathrm{F}}^2, \\ & \text{s.t. } \mathcal{X}_\Omega = \mathcal{T}_\Omega \\ & \hat{\mathcal{X}}_\Omega = \mathcal{T}_\Omega, \ \hat{\mathcal{X}}_{\bar{\Omega}} = \left(\hat{\mathcal{G}} \times_{n=1}^3 \hat{\mathbf{U}}_n - \frac{\mathcal{Y}_k}{\mu_k} \right)_{\bar{\Omega}} \end{aligned} \tag{16}$$

- Update $\mathcal{Y}$.

$$\begin{aligned} & \mathcal{Y}^{k+1} = \hat{\mathcal{Y}} + \mu^k \left(\hat{\mathcal{X}} - \hat{\mathcal{G}} \times_{n=1}^3 \hat{\mathbf{U}}_n \right), \\ & \mu^{k+1} = \rho \mu^k, \ \rho \in [1.1, 1.2]. \end{aligned} \tag{17}$$

The proposed IALM-based algorithm can be summarized in Algorithm 2.

Remark 4: Algorithm 2 exploits linearization to solve (2)**. The convergence of IALM has been proven for convex problems using linearization [24]. Since the Tucker decomposition and factor priors, deriving a theoretical guarantee for Algorithm 2 is difficult.**

IV. Numerical Experiments

This section presents the experimental results on real ITD, including the Abilene [1] $(121 * 288 * 7)$ and the GÉANT [2]

[1] https://doi.org/10.5281/zenodo.7725126
[2] https://totem.info.ucl.ac.be/dataset.html

Algorithm 2 IALM-based LSPTD

Input: Missing traffic tensor $\mathcal{T}$, observed entries Ω.
Output: Imputed traffic tensor $\hat{\mathcal{X}}$.
Initialize: $\mathcal{G}^0, \{\mathbf{U}_n^0\}$ $(1 \leq n \leq 3)$, $0 < \alpha < 1$, $\mu^0 = 1e^{-5}$, and define $\mathcal{Z}^0$ as null tensor;
$\mathcal{X}_\Omega = \mathcal{T}_\Omega$, $\mathcal{X}_{\bar{\Omega}} = \mathcal{Z}^0_{\bar{\Omega}}$;
while $k < K$ **do**
  Optimize $\mathcal{G}^{k+1}$ via (14) with other variables fixed;
  Optimize all $\mathbf{U}_n^{k+1}$ via (15) with other variables fixed;
  Optimize $\mathcal{X}^{k+1}$ with other variables fixed;
  Update $\mathcal{Y}^{k+1}$ and $\mu^{k+1} = \rho\mu^k$, $\rho \in [1.1, 1.2]$
  until $\left\|\mathcal{X}^{k+1} - \mathcal{X}^k\right\|_F \left\|\mathcal{X}^k\right\|_F^{-1} < 1e^{-3}$ are satisfied.
end while
return $\hat{\mathcal{X}}_\Omega = \mathcal{T}_\Omega$, $\hat{\mathcal{X}}_{\bar{\Omega}} = \left(\hat{\mathcal{G}} \times_{n=1}^{3} \hat{\mathbf{U}}_n - \frac{\mathcal{Y}_K}{\mu_K}\right)_{\bar{\Omega}}$

$(529 * 96 * 7)$ dataset. All experiments are performed using MATLAB 2023a on Windows 10 64-bit operating system on a workstation equipped with an Intel(R) Xeon(R) W-2123 CPU with 3.60 GHz, 64 GB RAM.

Performance evaluation: As reported in the literature, we adopt the normalized mean absolute error (NMAE, the lower the better) value to measure the imputation performance. Specifically, the NMAE is defined as (18)

$$\text{NMAE} = \frac{\sum_{(i_1,i_2,i_3)\in\bar{\Omega}} \left|\hat{\mathcal{X}}_{i_1i_2i_3} - \mathcal{X}^*_{i_1i_2i_3}\right|}{\sum_{(i_1,i_2,i_3)\in\bar{\Omega}} \left|\mathcal{X}^*_{i_1i_2i_3}\right|}, \qquad (18)$$

where $\hat{\mathcal{X}}$ and $\mathcal{X}^*$ correspond to the imputed and real tensor, respectively.

Implementation details: Our experiments consider the random missing (RM) with the sample ratio (SR) from 0.9 to 0.05, and structurally missing (SM) scenarios [10]. In all our experiments, we set $\alpha = 0.5$ and calculate the SVD ratios between mode-n unfolding matrices and spatiotemporal matrices ($\{\mathbf{L}_1, \mathbf{L}_2\}$ and $\mathbf{T}$ in (1)) to deliver β_n, $n = 1, 2, 3$.

Model comparison: We select four baselines: ManiRTD [10], LR-SETD [8], LATC [6], and TAS-LR [4], to demonstrate the robustness and efficiency of the proposed LSPTD model. It is noteworthy that TRMF is a direct matrix-based approach, while others are tensor-based. We reorganize the traffic tensor and assign the parameters of baselines as described in the reference papers.

Figure 2 illustrates the NMAE values and computation time for different SRs in the RM scenarios. The results demonstrate that LSPTD exhibits competitive performance. Particularly, it outperforms other methods when the sample ratio exceeds 20% and achieves comparable accuracy when the sample ratio is below 10%. Additionally, LSPTD significantly reduces computing time compared to other methods. We further consider the SM scenarios in Figure 2, the NMAE values demonstrate that SM is harder to model, still, LSPTD performs better in most cases. In conclusion, our proposed LSPTD method is both more efficient and effective for ITDI.

Ablation study: We first discuss the effect of the

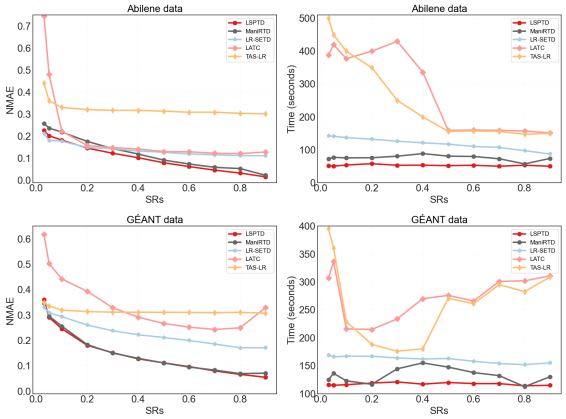

Fig. 2. RM results of ITDI methods based on NMAE values (left) and computing time (right) across different SRs for both the Abilene dataset (upper) and the GÉANT dataset (lower).

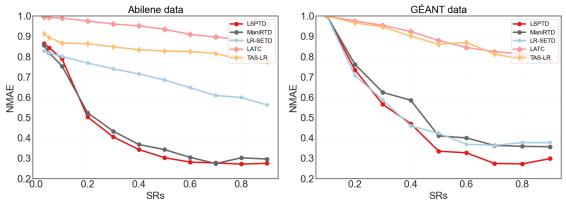

Fig. 3. SM results of ITDI methods based on NMAE values across different SRs for both the Abilene dataset (upper) and the GÉANT dataset (lower).

low-rank, spatial, and temporal priors of LSPTD on the performance in terms of NMAE for Abilene data. For simplicity, we denote 'LTD' (low-rank prior) with $\beta = (0, 0, 0)$, 'LTPTD'(low-rank and temporal priors) with $\beta = (0, 0, 1.23e^5)$, and 'LSPTD' (low-rank and spatiotemporal priors) with $\beta = 1e^5 * (0.22, 0.3, 1.23)$, respectively. Fig. 4 shows that LTD performs better than KBR and SBCD, which implies that our proposed global low rankness is more appropriate. Furthermore, using spatial priors leads to a 3% reduction in imputation error, whereas incorporating temporal priors results in an average performance enhancement of 8%, as depicted in Figure 4.

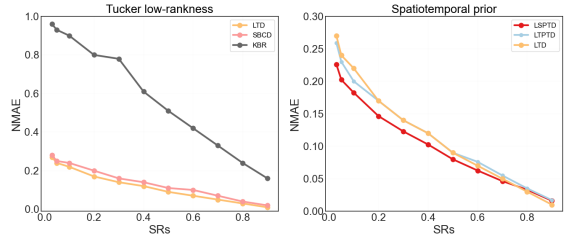

Fig. 4. NMAE versus SRs for KBR, SBCD, LTD, LTPTD, and LSPTD.

We investigate the performance of proposed Algorithms 1 and 2. Fig. 5 shows PALM-based algorithm has a better performance in accuracy, although it takes more computing time. The reason may be that we do re-update to make the objective nonincreasing. And the monotonicity of the

nonconvex objective plays a core role in performing stably.

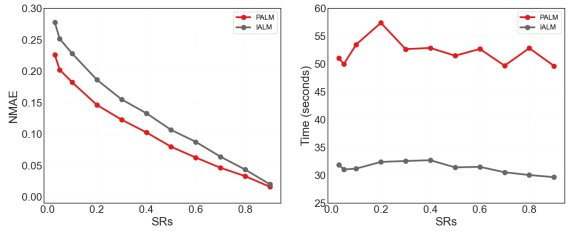

Fig. 5. Results of NMAE and computing time versus SRs via solving LSPTD model using Algorithm 1 and Algorithm 2.

V. Conclusion

This paper introduces a novel Tucker-based model with integrated low-rank and spatiotemporal priors for internet traffic data imputation. The proposed LSPTD model utilizes weighted factor matrix rank and core tensor sparsity to capture global low rankness. Additionally, it incorporates the factor GE and a Toeplitz matrix as spatiotemporal constraints to enhance model performance. To solve the LSPTD model, two optimization algorithms are specifically designed. A series of experiments validate the superiority of our proposal, showing its superior accuracy and efficiency compared to existing matrix-based and tensor-based methods. Actually, the current framework suffers a high computational cost for large-scale traffic data imputation. One can consider the fast Fourier transform to address this issue [25].

Acknowledgment: This research was funded by Shenzhen Science and Technology Plan platform and carrier special (Grant No. ZDSYS20210623092007023), Shenzhen Scientific Research Funding (Grant No. K22627501), and Guangdong Province Universities and Colleges Key Areas of Special Projects (Grant No. 2021222012).

References

[1] 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.

[2] K. Xie, L. Wang, X. Wang, G. Xie, J. Wen, and G. Zhang, "Accurate recovery of internet traffic data: A tensor completion approach," in *The 35th Annual IEEE International Conference on Computer Communications*, 2016, pp. 1–9.

[3] M. Roughan, Y. Zhang, W. Willinger, and L. Qiu, "Spatio-temporal compressive sensing and internet traffic matrices (extended version)," *IEEE/ACM Transactions on Networking*, vol. 20, no. 3, pp. 662–676, 2012.

[4] 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, 2019.

[5] B. Ran, H. Tan, Y. Wu, and P. J. Jin, "Tensor based missing traffic data completion with spatial–temporal correlation," *Physica A: Statistical Mechanics and its Applications*, vol. 446, pp. 54–63, 2016.

[6] 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*, pp. 1–10, 2021.

[7] Y. Wu, H. Tan, Y. Li, J. Zhang, and X. Chen, "A fused cp factorization method for incomplete tensors," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 3, pp. 751–764, 2019.

[8] C. Pan, C. Ling, H. He, L. Qi, and Y. Xu, "Low-rank and sparse enhanced tucker decomposition for tensor completion," 2020.

[9] Z. Zhang, C. Ling, H. He, and L. Q1, "A tensor train approach for internet traffic data completion," *Annals of Operations Research*, no. 6, pp. 860–889, 2021.

[10] W. Gong, Z. Huang, and L. Yang, "Manifold regularized tucker decomposition for spatiotemporal traffic data imputation," 2023.

[11] Y.-L. Chen, C.-T. Hsu, and H.-Y. M. Liao, "Simultaneous tensor decomposition and completion using factor priors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 577–591, 2014.

[12] Q. Xie, Q. Zhao, D. Meng, and Z. Xu, "Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 8, pp. 1888–1902, 2018.

[13] Q. Yu, X. Zhang, Y. Chen, and L. Qi, "Low tucker rank tensor completion using a symmetric block coordinate descent method," *Numerical Linear Algebra with Applications*, vol. 30, no. 3, p. e2464, 2023.

[14] T. G. Kolda and B. W. Bader, "Tensor decompositions and application," *SIAM Review*, vol. 5, no. 3, p. 455–500, 2009.

[15] A. B. Said and A. Erradi, "Spatiotemporal tensor completion for improved urban traffic imputation," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–14, 2021.

[16] P. Sure, C. P. Srinivasan, and C. N. Babu, "Spatio-temporal constraint-based low rank matrix completion approaches for road traffic networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 13 452–13 462, 2022.

[17] H. Tan, G. Feng, J. Feng, W. Wang, Y.-J. Zhang, and F. Li, "A tensor-based method for missing traffic data completion," *Transportation Research Part C: Emerging Technologies*, vol. 28, pp. 15–27, 2013.

[18] M. Yuan and C.-H. Zhang, "On tensor completion via nuclear norm minimization," *Foundations of Computational Mathematics*, vol. 16, p. 1031–1068, 2016.

[19] H. Tan, J. Feng, Z. Chen, F. Yang, and W. Wang, "Low multilinear rank approximation of tensors and application in missing traffic data," *Advances in Mechanical Engineering*, vol. 6, pp. 1575–1597, 2014.

[20] M. T. Bahadori, Q. R. Yu, and Y. Liu, "Fast multivariate spatio-temporal analysis via low rank tensor learning," in *Neural Information Processing Systems (NIPS)*, 2014, p. 3491–3499.

[21] X. Chen, Z. He, and J. Wang, "Spatial-temporal traffic speed patterns discovery and incomplete data recovery via svd-combined tensor decomposition," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 59–77, 2018.

[22] J. H. Goulart, A. Kibangou, and G. Favier, "Traffic data imputation via tensor completion based on soft thresholding of tucker core," *Transportation Research Part C Emerging Technologies*, vol. 85, pp. 348–362, 12 2017.

[23] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-Łojasiewicz inequality," *Mathematics of Operations Research*, vol. 35, no. 2, pp. 438–457, 2010.

[24] J. Yang and X. Yuan, "Linearized augmented lagrangian and alternating direction methods for nuclear norm minimization," *Mathematics of Computation*, vol. 82, pp. 301–329, 2012.

[25] X. Chen, Z. Cheng, N. Saunier, and L. Sun, "Laplacian convolutional representation for traffic time series imputation," 2022.