Structured Tensor RPCA: Anomaly Detection in Traffic Data

Xudong Wang[a], Luis F. Miranda-Moreno[a] and Lijun Sun[a,*]

[a] Department of Civil Engineering, McGill University, Montreal, Canada
xudong.wang2@mail.mcgill.ca, luis.miranda-moreno@mcgill.ca
lijun.sun@mcgill.ca
[*] Corresponding author

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)
September 02-03, 2024, Crete, Greece

April 11, 2024

Keywords: Anomaly detection, Hankel tensor, Metro system, Tensor RPCA

1 INTRODUCTION

Advancements in information and communication technologies (ICT) have enabled the collection of massive spatiotemporal traffic data, offering great opportunities to understand and manage transportation systems. Anomaly detection stands as a crucial task in traffic data analysis, distinguishing irregular occurrences such as passenger flow spikes, road congestion, or travel behavior changes from regular patterns. This detection assists transportation agencies in understanding system performance and adjusting traffic control plans promptly (Wang & Sun, 2021).

Robust Principal Component Analysis (RPCA) is widely utilized in anomaly detection due to its simplicity and non-parametric nature (Candès *et al.*, 2011). However, RPCA faces a fundamental limitation when dealing with spatiotemporal traffic data—it solely relies on the matrix structure and overlooks the spatiotemporal correlations within the data. To address this limitation, researchers have integrated additional regularizers (Wang *et al.*, 2018, Chen *et al.*, 2021) to capture the correlations. However, these regularizers are often constrained by their functional forms and may struggle to capture complex spatiotemporal patterns. Moreover, additional regularizers introduce new weight parameters, further complicating the model.

Motivated by the work in image processing (Jin & Ye, 2017), we propose an alternative solution – leveraging a specialized data structure (e.g., Hankel structure) – to capture the spatiotemporal correlations. Hankel structure is a data augmentation technique, where each descending skew-diagonal from left to right remains constant. Through data repetition, intrinsic temporal correlation can be naturally captured without requiring nonlocal similarity learning. Unlike the use of Hankel matrix RPCA in Jin & Ye (2017), we introduce a Hankel-structured tensor RPCA (HT-RPCA) framework for traffic anomaly detection. This tensor framework captures correlations in higher dimensions and accelerates estimation in the Fourier domain.

Mathematically speaking, we decompose the raw spatiotemporal traffic matrix into a low-rank matrix with Hankel constraint and a sparse matrix by solving a convex optimization problem. In this framework, the low-rank component represents regular patterns while sparisty component denotes irregular patterns (anomalies). We solve the HT-RPCA model by Alternating Direction Method of Multipliers (ADMM). Specifically, we apply tensor nuclear norm (Lu *et al.*, 2019) to approximate the tensor rank and l_1 norm to approximate sparsity. To evaluate the performance of HT-RPCA, we use metro passenger flow collected from Guangzhou, China, to detect unusual passenger flow at the station level and analyze anomaly propagation within the metro network.

2 METHODOLOGY

In this study, we introduce a temporal Hankelization operator $\mathcal{H}_\tau$ with a temporal delay embedding length τ. This operator transforms a given spatiotemporal matrix $\boldsymbol{X} \in \mathbb{R}^{N\times T}$ to a third-order Hankel tensor $\boldsymbol{\mathcal{X}} = \mathcal{H}_\tau(\boldsymbol{X}) \in \mathbb{R}^{N\times(T-\tau+1)\times\tau}$, defined as follows:

$$\boldsymbol{\mathcal{X}}_{:,:,t} = \boldsymbol{X}_{:,t:t+T-\tau+1} \in \mathbb{R}^{N\times(T-\tau+1)},\ t = 1,\ldots,\tau. \tag{1}$$

Correspondingly, the inverse Hankelization operation $\mathcal{H}_\tau^{-1}$ transforms a Hankel tensor $\boldsymbol{\mathcal{X}}$ back into a matrix $\hat{\boldsymbol{X}}$ by averaging the corresponding entries in the Hankel tensor (Wang *et al.*, 2023).

We denote the spatiotemporal data collected from N locations/sensors over T timestamps by $\boldsymbol{M} \in \mathbb{R}^{N\times T}$. We assume that the corrupted matrix $\boldsymbol{M}$ can be decomposed into a low-rank matrix $\boldsymbol{L} \in \mathbb{R}^{N\times T}$ with Hankel constraint, where $\boldsymbol{L} = \mathcal{H}_\tau^{-1}(\boldsymbol{\mathcal{L}})$. Here, $\boldsymbol{\mathcal{L}}$ is a low-rank Hankel tensor in $\mathbb{R}^{N\times(T-\tau+1)\times\tau}$, and $\boldsymbol{S} \in \mathbb{R}^{N\times T}$ is a sparse matrix. Using the temporal Hankelization operator, we transform the matrix-based RPCA problem into a tensor-based RPCA problem. Therefore, the HT-RPCA model can be formulated as the following optimization problem: $\min \boldsymbol{\mathcal{L}}, \boldsymbol{S}\ \operatorname{rank}(\boldsymbol{\mathcal{L}}) + \gamma\|\boldsymbol{S}\|_0$ with appropriate constraints. In our model, we use tensor nuclear norm (TNN), denoted as $\|\cdot\|_*$, calculated in the Fourier domain to approximate the tensor rank. This choice is motivated by its ability to provide a tight convex relaxation of the tensor average rank (Lu *et al.*, 2019). Additionally, we use the l_1 norm to approximate the sparsity of the matrix. The optimization problem for our HT-RPCA model is formulated as follows:

$$\min_{\boldsymbol{\mathcal{L}},\boldsymbol{S}}\ \|\boldsymbol{\mathcal{L}}\|_* + \gamma\|\boldsymbol{S}\|_1, \quad \text{s.t. } \boldsymbol{\mathcal{L}} = \mathcal{H}_\tau(\boldsymbol{L}) \text{ and } \boldsymbol{L} + \boldsymbol{S} = \boldsymbol{M}. \tag{2}$$

The convex optimization problem in (2) can be efficiently solved using the Alternating Direction Method of Multipliers (ADMM) framework. The augmented Lagrangian function is defined as:

$$\mathcal{L}(\boldsymbol{\mathcal{L}},\boldsymbol{S},\boldsymbol{E}) = \|\boldsymbol{\mathcal{L}}\|_* + \gamma\|\boldsymbol{S}\|_1 + \frac{\rho}{2}\|\boldsymbol{M} - \mathcal{H}_\tau^{-1}(\boldsymbol{\mathcal{L}}) - \boldsymbol{S}\|_F^2 + \langle \boldsymbol{M} - \mathcal{H}_\tau^{-1}(\boldsymbol{\mathcal{L}}) - \boldsymbol{S}, \boldsymbol{E}\rangle, \tag{3}$$

where $\rho > 0$ is a penalty parameter and $\boldsymbol{E}$ is a dual variable. The optimization problem (2) can be solved iteratively using (3). The updates for variables $\boldsymbol{\mathcal{L}}$, $\boldsymbol{S}$, and $\boldsymbol{E}$ at the ℓth iteration are as follows:

$$\begin{aligned}
\boldsymbol{\mathcal{L}}^{\ell+1} :&= \arg\min_{\boldsymbol{\mathcal{L}}} \frac{1}{\rho}\|\boldsymbol{\mathcal{L}}^\ell\|_* + \frac{1}{2}\|\boldsymbol{\mathcal{L}}^\ell - \mathcal{H}_\tau(\boldsymbol{M}^\ell - \boldsymbol{S}^\ell + \frac{1}{\rho^\ell}\boldsymbol{E}^\ell)\|_F^2 \\
&= \mathcal{D}_{1/\rho}(\mathcal{H}_\tau(\boldsymbol{M}^\ell - \boldsymbol{S}^\ell + \frac{1}{\rho^\ell}\boldsymbol{E}^\ell)). \\
\boldsymbol{S}^{\ell+1} :&= \arg\min_{\boldsymbol{S}} \frac{\gamma}{\rho}\|\boldsymbol{S}^\ell\|_1 + \frac{1}{2}\|\boldsymbol{S}^\ell - (\boldsymbol{M}^\ell - \mathcal{H}_\tau^{-1}(\boldsymbol{\mathcal{L}}^{\ell+1}) + \frac{1}{\rho^\ell}\boldsymbol{E}^\ell)\|_F^2 \\
&= \mathcal{S}_{\gamma/\rho}(\boldsymbol{M}^\ell - \mathcal{H}_\tau^{-1}(\boldsymbol{\mathcal{L}}^{\ell+1}) + \frac{1}{\rho^\ell}\boldsymbol{E}^\ell). \\
\boldsymbol{E}^{\ell+1} &= \boldsymbol{E}^\ell + \rho^\ell(\boldsymbol{M} - \mathcal{H}_\tau^{-1}(\boldsymbol{\mathcal{L}}^{\ell+1}) - \boldsymbol{S}^{\ell+1}).
\end{aligned} \tag{4}$$

Here, $\mathcal{D}.(\cdot)$ denotes the tensor singular value thresholding (Lu *et al.*, 2019), and $\mathcal{S}.(\cdot)$ denotes the soft shrinkage operator (Candès *et al.*, 2011). To accelerate the estimation process, we adjust the penalty parameter as $\rho^{\ell+1} = \beta\rho^\ell$, where $\beta \in [1.0, 1.2]$. The algorithm stops when it meets the convergence criterion, with a stopping threshold denoted as tol:

$$\frac{||\boldsymbol{M} - \mathcal{H}_\tau^{-1}(\boldsymbol{\mathcal{L}}^{\ell+1}) - \boldsymbol{S}^{\ell+1}||_F}{||\boldsymbol{M}||_F} < tol, \tag{5}$$

The proposed model has three parameters: the Hankel delay embedding length τ, the low-rank and sparse trade-off parameter γ, and the penalty parameter ρ. Given that traffic data often display weekly patterns due to regular human behavior, τ can be set to the length of a week to capture this periodicity. The choice of γ depends on the specific application, and ρ is typically a small number (e.g., 1×10^{-5}) as it updates at each iteration.

3 RESULTS

To evaluate HT-RPCA's performance, we use boarding passenger flow data from Guangzhou metro stations in China, covering 159 stations in July 2017 at a 15-minute resolution. Represented as a matrix $\boldsymbol{M} \in \mathbb{R}^{159\times(72\times20)}$, the dataset excludes data from 0:00 a.m. to 5:45 a.m. (no services) and weekends (travel behaviors differ). We compare the proposed model with RPCA (Candès *et al.*, 2011) and RPCA with Toepliz temporal regularizer (RPCA-TV) (Wang *et al.*, 2018). Given the unavailability of anomaly labels in public transit, we define anomalies as deviations from the expected passenger flow, excluding periodic fluctuations. Specifically, an anomaly at station n at time t on day k occurs when

$$M_{n,t}^{k} > \bar{M}_{n,t} + \xi\sigma_{n,t}, \quad \text{or} \quad M_{n,t}^{k} < \bar{M}_{n,t} - \xi\sigma_{n,t}, \tag{6}$$

where $\bar{M}_{n,t} = \sum_{k=1}^{K} M_{n,t}^{k}/K$ is the average passenger boarding flow in station n at time t during K days, $\sigma_{n,t}$ is its standard derivation, and ξ is a parameter to control the standard derivation.

Figure 1 shows the results of anomaly detection from the sparse matrix $\boldsymbol{S}$ of two selected stations (JNX and XMK). Two conclusions can be drawn: (i) All three models can detect almost all anomalies except for the one highlighted in the green patch (23:45 on July 26th at XMK station). Detecting anomalies accurately becomes challenging when passenger flow is minimal, as in this case where only 4 passengers are present, making it inaccurate to rely on statistical measurements like Equation (6). (ii) The anomaly passenger flow obtained from the RPCA and RPCA-TV models fluctuates much more compared to the proposed method. This indicates that the proposed method can accurately detect anomalies with fewer false alarms. This advantage stems from incorporating the Hankel tensor structure to capture more correlations from higher dimensions, such as periodic information within the data.

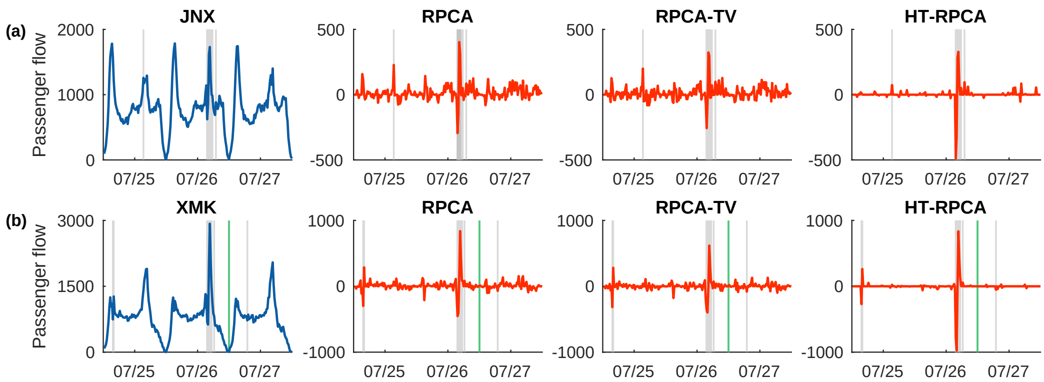

Figure 1 – *The anomaly detection results in metro passenger flow. Blue line: the raw passenger flow; red line: the anomaly passenger flow obtained from the models. The gray patches represent the anomaly occurrence defined in* (6) *when* $\xi = 2$.

We observe a similar anomaly pattern at stations JNX and XMK, beginning at 17:30 and ending at 18:30. Building on this observation, we proceed to analyze anomaly propagation through the metro network. Figure 2 illustrates the spatial distribution of anomalies from 17:00 to 19:00 on July 26th. Prior to anomalies occurring at JNX and XMK, the boarding passenger flow at several nearby stations experiences a slight increase at 17:15. Subsequently, the boarding flow at some stations, including JNX and XMK, abruptly drops (indicated by purple dots) at 17:30. This phenomenon rapidly spreads to more stations along the same metro line within 15 minutes. After the short interruption, the boarding flow of affected stations starts to increase until 18:30. Interestingly, the anomaly has delayed occurrence at a few stations far from JNX and XMK at 18:00 and 18:15. The duration of the entire anomaly spans 1 hour, consistent with the raw boarding flow data for JNX and XMK depicted by the blue line in Figure 1.

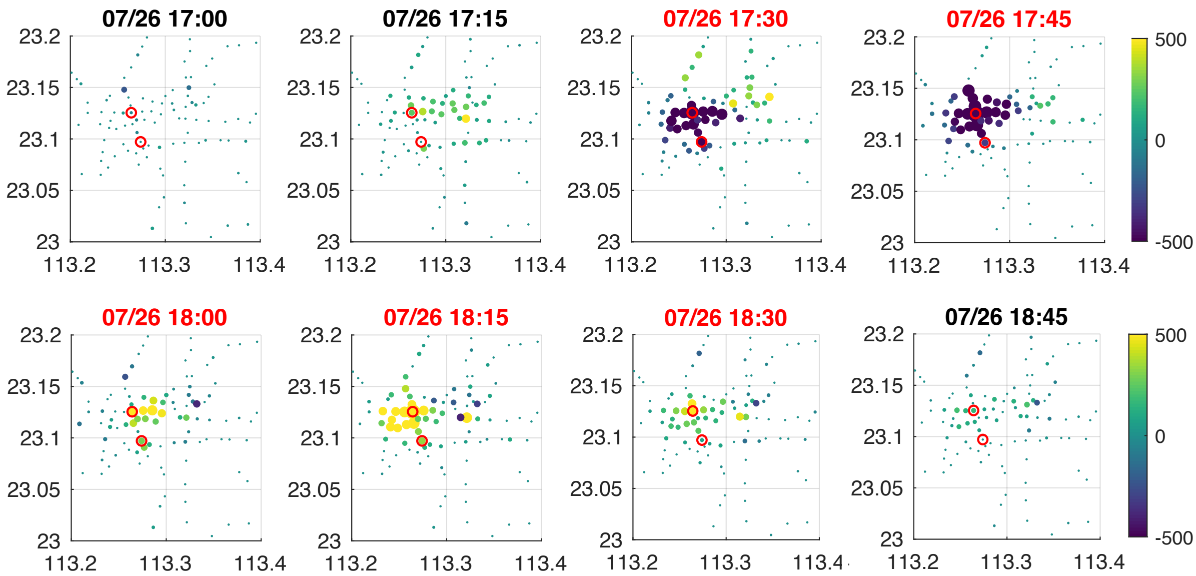

Figure 2 – *The spatial distribution of anomalies on July 26th. The red circles are JNX and XMK, and the red title represents anomaly timestamp.*

4 DISCUSSION

This study proposes an enhanced tensor version of RPCA model for anomaly detection in spatiotemporal traffic data. By incorporating temporal Hankel delay embedding, the model can capture additional dependencies, such as periodic information, thereby enhancing its robustness. There are some directions for future work. Firstly, we can incorporate the causality of anomaly into a forecasting model to improve prediction accuracy. Secondly, the computational cost of updating the tensor nuclear norm is significant due to SVD computations. Hence, we can apply faster SVD strategies or nonconvex methods to accelerate the algorithm. Thirdly, we can also leverage the spatial constraint, e.g., topology information, into the model to capture the spatial correlations. These enhancements will further strengthen the model's capabilities and extend its applicability to real-world traffic analysis scenarios.

References

Candès, Emmanuel J, Li, Xiaodong, Ma, Yi, & Wright, John. 2011. Robust principal component analysis? *Journal of the ACM (JACM)*, **58**(3), 1–37.

Chen, Xinyu, Lei, Mengying, Saunier, Nicolas, & Sun, Lijun. 2021. Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation. *IEEE Transactions on Intelligent Transportation Systems*, **23**(8), 12301–12310.

Jin, Kyong Hwan, & Ye, Jong Chul. 2017. Sparse and low-rank decomposition of a Hankel structured matrix for impulse noise removal. *IEEE Transactions on Image Processing*, **27**(3), 1448–1461.

Lu, Canyi, Feng, Jiashi, Chen, Yudong, Liu, Wei, Lin, Zhouchen, & Yan, Shuicheng. 2019. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE transactions on pattern analysis and machine intelligence*, **42**(4), 925–938.

Wang, Xudong, & Sun, Lijun. 2021. Diagnosing spatiotemporal traffic anomalies with low-rank tensor autoregression. *IEEE Transactions on Intelligent Transportation Systems*, **22**(12), 7904–7913.

Wang, Xudong, Wu, Yuankai, Zhuang, Dingyi, & Sun, Lijun. 2023. Low-rank Hankel tensor completion for traffic speed estimation. *IEEE Transactions on Intelligent Transportation Systems*, **24**(5), 4862–4871.

Wang, Xuehui, Zhang, Yong, Liu, Hao, Wang, Yang, Wang, Lichun, & Yin, Baocai. 2018. An improved robust principal component analysis model for anomalies detection of subway passenger flow. *Journal of advanced transportation*, **2018**.