POLYTECHNIQUE MONTRÉAL
UNIVERSITÉ D'INGÉNIERIE

IVADO

Matrix and Tensor Models for Spatiotemporal Traffic Data Imputation and Forecasting

Ph.D. Defense

Xinyu Chen

Polytechnique Montreal, Canada

December 11, 2023

President
Prof. Francesco Ciari
Polytechnique Montréal

Supervisor
Prof. Nicolas Saunier
Polytechnique Montréal

Co-supervisor
Prof. Lijun Sun
McGill University

Member
Prof. James Goulet
Polytechnique Montréal

External member
Prof. Guillaume Rabusseau
Université de Montréal

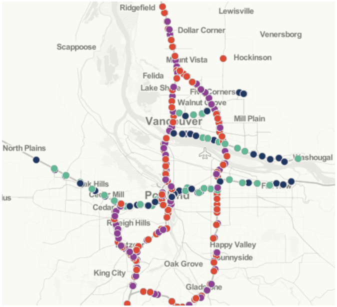
Outline

- **Background**
- **Literature Review**
- **Nonstationary Temporal Matrix Factorization**
- **Low-Rank Autoregressive Tensor Completion**
- **Laplacian Convolutional Representation**
  - Motivation
  - Reformulate Laplacian Regularization
  - Traffic Time Series Imputation
- **Hankel Tensor Factorization**
  - Motivation
  - Hankel Structure
- **Experiments**
- **Conclusion**

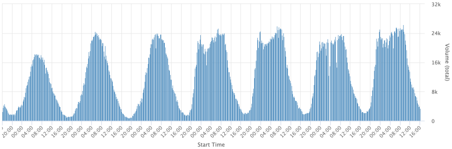
Multivariate Traffic Time Series

Many spatiotemporal traffic time series data are in the form of **matrix**.

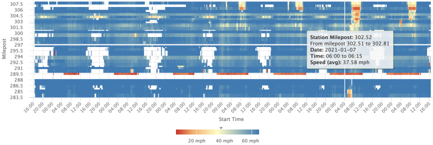
- **Example**: Portland highway traffic data[1].



Highway network & sensor locations



Traffic volume



Traffic speed

- $\boldsymbol{X} \in \mathbb{R}^{N \times T}$ with N spatial locations $\times$ T time steps
- Traffic volume/speed shows strong spatial/temporal dependencies

[1] https://portal.its.pdx.edu/home

Multiple Data Behaviors

Spatiotemporal traffic data are time series, but they involve multiple data behaviors.

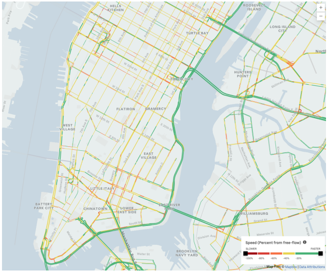
- Incompleteness & sparsity
- High-dimensionality
- Multidimensionality
- Noises & outliers
- Time-varying behavior
- Nonstationarity
- ......

In addition, spatiotemporal correlations are also very important.

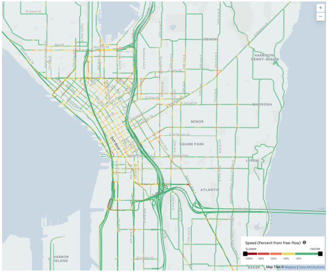
Multiple Data Behaviors

Sparsity & high-dimensionality

- Uber (hourly) movement speed data[2]



NYC movement



Seattle movement

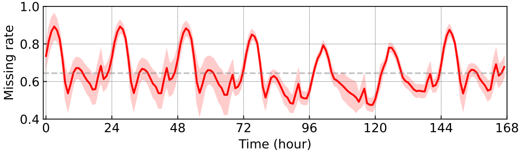
- The average speed on a given road segment for each hour of each day.
- Hourly speeds are computed when road segments have 5+ unique trips.
- **Issue**: insufficient sampling of ridesharing vehicles on the road network.

[2] https://movement.uber.com/

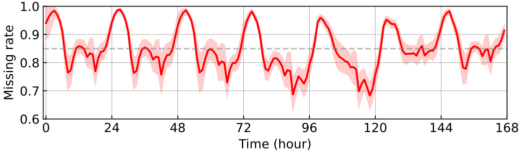
Multiple Data Behaviors

Sparsity & high-dimensionality

- **NYC** movement speed data (2019)
  - **98,210** road segments & 8,760 time steps (hours)
  - Overall missing rate: **64.43%**

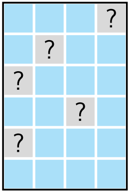


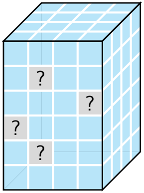
- **Seattle** movement speed data (2019)
  - **63,490** road segments & 8,760 time steps (hours)
  - Overall missing rate: **84.95%**



Problem Formulation

- **Objective A:** Given a multivariate time series data like $\boldsymbol{Y} \in \mathbb{R}^{N \times T}$ or a multidimensional time series data like $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{M \times N \times T}$ with the observed index set Ω, impute the missing values of the data.



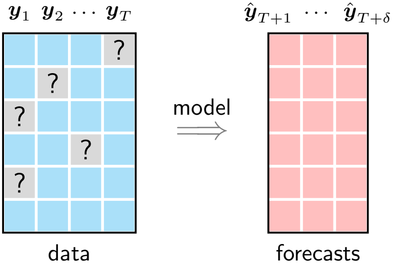


[Q]

- How to reconstruct missing values from observed data?
  - Matrix completion: From $\mathcal{P}_{\Omega}(\boldsymbol{Y})$ (observed) to $\mathcal{P}_{\Omega}^{\perp}(\boldsymbol{Y})$ (unobserved)
  - Tensor completion: From $\mathcal{P}_{\Omega}(\boldsymbol{\mathcal{Y}})$ (observed) to $\mathcal{P}_{\Omega}^{\perp}(\boldsymbol{\mathcal{Y}})$ (unobserved)
- How to make use of spatiotemporal correlations?
- How to make use of traffic time series dynamics?

Problem Formulation

- **Objective B:** Given a partially observed data $\boldsymbol{Y} \in \mathbb{R}^{N \times T}$ consisting of time series $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T \in \mathbb{R}^N$, forecast data points $\boldsymbol{y}_{T+\delta}, \delta \in \mathbb{N}^+$.

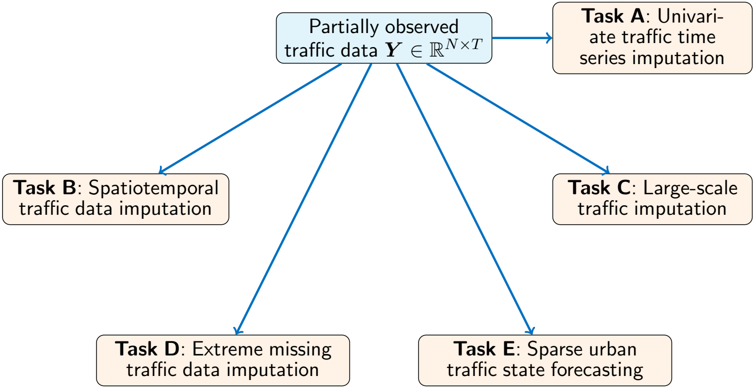


[Q]

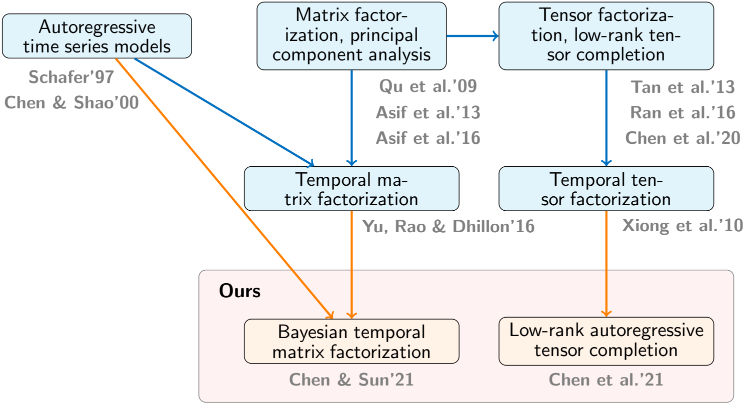
- How to learn from *high-dimensional* and *sparse* data?
- How to model *nonstationarity* in time series?
- How to perform forecasting on these time series?

Whole Picture

We are working on **spatiotemporal traffic data modeling**.



Spatiotemporal Traffic Data Imputation



Spatiotemporal Traffic Forecasting

Large-scale traffic data, e.g., PeMS traffic speed data and Uber movement speed data

High-dimensional forecasting

- Transfer learning — Mallick et al.'21
- Temporal matrix factorization — Chen & Sun'21
- High-dimensional vector autoregressive model — Wang et al.'21

Multidimensional forecasting

- Matrix autoregressive model — Chen, Xiao & Yang'21; Hsu, Huang & Tsay'21
- Autoregressive tensor factorization — Takeuchi, Kashima & Ueda'17

Forecasting on incomplete data

- AR predictor — Anava, Hazan & Zeevi'15
- Temporal matrix factorization — Yu, Rao & Dhillon'16; Chen & Sun'21
- Temporal tensor factorization — Xiong et al.'10; Chen & Sun'21

Tensor Factorization

- Revisit tensor factorization

1927 — Higher-Order SVD

Frank Lauren Hitchcock

1960s — Tucker Decomposition

Ledyard R. Tucker

1970 — CP Decomposition

J. Douglas Carroll
Jih-Jie Chang
Richard A. Harshman

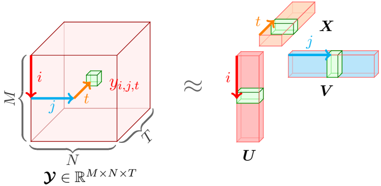
2009 — Tensor Decompositions and Applications

Tamara G. Kolda

2011 — Tensor-Train Decomposition

Ivan Oseledets

- **CP tensor factorization**: Factorize $\boldsymbol{\mathcal{Y}}$ into the combination of three rank-R factor matrices (i.e., low-dimensional latent factors).



$$\begin{cases} y_{i,j,t} \approx \sum_{r=1}^{R} u_{i,r} v_{j,r} x_{t,r} \\ \boldsymbol{\mathcal{Y}} \approx \sum_{r=1}^{R} \boldsymbol{u}_r \otimes \boldsymbol{v}_r \otimes \boldsymbol{x}_r \end{cases}$$

Matrix/Tensor Completion

Matrix nuclear norm ($\|X\|_*$) minimization
Candès & Recht'09

Singular value thresholding
Cai et al.'10

Truncated nuclear norm ($\|X\|_{r,*}, r \in \mathbb{N}^+$) minimization
Zhang et al.'12
Hu et al.'12

Tensor nuclear norm ($\|\boldsymbol{\mathcal{X}}\|_*$) minimization
Liu et al.'13

Generalized nonconvex nonsmooth low-rank minimization
Lu et al.'14

Tensor nuclear norm minimization with linear transform
Lu et al.'19

Low-rank Hankel matrix/tensor ($\mathcal{H}_\tau(\cdot)$) completion
Yokota et al.'18
Cai et al.'21

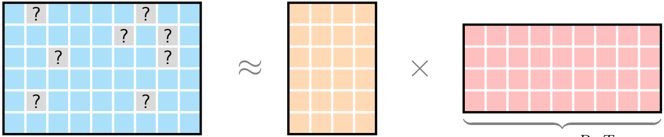
Circulant/convolutional nuclear norm ($\mathcal{C}(\cdot), \mathcal{C}_\tau(\cdot)$) minimization
Liu'22
Liu & Zhang'23

This research

- Integrate temporal modeling techniques (e.g., temporal smoothing and time series autoregression) into low-rank matrix and tensor methods
- Implement spatiotemporal traffic data imputation and forecasting on partially observed data

Matrix Factorization

A simple approach to reconstruct missing values.



$$\underbrace{\boldsymbol{Y}}_{\boldsymbol{Y}\in\mathbb{R}^{N\times T}} \approx \underbrace{\boldsymbol{W}^\top}_{\boldsymbol{W}^\top\in\mathbb{R}^{N\times R}} \times \underbrace{\boldsymbol{X}}_{\boldsymbol{X}\in\mathbb{R}^{R\times T}}$$

MF (Koren et al.'09)

Estimating low-dimensional $\boldsymbol{W},\boldsymbol{X}$:

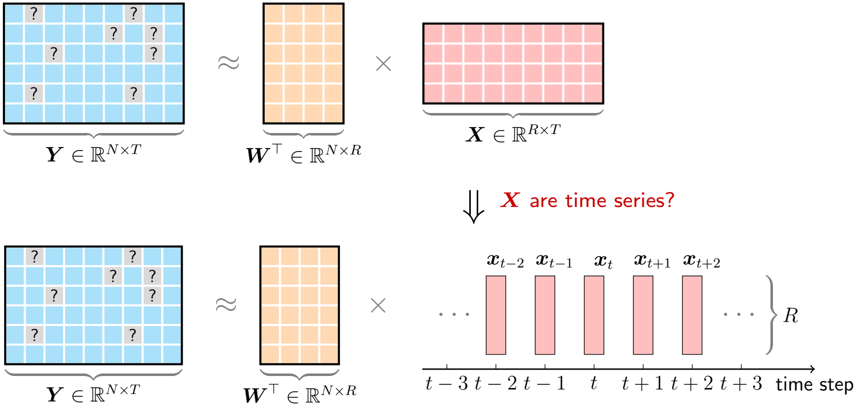
$$\min_{\boldsymbol{W},\boldsymbol{X}}\ \frac{1}{2}\|\mathcal{P}_\Omega(\boldsymbol{Y}-\boldsymbol{W}^\top\boldsymbol{X})\|_F^2$$

on data $\boldsymbol{Y}$ w/ observed index set Ω.

- ✓ Learn from sparse data
- ✗ Temporal correlations
- ✗ Time series forecasting

Temporal Matrix Factorization

Vector autoregression (VAR) on the temporal factor matrix.



Why? $\boldsymbol{X} \in \mathbb{R}^{R \times T}$ is the low-dimensional representation of time series dynamics of $\boldsymbol{Y} \in \mathbb{R}^{N \times T}$.

Temporal Matrix Factorization

MF (Koren et al.'09)

Estimating low-dimensional $\boldsymbol{W}, \boldsymbol{X}$:

$$\min_{\boldsymbol{W},\boldsymbol{X}} \frac{1}{2}\|\mathcal{P}_{\Omega}(\boldsymbol{Y}-\boldsymbol{W}^{\top}\boldsymbol{X})\|_F^2$$

on data $\boldsymbol{Y}$ w/ observed index set Ω.

$+$

dth-order VAR

$$\boldsymbol{x}_t=\sum_{k=1}^{d}\boldsymbol{A}_k\boldsymbol{x}_{t-k}+\boldsymbol{\epsilon}_t$$

w/ coefficients $\{\boldsymbol{A}_k\}$.

$\Downarrow$ Yu et al.'16, Chen & Sun'21

$$\min_{\boldsymbol{W},\boldsymbol{X},\{\boldsymbol{A}_k\}_{k=1}^{d}} \frac{1}{2}\underbrace{\|\mathcal{P}_{\Omega}(\boldsymbol{Y}-\boldsymbol{W}^{\top}\boldsymbol{X})\|_F^2}_{\text{MF on data } \boldsymbol{Y}}+\frac{\gamma}{2}\underbrace{\sum_{t=d+1}^{T}\left\|\boldsymbol{x}_t-\sum_{k=1}^{d}\boldsymbol{A}_k\boldsymbol{x}_{t-k}\right\|_2^2}_{\text{VAR on temporal factors } \boldsymbol{X}}$$

Nonstationary Temporal Matrix Factorization

Nonstationary temporal matrix factorization (NoTMF)

Given any partially observed time series data $\boldsymbol{Y}\in\mathbb{R}^{N\times T}$ with observed index set Ω, then we assume a season-m differencing on the latent temporal factors:

$$\min_{\boldsymbol{W},\boldsymbol{X},\{\boldsymbol{A}_k\}_{k=1}^{d}} \frac{1}{2}\|\mathcal{P}_{\Omega}(\boldsymbol{Y}-\boldsymbol{W}^{\top}\boldsymbol{X})\|_F^2+\frac{\rho}{2}(\|\boldsymbol{W}\|_F^2+\|\boldsymbol{X}\|_F^2) + \frac{\gamma}{2}\sum_{t=d+m+1}^{T}\left\|(\boldsymbol{x}_t-\boldsymbol{x}_{t-m})-\sum_{k=1}^{d}\boldsymbol{A}_k(\boldsymbol{x}_{t-k}-\boldsymbol{x}_{t-m-k})\right\|_2^2$$

- First-order differencing $\boldsymbol{x}_t'=\boldsymbol{x}_t-\boldsymbol{x}_{t-1}$.
- Second-order differencing $\boldsymbol{x}_t''=(\boldsymbol{x}_t-\boldsymbol{x}_{t-1})-(\boldsymbol{x}_{t-1}-\boldsymbol{x}_{t-2})$.
- Twice-differenced series $\boldsymbol{x}_t''=(\boldsymbol{x}_t-\boldsymbol{x}_{t-m})-(\boldsymbol{x}_{t-1}-\boldsymbol{x}_{t-m-1})$.

✓ Stationarizing a time series with differencing can improve the prediction.[3]

[3] Stationarity and differencing: https://otexts.com/fpp2/stationarity.html

Nonstationary Temporal Matrix Factorization

Rewrite VAR in the form of matrix

Temporal operators

For any multivariate time series $\boldsymbol{X}\in\mathbb{R}^{R\times T}$ with $m,d\in\mathbb{N}^{+}$, if we define temporal operators as

$$\begin{aligned}\boldsymbol{\Psi}_k \triangleq & \begin{bmatrix}\boldsymbol{0}_{(T-d-m)\times(d-k)} & -\boldsymbol{I}_{T-d-m} & \boldsymbol{0}_{(T-d-m)\times(k+m)}\end{bmatrix} \\ & +\begin{bmatrix}\boldsymbol{0}_{(T-d-m)\times(d+m-k)} & \boldsymbol{I}_{T-d-m} & \boldsymbol{0}_{(T-d-m)\times k}\end{bmatrix} \\ & \in\mathbb{R}^{(T-d-m)\times T},\, k=0,1,\ldots,d\end{aligned}$$

then

$$\sum_{t=d+m+1}^{T}\|(\boldsymbol{x}_t-\boldsymbol{x}_{t-m})-\sum_{k=1}^{d}\boldsymbol{A}_k(\boldsymbol{x}_{t-k}-\boldsymbol{x}_{t-m-k})\|_2^2$$

$$\equiv\|\boldsymbol{X}\boldsymbol{\Psi}_0^\top-\sum_{k=1}^{d}\boldsymbol{A}_k\boldsymbol{X}\boldsymbol{\Psi}_k^\top\|_F^2\triangleq\|\boldsymbol{X}\boldsymbol{\Psi}_0^\top-\boldsymbol{A}(\boldsymbol{I}_d\otimes\boldsymbol{X})\boldsymbol{\Psi}^\top\|_F^2$$

where $\boldsymbol{A}\triangleq\begin{bmatrix}\boldsymbol{A}_1 & \cdots & \boldsymbol{A}_d\end{bmatrix}$ and $\boldsymbol{\Psi}\triangleq\begin{bmatrix}\boldsymbol{\Psi}_1 & \cdots & \boldsymbol{\Psi}_d\end{bmatrix}$.

Nonstationary Temporal Matrix Factorization

Rewrite NoTMF:

$$\begin{aligned}\min_{\boldsymbol{W},\boldsymbol{X},\boldsymbol{A}}\ & \frac{1}{2}\|\mathcal{P}_{\Omega}(\boldsymbol{Y}-\boldsymbol{W}^\top\boldsymbol{X})\|_F^2+\frac{\rho}{2}(\|\boldsymbol{W}\|_F^2+\|\boldsymbol{X}\|_F^2) \\ & +\frac{\gamma}{2}\|\boldsymbol{X}\boldsymbol{\Psi}_0^\top-\boldsymbol{A}(\boldsymbol{I}_d\otimes\boldsymbol{X})\boldsymbol{\Psi}^\top\|_F^2\end{aligned}$$

Alternating minimization method:

- w.r.t. $\boldsymbol{W}$:

$$\frac{\partial f}{\partial\boldsymbol{W}}=-\boldsymbol{X}\mathcal{P}_{\Omega}^\top(\boldsymbol{Y}-\boldsymbol{W}^\top\boldsymbol{X})+\rho\boldsymbol{W}=\boldsymbol{0}\qquad\text{(Least squares)}$$

- w.r.t. $\boldsymbol{X}$:

$$\frac{\partial f}{\partial\boldsymbol{X}}=-\boldsymbol{W}\mathcal{P}_{\Omega}(\boldsymbol{Y}-\boldsymbol{W}^\top\boldsymbol{X})+\rho\boldsymbol{X}+\gamma\sum_{k=0}^{d}\boldsymbol{A}_k^\top\left(\sum_{h=0}^{d}\boldsymbol{A}_h\boldsymbol{X}\boldsymbol{\Psi}_h^\top\right)\boldsymbol{\Psi}_k=\boldsymbol{0}$$

  This generalized Sylvester equation can be solved by **conjugate gradient**.

- w.r.t. $\boldsymbol{A}$:

$$\boldsymbol{A}=\boldsymbol{X}\boldsymbol{\Psi}_0^\top[(\boldsymbol{I}_d\otimes\boldsymbol{X})\boldsymbol{\Psi}^\top]^\dagger\qquad\text{(Least squares)}$$

Nonstationary Temporal Matrix Factorization

NoTMF forecasting on streaming data?

- NoTMF: Use $\boldsymbol{Y}_t$ to estimate $\boldsymbol{W}, \boldsymbol{X}, \boldsymbol{A}$.

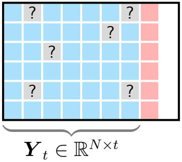
Implementation

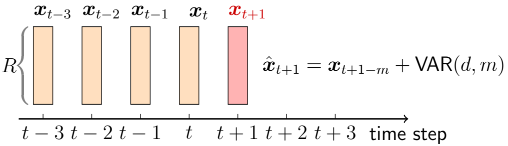
- Estimate $\boldsymbol{W}, \boldsymbol{X}, \boldsymbol{A}$
- Forecast $\hat{\boldsymbol{x}}_{t+1}$ with VAR
- Return $\hat{\boldsymbol{y}}_{t+1} = \boldsymbol{W}^\top \hat{\boldsymbol{x}}_{t+1}$

✓ Sparse input $\boldsymbol{Y}_t$

✓ Low-dimensional temporal factors

✓ Forecast in latent spaces





Nonstationary Temporal Matrix Factorization

NoTMF forecasting on streaming data?

- Online forecasting (Gultekin & Paisley'18): Fix $\boldsymbol{W}$ and use $\boldsymbol{Y}_{t+1}$ to update $\boldsymbol{X}, \boldsymbol{A}$.

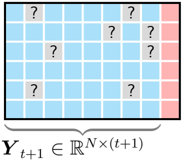
Implementation

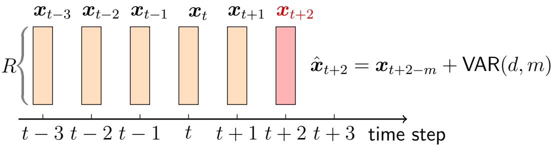
- Estimate $\boldsymbol{X}, \boldsymbol{A}$
- Forecast $\hat{\boldsymbol{x}}_{t+2}$ with VAR
- Return $\hat{\boldsymbol{y}}_{t+2} = \boldsymbol{W}^\top \hat{\boldsymbol{x}}_{t+2}$

✓ Sparse input $\boldsymbol{Y}_{t+1}$

✓ Fixed spatial factors $\boldsymbol{W}$

✓ Forecast in latent spaces





Matrix/Tensor Completion

Cornerstone: Nuclear norm minimization in matrix/tensor completion

LRMC (Candès & Recht'09)

Estimating the matrix $\boldsymbol{X}$:

$$\min_{\boldsymbol{X}} \ \|\boldsymbol{X}\|_*$$
$$\text{s.t. } \mathcal{P}_\Omega(\boldsymbol{X}) = \mathcal{P}_\Omega(\boldsymbol{Y})$$

on data $\boldsymbol{Y}$ w/ observed index set Ω.

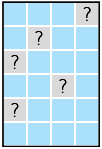
vs.

LRTC (Liu et al.'13)

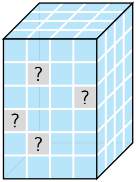
Estimating the tensor $\boldsymbol{\mathcal{X}}$:

$$\min_{\boldsymbol{\mathcal{X}}} \ \|\boldsymbol{\mathcal{X}}\|_*$$
$$\text{s.t. } \mathcal{P}_\Omega(\boldsymbol{\mathcal{X}}) = \mathcal{P}_\Omega(\boldsymbol{\mathcal{Y}})$$

on data $\boldsymbol{\mathcal{Y}}$ w/ observed index set Ω.



$\mathcal{P}_\Omega(\boldsymbol{Y}) \in \mathbb{R}^{N\times T}$

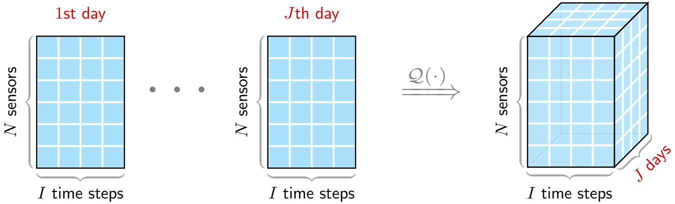


$\mathcal{P}_\Omega(\boldsymbol{\mathcal{Y}}) \in \mathbb{R}^{N\times I\times J}$

- **Limitation**: Only cover global consistency

Low-Rank Autoregressive Tensor Completion

- Introduce traffic tensors with day dimension[4] (Tan et al.'13, Chen et al.'19, ...)



[4] There are $T = IJ$ time steps in total.

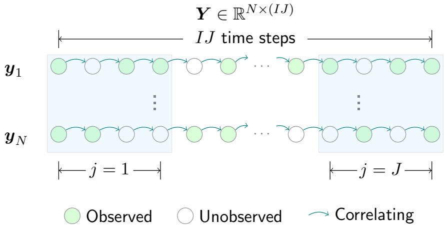
Low-Rank Autoregressive Tensor Completion

- Build temporal correlations with autoregression

On the time series $\boldsymbol{Y} \in \mathbb{R}^{N \times T}$:

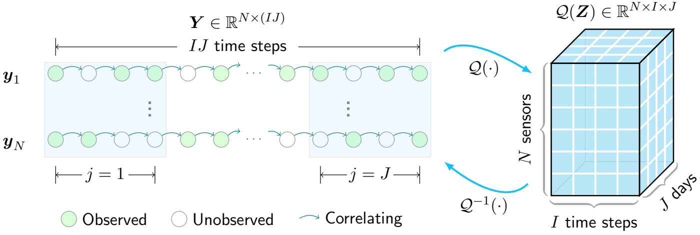
$$\boxed{\|\boldsymbol{Y}\|_{\boldsymbol{A},\mathcal{H}} \triangleq \sum_{n,t}\Big(y_{n,t} - \sum_{k} a_{n,k} y_{n,t-h_k}\Big)^2}$$

with the time lag set $\mathcal{H} = \{h_1, \ldots, h_d\}$ and the coefficient matrix $\boldsymbol{A} \in \mathbb{R}^{N \times d}$.



Low-Rank Autoregressive Tensor Completion

Local consistency w/ autoregression **Global consistency** w/ tensor structure



LATC

Optimization problem:

$$\min_{\boldsymbol{Z},\boldsymbol{A}} \ \|\mathcal{Q}(\boldsymbol{Z})\|_{r,*} + \frac{\gamma}{2}\|\boldsymbol{Z}\|_{\boldsymbol{A},\mathcal{H}}$$

$$\text{s.t. } \mathcal{P}_\Omega(\boldsymbol{Z}) = \mathcal{P}_\Omega(\boldsymbol{Y})$$

on data $\boldsymbol{Y}$ w/ observed index set Ω.

$\Rightarrow$

Two subproblems

$$\begin{cases} \boldsymbol{Z} := \underset{\mathcal{P}_\Omega(\boldsymbol{Z})=\mathcal{P}_\Omega(\boldsymbol{Y})}{\arg\min} \ \|\mathcal{Q}(\boldsymbol{Z})\|_{r,*} + \frac{\gamma}{2}\|\boldsymbol{Z}\|_{\boldsymbol{A},\mathcal{H}} \\ \boldsymbol{A} := \frac{1}{2}\|\boldsymbol{Z}\|_{\boldsymbol{A},\mathcal{H}} \qquad \text{(Least squares)} \end{cases}$$

Low-Rank Autoregressive Tensor Completion

$\boldsymbol{Z}$-subproblem:

$$\boldsymbol{Z} := \underset{\mathcal{P}_{\Omega}(\boldsymbol{Z})=\mathcal{P}_{\Omega}(\boldsymbol{Y})}{\arg\min} \quad \|\mathcal{Q}(\boldsymbol{Z})\|_{r,*} + \frac{\gamma}{2}\|\boldsymbol{Z}\|_{\boldsymbol{A},\mathcal{H}}$$

- Augmented Lagrangian function:[5]

$$\mathcal{L}(\boldsymbol{\mathcal{X}},\boldsymbol{Z},\boldsymbol{\mathcal{W}}) = \|\boldsymbol{\mathcal{X}}\|_{r,*}+\frac{\gamma}{2}\|\boldsymbol{Z}\|_{\boldsymbol{A},\mathcal{H}}+\frac{\lambda}{2}\|\boldsymbol{\mathcal{X}}-\mathcal{Q}(\boldsymbol{Z})\|_F^2+\langle\boldsymbol{\mathcal{W}},\boldsymbol{\mathcal{X}}-\mathcal{Q}(\boldsymbol{Z})\rangle+\pi(\boldsymbol{Z})$$

Implementation

Repeat

- Compute $\boldsymbol{Z}$
- Compute $\boldsymbol{A}$

$\Longrightarrow$

Implementation

Repeat

- Repeat
  - Compute $\boldsymbol{\mathcal{X}}$
  - Compute $\boldsymbol{Z}$
  - Compute $\boldsymbol{\mathcal{W}}$
- Compute $\boldsymbol{A}$

[5] $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{N\times I\times J}$ (Lagrange multiplier); The indicator function:

$$\pi(\boldsymbol{Z}) = \begin{cases} 0, & \text{if } \mathcal{P}_{\Omega}(\boldsymbol{Z}) = \mathcal{P}_{\Omega}(\boldsymbol{Y}), \\ +\infty, & \text{otherwise.} \end{cases}$$

Low-Rank Autoregressive Tensor Completion

- Augmented Lagrangian function:

$$\mathcal{L}(\boldsymbol{\mathcal{X}},\boldsymbol{Z},\boldsymbol{\mathcal{W}})=\|\boldsymbol{\mathcal{X}}\|_{r,*}+\frac{\gamma}{2}\|\boldsymbol{Z}\|_{\boldsymbol{A},\mathcal{H}}+\frac{\lambda}{2}\|\boldsymbol{\mathcal{X}}-\mathcal{Q}(\boldsymbol{Z})\|_F^2+\langle\boldsymbol{\mathcal{W}},\boldsymbol{\mathcal{X}}-\mathcal{Q}(\boldsymbol{Z})\rangle+\pi(\boldsymbol{Z})$$

- The ADMM[6] scheme:

$$\begin{cases}\boldsymbol{\mathcal{X}}:=\arg\min\limits_{\boldsymbol{\mathcal{X}}}\ \mathcal{L}(\boldsymbol{\mathcal{X}},\boldsymbol{Z},\boldsymbol{\mathcal{W}}) & \text{(Truncated nuclear norm minimization)}\\ \boldsymbol{Z}:=\arg\min\limits_{\boldsymbol{Z}}\ \mathcal{L}(\boldsymbol{\mathcal{X}},\boldsymbol{Z},\boldsymbol{\mathcal{W}}) & \text{(Generalized Sylvester equation)}\\ \boldsymbol{\mathcal{W}}:=\boldsymbol{\mathcal{W}}+\lambda(\boldsymbol{\mathcal{X}}-\mathcal{Q}(\boldsymbol{Z})) & \text{(Standard update)}\end{cases}$$

✓ Solution to $\boldsymbol{\mathcal{X}}$: singular value thresholding

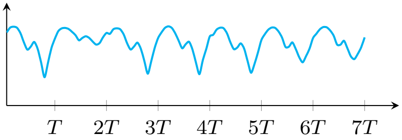
✓ Solution to $\boldsymbol{Z}$: conjugate gradient

[6] Alternating Direction Method of Multipliers.

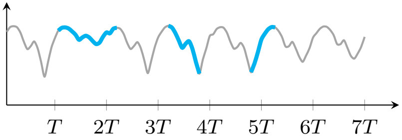
Laplacian Convolutional Representation

Motivation: Time series imputation

- Global trends (e.g., long-term quasi-seasonality & daily/weekly rhythm):



- Local trends (e.g., short-term time series trends):

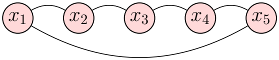


- **[Question]** How to characterize both global and local trends in sparse time series data?

Laplacian Convolutional Representation

Local trend modeling: Reformulate temporal regularization with circular convolution.

- Intuition of (circulant) Laplacian matrix.



Undirected and circulant graph

$\xrightarrow{\text{Modeling}}$

$$\boldsymbol{L}=\begin{bmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ -1 & 0 & 0 & -1 & 2 \end{bmatrix}$$

(Circulant) Laplacian matrix

- Define Laplacian kernel:

$$\boldsymbol{\ell} \triangleq (2,-1,0,0,-1)^\top$$

$$\Downarrow$$

$$\boldsymbol{\ell} \triangleq (\underbrace{2\tau}_{\text{degree}}, \underbrace{-1,\cdots,-1}_{\tau}, 0, \cdots, 0, \underbrace{-1,\cdots,-1}_{\tau})^\top \in \mathbb{R}^T$$

for any time series $\boldsymbol{x}=(x_1,\ldots,x_T)^\top \in \mathbb{R}^T$.

- (Laplacian) Temporal regularization:

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

Laplacian Convolutional Representation

Global trend modeling
Circulant matrix definition
Literature review of circulant/convolution matrix

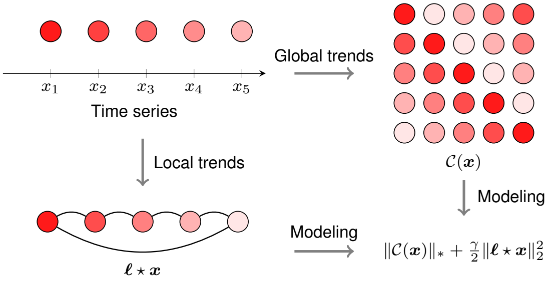
Laplacian Convolutional Representation

Laplacian Convolutional Representation (LCR)

For any partially observed time series $\boldsymbol{y} \in \mathbb{R}^T$ with observed index set Ω, LCR utilizes circulant matrix and Laplacian kernel to characterize **global and local trends** in time series, respectively, i.e.,

$$\min_{\boldsymbol{x}} \; \|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\gamma}{2}\|\boldsymbol{\ell} \star \boldsymbol{x}\|_2^2$$
$$\text{s.t. } \|\mathcal{P}_\Omega(\boldsymbol{x} - \boldsymbol{y})\|_2 \leq \epsilon$$

where $\mathcal{C} : \mathbb{R}^T \to \mathbb{R}^{T \times T}$ denotes the circulant operator.



Laplacian Convolutional Representation

- Augmented Lagrangian function:

$$\mathcal{L}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{w}) = \|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\gamma}{2}\|\boldsymbol{\ell}\star\boldsymbol{x}\|_2^2 + \frac{\lambda}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 + \langle\boldsymbol{w},\boldsymbol{x}-\boldsymbol{z}\rangle + \frac{\eta}{2}\|\mathcal{P}_\Omega(\boldsymbol{z}-\boldsymbol{y})\|_2^2$$

  where $\boldsymbol{w}\in\mathbb{R}^T$ is the Lagrange multiplier, and $\langle\cdot,\cdot\rangle$ denotes the inner product.
- The ADMM scheme:

$$\begin{cases} \boldsymbol{x} := \arg\min\limits_{\boldsymbol{x}}\ \mathcal{L}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{w}) & \text{(Nuclear norm minimization)} \\ \boldsymbol{z} := \arg\min\limits_{\boldsymbol{z}}\ \mathcal{L}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{w}) & \text{(Closed-form solution)} \\ \quad = \frac{1}{\lambda+\eta}\mathcal{P}_\Omega(\lambda\boldsymbol{x}+\boldsymbol{w}+\eta\boldsymbol{y}) + \frac{1}{\lambda}\mathcal{P}_\Omega^\perp(\lambda\boldsymbol{x}+\boldsymbol{w}) & \\ \boldsymbol{w} := \boldsymbol{w}+\lambda(\boldsymbol{x}-\boldsymbol{z}) & \text{(Standard update)} \end{cases}$$

- Optimize $\boldsymbol{x}$?

$$\|\mathcal{C}(\boldsymbol{x})\|_* = \|\mathcal{F}(\boldsymbol{x})\|_1 \qquad \& \qquad \frac{1}{2}\|\boldsymbol{\ell}\star\boldsymbol{x}\|_2^2 = \frac{1}{2T}\|\mathcal{F}(\boldsymbol{\ell})\circ\mathcal{F}(\boldsymbol{x})\|_2^2$$

  Nuclear norm minimization $\Rightarrow$ ℓ_1**-norm minimization with FFT** (in $\mathcal{O}(T\log T)$ time).

Laplacian Convolutional Representation

- Optimize $\boldsymbol{x}$ via FFT (in $\mathcal{O}(T\log T)$ time):

$$\begin{aligned} \boldsymbol{x} &:= \arg\min_{\boldsymbol{x}}\ \|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\gamma}{2}\|\boldsymbol{\ell}\star\boldsymbol{x}\|_2^2 + \frac{\lambda}{2}\|\boldsymbol{x}-\boldsymbol{z}+\boldsymbol{w}/\lambda\|_2^2 \\ \Longrightarrow \hat{\boldsymbol{x}} &:= \arg\min_{\hat{\boldsymbol{x}}}\ \|\hat{\boldsymbol{x}}\|_1 + \frac{\gamma}{2T}\|\hat{\boldsymbol{\ell}}\circ\hat{\boldsymbol{x}}\|_2^2 + \frac{\lambda}{2T}\|\hat{\boldsymbol{x}}-\hat{\boldsymbol{z}}+\hat{\boldsymbol{w}}/\lambda\|_2^2 \end{aligned}$$

where we introduce $\{\hat{\boldsymbol{\ell}},\hat{\boldsymbol{x}},\hat{\boldsymbol{z}},\hat{\boldsymbol{w}}\}\triangleq\mathcal{F}\{\boldsymbol{\ell},\boldsymbol{x},\boldsymbol{z},\boldsymbol{w}\}$ (i.e., FFT).

ℓ_1-norm Minimization in Complex Space (Liu & Zhang'23)

For any optimization problem in the form of ℓ_1-norm minimization in complex space:

$$\min_{\hat{\boldsymbol{x}}}\ \|\hat{\boldsymbol{x}}\|_1 + \frac{\omega}{2}\|\hat{\boldsymbol{x}}-\hat{\boldsymbol{h}}\|_2^2$$

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

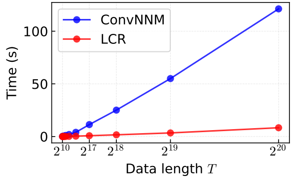
$$\hat{x}_t := \frac{\hat{h}_t}{|\hat{h}_t|}\cdot\max\{0,|\hat{h}_t|-1/\omega\}, t=1,\dots,T.$$

Laplacian Convolutional Representation

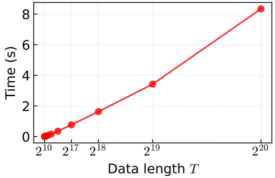
Empirical time complexity

On the synthetic data $\boldsymbol{y} \in \mathbb{R}^T$ with $T \in \{2^{10}, 2^{11}, \ldots, 2^{20}\}$

- Ours: **LCR**
  - An FFT implementation in $\mathcal{O}(T \log T)$
  - The logarithmic factor $\log T$ makes the FFT highly efficient
- Baseline: **ConvNNM**[7] (Liu'22, Liu & Zhang'23)
  - Convolution matrix $\mathcal{C}_{\tilde{\tau}}(\boldsymbol{y}) \in \mathbb{R}^{T \times \tilde{\tau}}$ with kernel size $\tilde{\tau} \in \mathbb{N}^+$
  - Singular value thresholding in $\mathcal{O}(\tilde{\tau}^2 T)$



ConvNNM vs. LCR



LCR

[7] Convolution nuclear norm minimization.

Two-Dimensional LCR (LCR-2D)

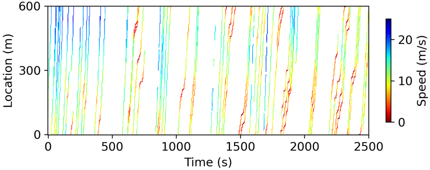
For any partially observed time series $\boldsymbol{Y} \in \mathbb{R}^{N \times T}$ with observed index set Ω, LCR can be formulated as follows,

$$\begin{aligned} \min_{\boldsymbol{X}} \; & \|\mathcal{C}(\boldsymbol{X})\|_* + \frac{\gamma}{2}\|(\boldsymbol{\ell}_s\boldsymbol{\ell}^\top) \star \boldsymbol{X}\|_F^2 \\ \text{s.t. } & \|\mathcal{P}_\Omega(\boldsymbol{X} - \boldsymbol{Y})\|_F \leq \epsilon \end{aligned}$$

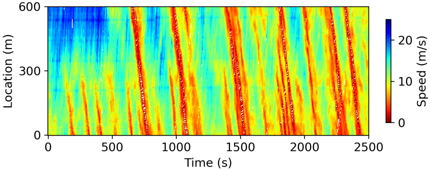
where $\mathcal{C} : \mathbb{R}^{N \times T} \to \mathbb{R}^{N \times N \times T \times T}$ denotes the circulant operator.

Motivation: Spatiotemporal data reconstruction

- Speed field reconstruction problem in vehicular traffic flow.



200-by-500 matrix (NGSIM) $\Downarrow$ Reconstruct speed field from 5% sparse trajectories?



- How to learn from sparse spatiotemporal data?
- How to characterize spatial/temporal dependencies?

Hankel Tensor Factorization

- Hankel matrix
  - Given $\boldsymbol{x}=(1,2,3,4,5)^\top$ and window length $\tau=2$, we have

$$\mathcal{H}_\tau(\boldsymbol{x})=\begin{bmatrix}1 & 2\\ 2 & 3\\ 3 & 4\\ 4 & 5\end{bmatrix}\in\mathbb{R}^{4\times 2}$$

Hankel matrix (Source: Twitter)

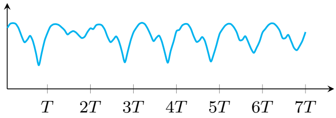
Hankel Tensor Factorization

- Hankel matrix
  - On time series $\boldsymbol{y}=(y_1,y_2,\ldots,y_5)^\top$ with $\tau=2$:

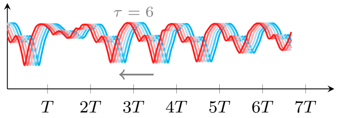
$$\mathcal{H}_\tau(\boldsymbol{y})=\begin{bmatrix} y_1 & y_2 \\ y_2 & y_3 \\ y_3 & y_4 \\ y_4 & y_5 \end{bmatrix}\approx\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}\otimes\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\Longrightarrow\quad \hat{\boldsymbol{y}}=\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \hat{y}_4 \\ \hat{y}_5 \end{bmatrix}=\mathcal{H}_\tau^{-1}\left(\begin{bmatrix} v_1x_1 & v_1x_2 \\ v_2x_1 & v_2x_2 \\ v_3x_1 & v_3x_2 \\ v_4x_1 & v_4x_2 \end{bmatrix}\right)=\begin{bmatrix} v_1x_1 \\ (v_1x_2+v_2x_1)/2 \\ (v_2x_2+v_3x_1)/2 \\ (v_3x_2+v_4x_1)/2 \\ v_4x_2 \end{bmatrix}$$

  - Automatic temporal modeling.



Traffic time series



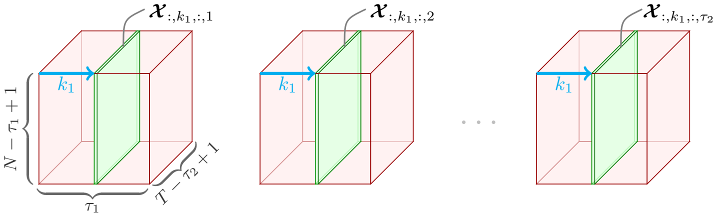
Hankel matrix

Hankel Tensor Factorization

- Hankel tensor: Given any matrix $\boldsymbol{X} \in \mathbb{R}^{N \times T}$, we have

$$\boldsymbol{\mathcal{X}} \triangleq \mathcal{H}_{\tau_1,\tau_2}(\boldsymbol{X})$$

  - Window lengths: $\tau_1, \tau_2 \in \mathbb{N}^{+}$;
  - Tensor size: $(N-\tau_1+1) \times \tau_1 \times (T-\tau_2+1) \times \tau_2$;



(Figure) 4th order Hankel tensor: A sequence of third-order tensors.

  - Slice: $\boldsymbol{\mathcal{X}}_{:,k_1,:,k_2}, \forall k_1, k_2$;
  - Slice size: $(N-\tau_1+1) \times (T-\tau_2+1)$.

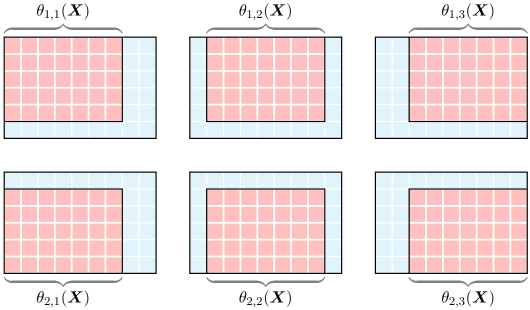
Hankel Tensor Factorization

Hankel indexing:

- Sampling function for the Hankelization:

$$\theta_{k_1,k_2}(\boldsymbol{X}) \triangleq [\mathcal{H}_{\tau_1,\tau_2}(\boldsymbol{X})]_{:,k_1,:,k_2},$$

  referring to the tensor slice with $k_1 \in \{1,\ldots,\tau_1\}$, $k_2 \in \{1,\ldots,\tau_2\}$.
- **[Importance]** Developing memory-efficient algorithms.



- Tensor slices $\theta_{k_1,k_2}(\boldsymbol{X})$ vs. data matrix $\boldsymbol{X}$

Hankel Tensor Factorization

Ours:

- Convolutional tensor decomposition (circular convolution $\star_{\text{row}}$):

$$\theta_{k_1,k_2}(\boldsymbol{Y}) \approx (\boldsymbol{Q} \star_{\text{row}} \boldsymbol{s}_{k_1}^\top)(\boldsymbol{U} \star_{\text{row}} \boldsymbol{v}_{k_2}^\top)^\top$$

Baselines:

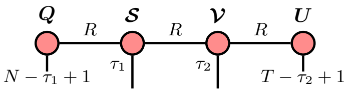
- CP tensor decomposition (Khatri-Rao product $\odot$):

$$\theta_{k_1,k_2}(\boldsymbol{Y}) \approx (\boldsymbol{Q} \odot \boldsymbol{s}_{k_1}^\top)(\boldsymbol{U} \odot \boldsymbol{v}_{k_2}^\top)^\top$$

- Tensor-train decomposition:

$$\theta_{k_1,k_2}(\boldsymbol{Y}) \approx (\boldsymbol{Q}\boldsymbol{S}_{k_1})(\boldsymbol{U}\boldsymbol{V}_{k_2})^\top$$

  - $\{\boldsymbol{S}_{k_1}, \boldsymbol{V}_{k_2}\}$ are **circulant matrices** $\Rightarrow$ convolutional decomposition
  - $\{\boldsymbol{S}_{k_1}, \boldsymbol{V}_{k_2}\}$ are **diagonal matrices** $\Rightarrow$ CP decomposition



Hankel Tensor Factorization

HTF (convolutional decomposition)

- Optimization problem:

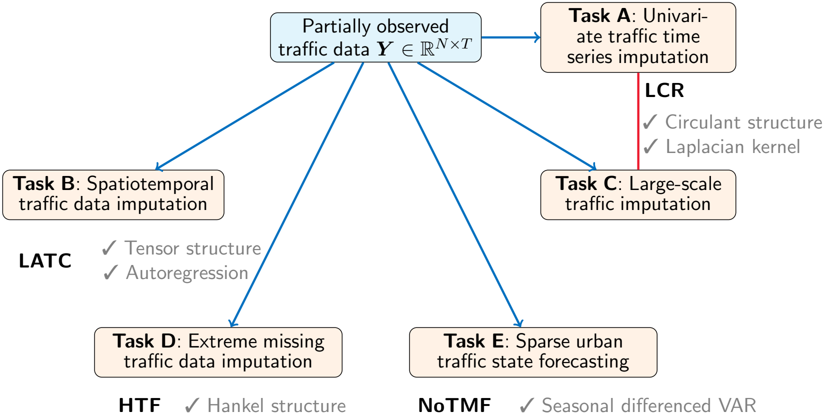
$$\min_{\boldsymbol{Q},\boldsymbol{S},\boldsymbol{U},\boldsymbol{V}} \ \frac{1}{2}\sum_{k_1,k_2}\left\|\mathcal{P}_{\Omega_{k_1,k_2}}\left(\theta_{k_1,k_2}(\boldsymbol{Y})-(\boldsymbol{Q}\star_{\text{row}}\boldsymbol{s}_{k_1})(\boldsymbol{U}\star_{\text{row}}\boldsymbol{v}_{k_2})^\top\right)\right\|_F^2 + \frac{\rho}{2}\left(\|\boldsymbol{Q}\|_F^2+\|\boldsymbol{S}\|_F^2+\|\boldsymbol{U}\|_F^2+\|\boldsymbol{V}\|_F^2\right)$$

- Alternating minimization:

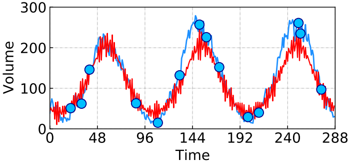
$$\begin{cases} \boldsymbol{Q}:=\{\boldsymbol{Q}\mid \frac{\partial f}{\partial \boldsymbol{Q}}=\boldsymbol{0}\} & \text{(conjugate gradient)} \\ \boldsymbol{s}_{k_1}:=\{\boldsymbol{s}_{k_1}\mid \frac{\partial f}{\partial \boldsymbol{s}_{k_1}}=\boldsymbol{0}\},\ \forall k_1 & \text{(conjugate gradient)} \\ \boldsymbol{U}:=\{\boldsymbol{U}\mid \frac{\partial f}{\partial \boldsymbol{U}}=\boldsymbol{0}\} & \text{(conjugate gradient)} \\ \boldsymbol{v}_{k_2}:=\{\boldsymbol{v}_{k_2}\mid \frac{\partial f}{\partial \boldsymbol{v}_{k_2}}=\boldsymbol{0}\},\ \forall k_2 & \text{(conjugate gradient)} \end{cases}$$

Whole Picture

We are working on **spatiotemporal traffic data modeling**.



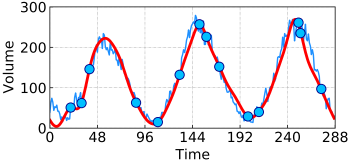
Univariate Traffic Time Series Imputation



CircNNM:

$$\min_{\boldsymbol{x}} \ \|\mathcal{C}(\boldsymbol{x})\|_*$$
$$\text{s. t.} \ \|\mathcal{P}_\Omega(\boldsymbol{x}-\boldsymbol{y})\|_2 \leq \epsilon$$

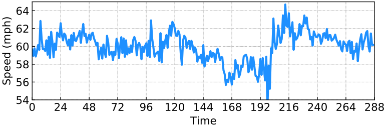
$\Downarrow$ Plus temporal regularization (TR)



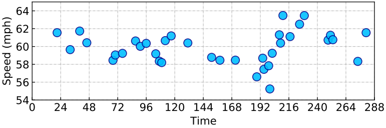
LCR:

$$\min_{\boldsymbol{x}} \ \|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\gamma}{2}\|\boldsymbol{\ell}\star\boldsymbol{x}\|_2^2$$
$$\text{s. t.} \ \|\mathcal{P}_\Omega(\boldsymbol{x}-\boldsymbol{y})\|_2 \leq \epsilon$$

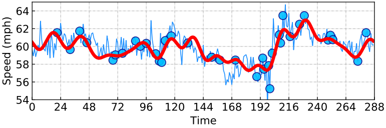
Univariate Traffic Time Series Imputation



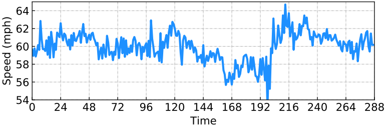
Mask **90%** observations



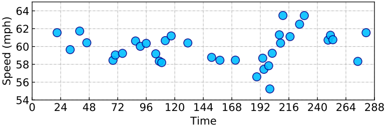
Reconstruct time series



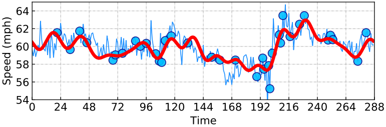
Univariate Traffic Time Series Imputation

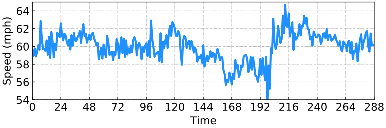


Mask **90%** observations

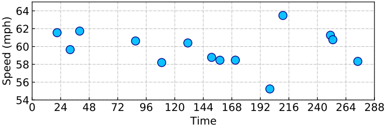


Reconstruct time series

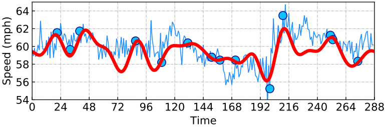




Mask **95%** observations



Reconstruct time series



Spatiotemporal Traffic Data Imputation

LATC vs. baseline models (in MAPE/RMSE)

- On the Seattle freeway traffic speed dataset ($\boldsymbol{Y} \in \mathbb{R}^{323 \times 8064}$)

Missing rate	LATC	LAMC	LRTC-TNN	BTMF	SPC
30%, Random Missing	**4.90**/**3.16**	5.98/3.73	4.99/3.20	5.91/3.72	5.92/3.62
70%, Random Missing	**5.96**/**3.71**	8.02/4.70	6.10/3.77	6.47/3.98	7.38/4.30
90%, Random Missing	**7.46**/**4.50**	10.56/5.91	8.08/4.80	8.17/4.81	9.75/5.31
30%, Nonrandom Missing	7.10/4.33	6.99/4.25	**6.85**/**4.21**	9.26/5.36	8.87/4.99
70%, Nonrandom Missing	9.40/5.40	9.75/5.60	**9.23**/**5.35**	10.47/6.15	11.32/5.92
30%, Block-out Missing	**9.43**/**5.36**	27.05/13.66	9.52/5.41	14.33/13.60	11.30/5.84

- On the Portland highway traffic volume dataset ($\boldsymbol{Y} \in \mathbb{R}^{1156 \times 2976}$)

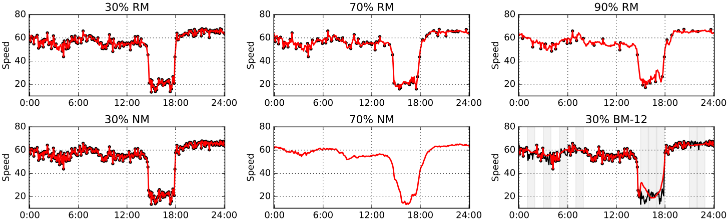
Missing rate	LATC	LAMC	LRTC-TNN	BTMF	SPC
30%, Random Missing	**16.95**/**15.99**	17.93/16.03	17.27/16.08	18.22/19.14	21.29/56.73
70%, Random Missing	**19.59**/**18.70**	21.26/19.37	19.99/18.73	19.96/22.21	24.35/43.32
90%, Random Missing	23.15/22.83	25.64/23.75	**22.90**/**22.68**	23.90/25.71	28.45/39.65
30%, Nonrandom Missing	**19.48**/19.14	19.93/19.69	19.59/**18.91**	19.55/20.38	26.96/60.33
70%, Nonrandom Missing	27.67/45.03	25.75/28.25	30.26/60.85	**23.86**/**26.74**	33.42/47.34
30%, Block-out Missing	**24.01**/**23.50**	29.21/27.60	31.74/74.42	27.85/25.68	31.01/60.33

- LATC vs. LAMC: The significance of tensor representation
- LATC vs. LRTC-TNN: The significance of temporal autoregression

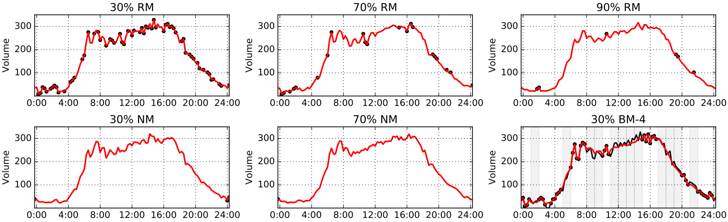
Spatiotemporal Traffic Data Imputation

LATC imputation

- Seattle freeway traffic speed data



- Portland highway traffic volume data



Large-Scale Traffic Data Imputation

LCR vs. baseline models (in MAPE/RMSE)

- PeMS-4W: California freeway traffic speed dataset ($\boldsymbol{Y} \in \mathbb{R}^{11160 \times 8064}$)

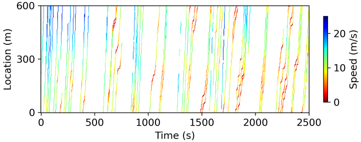
Model	Missing rate			
	30%	50%	70%	90%
LCR-2D	1.50/**1.49**	1.76/**1.69**	**2.07**/**2.06**	**3.19**/**3.05**
LCR$_N$	**1.48**/1.50	**1.73**/1.73	**2.07**/2.12	3.24/3.22
LCR	1.50/**1.49**	1.76/**1.69**	2.08/2.07	3.21/3.06
CTNNM	2.26/1.84	2.67/2.14	3.40/2.66	5.22/3.90
CircNNM	2.26/1.84	2.69/2.15	3.43/2.67	5.34/3.96
LRMC	2.04/1.80	2.43/2.12	3.08/2.66	6.05/4.43
HaLRTC	1.98/1.73	2.22/1.98	2.84/2.49	4.39/3.66
LRTC-TNN	1.68/1.55	1.93/1.77	2.33/2.14	3.40/3.10
NoTMF	2.95/2.65	3.05/2.73	3.33/2.97	5.22/4.71

Results

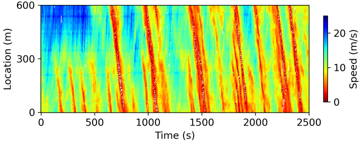
- LCR-2D $>$ CTNNM: The importance of temporal regularization
- CTNNM $\geq$ CircNNM: Ciculant tensor is superior to circulant matrix
- LCR $>$ LRMC/LRTC: The importance of global/local modeling

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

Extreme Missing Traffic Data Imputation

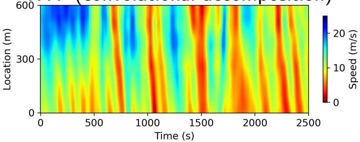


Sparse speed field



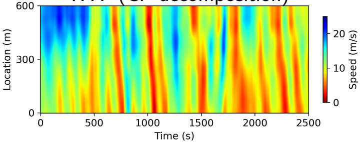
Ground truth speed field

HTF (convolutional decomposition)



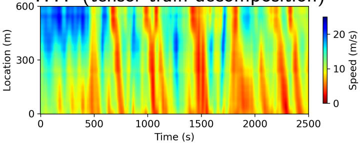
MAPE = **51.92%**

HTF (CP decomposition)



MAPE = 53.93%

HTF (tensor-train decomposition)



MAPE = 56.48%

Sparse Urban Traffic State Forecasting

NoTMF

Conclusion

- Data (large-scale, high-dimensional, city-wide, sparse)
- Modeling (meaningfulness and importance of temporal correlations)

Conclusion

Low-Rank Autoregressive Tensor Completion (LATC):

- (Highlight) Global & local consistency
  - ✓ Tensor structure $\|\mathcal{Q}(\boldsymbol{\mathcal{Z}})\|_{r,*}$ ✓ Autoregression $\|\boldsymbol{Z}\|_{\boldsymbol{A},\mathcal{H}}$

Laplacian Convolutional Representation (LCR):

- (Solution) Time series trend modeling in the low-rank framework?
  - Global time series trend modeling: $\|\mathcal{C}(\boldsymbol{x})\|_*$
  - Local time series trend modeling: $\|\boldsymbol{\ell}\star\boldsymbol{x}\|_2^2$
- (Highlight) A unified framework with the FFT implementation.

Hankel Tensor Factorization:

- (Highlight) Memory-efficient Hankel indexing & convolutional parameterization.

Collaborators

Dr. HanQin Cai

Xiaoxu Chen

Dr. Zhanhong Cheng

Chengyuan Zhang

Dr. Xi-Le Zhao

References

A short list:

- **[Candès & Recht'09]** Exact matrix completion via convex optimization. Foundations of Computational Mathematics, 9 (6), 717-772.
- **[Cai et al.'10]** A singular value thresholding algorithm for matrix completion
- **[Zhang et al.'12]** Matrix completion by truncated nuclear norm regularization
- **[Hu et al.'12]** Fast and accurate matrix completion via truncated nuclear norm regularization
- **[Lu et al.'14]** Generalized Nonconvex Nonsmooth Low-Rank Minimization
- **[Lu et al.'19]** Tensor Robust Principal Component Analysis with A New Tensor Nuclear Norm
- **[Yokota et al.'18]** Missing Slice Recovery for Tensors Using a Low-rank Model in Embedded Space
- **[Cai et al.'21]** Accelerated Structured Alternating Projections for Robust Spectrally Sparse Signal Recovery
- **[Liu'22]**
- **[Liu & Zhang'23]**

Thanks for your attention!

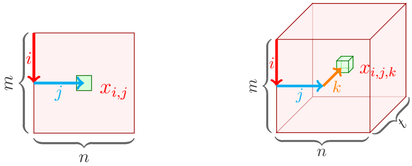
Any Questions?

About me:

- Homepage: https://xinychen.github.io
- GitHub: https://github.com/xinychen
- How to reach me: chenxy346@gmail.com

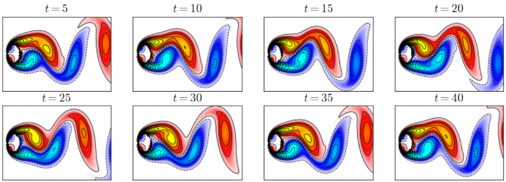
What Is Tensors?

- What is tensor? $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ vs. $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m\times n\times t}$



- Tensors are everywhere!

Color image with RGB channels



Dynamical system (fluid flow)

Convolution & FFT

Flipping Operation in LCR

Results on speed fields

Training, Validation & Testing

Tuning Hyperparameters