POLYTECHNIQUE MONTRÉAL
UNIVERSITÉ D'INGÉNIERIE

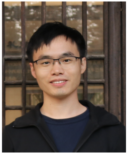
IVADO

Low-Rank Matrix and Tensor Methods for Spatiotemporal Traffic Data Modeling

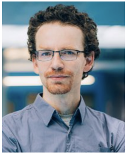
Xinyu Chen

University of Montreal, Canada

May 22, 2023



Ph.D. Candidate
Xinyu Chen



Supervisor
Prof. Nicolas Saunier

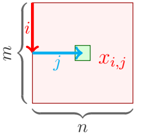
Co-supervisor
Prof. Lijun Sun (McGill)

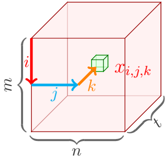
Outline

- **Spatiotemporal Traffic Data**
  Preliminaries: What are matrix & tensor?
  Traffic data & representation & problems
- **Spatiotemporal Traffic Data Imputation**
  Laplacian convolutional representation
- **Sparse Traffic Flow Forecasting**
  Uber movement speed data
  Temporal matrix factorization
- **Dynamic Pattern Discovery**
  Human mobility changes over COVID-19
  Time-varying low-rank autoregression
- **Conclusion**

What are Matrix & Tensor?

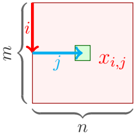
- What is tensor? $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ vs. $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m \times n \times t}$

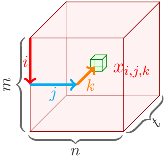




What are Matrix & Tensor?

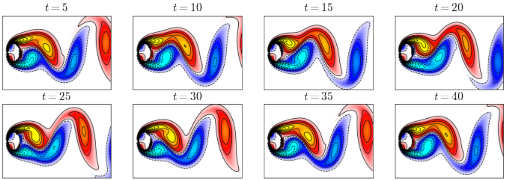
- What is tensor? $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ vs. $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m\times n\times t}$





- Tensors are everywhere!

Color image with RGB channels

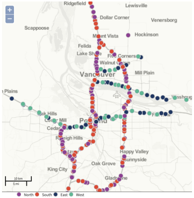


Dynamical system (fluid flow)

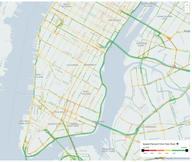
Spatiotemporal Traffic Data

- Spatiotemporal traffic data are indeed matrices or tensors.

Network

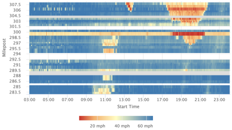


Highway network

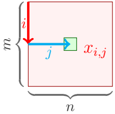


Urban road network

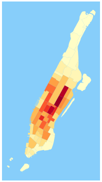
Data & Representation



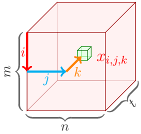
Speed field



Matrix

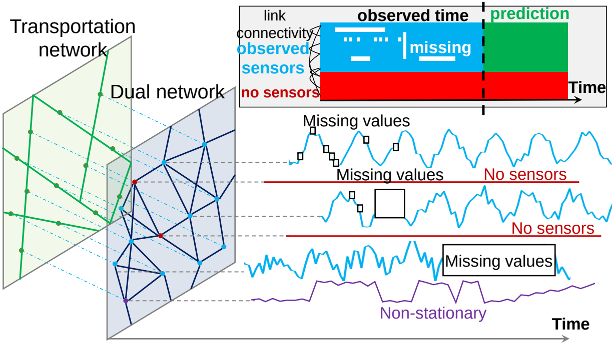


OD matrices over time



Tensor

Spatiotemporal Traffic Data



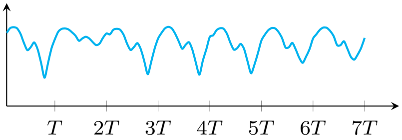
Spatiotemporal Traffic Data Imputation

❶ X. Chen, Z. Cheng, N. Saunier, L. Sun (2022). Laplacian convolutional representation for traffic time series imputation. arXiv preprint arXiv:2212.01529.

Laplacian Convolutional Representation

Motivation:

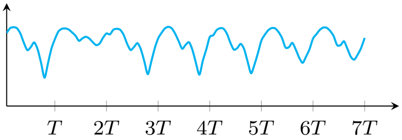
- Global trends (e.g., long-term quasi-seasonality & daily/weekly rhythm):



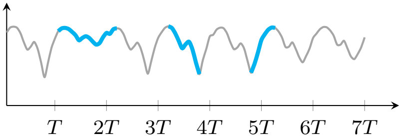
Laplacian Convolutional Representation

Motivation:

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



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

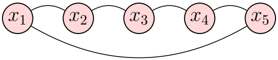


- **Question**: How to characterize both global and local trends in sparse traffic data?

Laplacian Convolutional Representation

Reformulate Laplacian regularization with circular convolution.

- Intuition of (circulant) Laplacian matrix.



Undirected and circulant graph

$\xrightarrow{\text{Modeling}}$

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

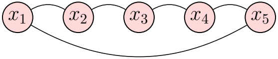
(Circulant) Laplacian matrix

- Laplacian kernel: $\boldsymbol{\ell}=(2,-1,0,0,-1)^\top$.

Laplacian Convolutional Representation

Reformulate Laplacian regularization with circular convolution.

- Intuition of (circulant) Laplacian matrix.



Undirected and circulant graph

$$\xrightarrow{\text{Modeling}} \quad \boldsymbol{L} = \begin{bmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ -1 & 0 & 0 & -1 & 2 \end{bmatrix}$$

(Circulant) Laplacian matrix

- Laplacian kernel: $\boldsymbol{\ell} = (2, -1, 0, 0, -1)^\top$.
- Define Laplacian kernel:

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

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

- Local trend modeling via (Laplacian) temporal regularization:

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

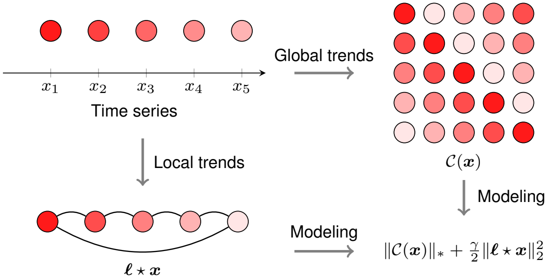
Laplacian Convolutional Representation

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.,

$$\begin{aligned}\min_{\boldsymbol{x}} \; & \|\mathcal{C}(\boldsymbol{x})\|_* + \gamma \cdot \mathcal{R}_\tau(\boldsymbol{x}) \\ \text{s.t. } & \|\mathcal{P}_\Omega(\boldsymbol{x} - \boldsymbol{y})\|_2 \leq \epsilon\end{aligned}$$

where $\mathcal{C}: \mathbb{R}^T \to \mathbb{R}^{T \times T}$ denotes the circulant operator. $\|\cdot\|_*$ denotes the nuclear norm of matrix, namely, the sum of singular values.



Laplacian Convolutional Representation

- LCR model:

$$\begin{aligned}\min_{\boldsymbol{x}}\ & \|\mathcal{C}(\boldsymbol{x})\|_{*}+\gamma\cdot\mathcal{R}_{\tau}(\boldsymbol{x}) \\ \text{s.t. } & \|\mathcal{P}_{\Omega}(\boldsymbol{x}-\boldsymbol{y})\|_{2}\leq\epsilon\end{aligned}$$

- Augmented Lagrangian function:

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

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

Laplacian Convolutional Representation

- LCR model:

$$\begin{aligned}\min_{\boldsymbol{x}}\ & \|\mathcal{C}(\boldsymbol{x})\|_{*}+\gamma\cdot\mathcal{R}_{\tau}(\boldsymbol{x}) \\ \text{s.t. } & \|\mathcal{P}_{\Omega}(\boldsymbol{x}-\boldsymbol{y})\|_{2}\leq\epsilon\end{aligned}$$

- Augmented Lagrangian function:

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

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

- The ADMM scheme:

$$\begin{cases}\boldsymbol{x}:=\arg\min\limits_{\boldsymbol{x}}\ \mathcal{L}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{w}) \\ \boldsymbol{z}:=\arg\min\limits_{\boldsymbol{z}}\ \mathcal{L}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{w}) \\ \quad=\dfrac{1}{\lambda+\eta}\mathcal{P}_{\Omega}(\lambda\boldsymbol{x}+\boldsymbol{w}+\eta\boldsymbol{y})+\dfrac{1}{\lambda}\mathcal{P}_{\Omega}^{\perp}(\lambda\boldsymbol{x}+\boldsymbol{w}) \\ \boldsymbol{w}:=\boldsymbol{w}+\lambda(\boldsymbol{x}-\boldsymbol{z})\end{cases}$$

Laplacian Convolutional Representation

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

$$\boldsymbol{x} := \arg\min_{\boldsymbol{x}} \ \|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\gamma}{2}\|\boldsymbol{\ell}\star\boldsymbol{x}\|_2^2 + \frac{\lambda}{2}\|\boldsymbol{x}-\boldsymbol{z}+\boldsymbol{w}/\lambda\|_2^2$$

Laplacian Convolutional Representation

- Optimize $\boldsymbol{x}$ via fast Fourier transform (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 $\{\hat{\boldsymbol{\ell}},\hat{\boldsymbol{x}},\hat{\boldsymbol{z}},\hat{\boldsymbol{w}}\}$ refers to $\{\boldsymbol{\ell},\boldsymbol{x},\boldsymbol{z},\boldsymbol{w}\}$ in the frequency domain.

Laplacian Convolutional Representation

- Optimize $\boldsymbol{x}$ via fast Fourier transform (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 $\{\hat{\boldsymbol{\ell}},\hat{\boldsymbol{x}},\hat{\boldsymbol{z}},\hat{\boldsymbol{w}}\}$ refers to $\{\boldsymbol{\ell},\boldsymbol{x},\boldsymbol{z},\boldsymbol{w}\}$ in the frequency domain.

ℓ_1-norm Minimization in Complex Space (Liu & Zhang'22)

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

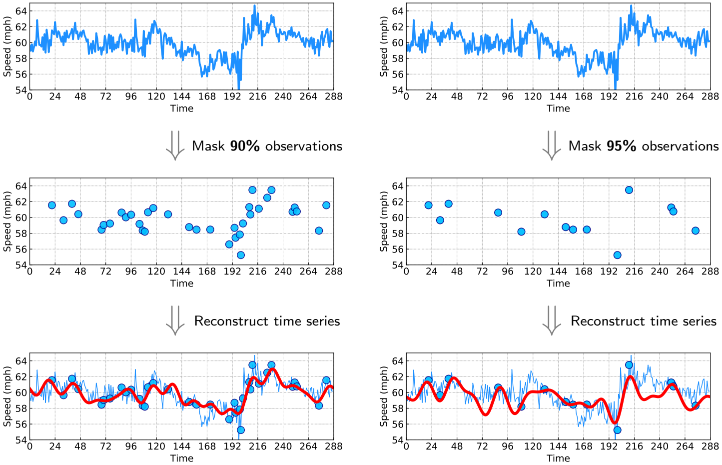
$$\min_{\hat{\boldsymbol{x}}}\ \|\hat{\boldsymbol{x}}\|_1 + \frac{\omega}{2}\|\hat{\boldsymbol{x}}-\hat{\boldsymbol{h}}\|_2^2$$

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

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

Laplacian Convolutional Representation

- On traffic speed time series:



- LCR can characterize both global and local trends and produce accurate results.

Sparse Traffic Flow Forecasting

❷ X. Chen, L. Sun (2022). Bayesian temporal factorization for multidimensional time series prediction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44 (9): 4659–4673.

- ∘ $100 + citationsonGoogleScholar$
- ∘ $ESIhighlycitedpaper(top1\%)$

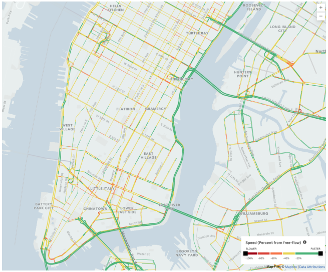
∘ $ESIhotpaper(top0.1\%)\underline{X.Chen}, C.Zhang, X.-$
$L.Zhao, N.Saunier, L.Sun(2022).Nonstationarytemporalmatrixfactorizationfor$
2203.10651.

(Under 2nd review at Transportation Research Part C: Emerging Technologies)

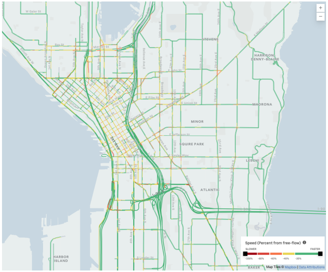
Sparse Traffic Flow Forecasting

Motivation:

- Uber (hourly) movement speed data[1]



NYC movement



Seattle movement

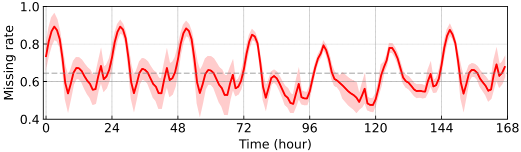
- The average speed on a given road segment for each hour of each day.
- Hourly speeds are computed when road segments have 5+ unique trips.
- **Issue**: insufficient sampling of ridesharing vehicles on the road network.

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

Sparse Traffic Flow Forecasting

High-dimensionality & Sparsity

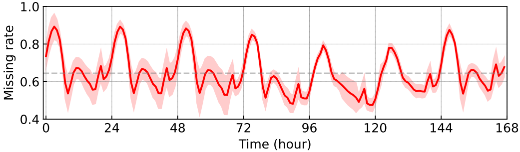
- **NYC** movement speed data (2019)
  - **98,210** road segments & 8,760 time steps (hours)
  - Whole missing rate: **64.43%**



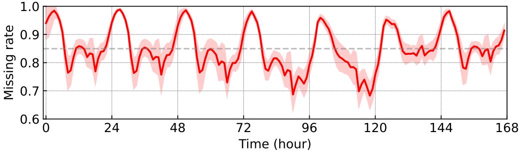
Sparse Traffic Flow Forecasting

High-dimensionality & Sparsity

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

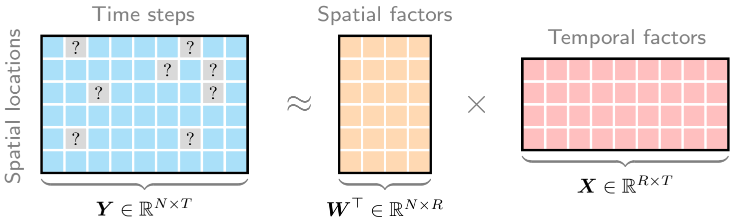


- **Seattle** movement speed data (2019)
  - **63,490** road segments & 8,760 time steps (hours)
  - Whole missing rate: **84.95%**



Sparse Traffic Flow Forecasting

- Spatiotemporal data can be reconstructed by low-dimensional latent factors!



- MF optimization problem

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

  with factor matrices $\boldsymbol{W}$ and $\boldsymbol{X}$.

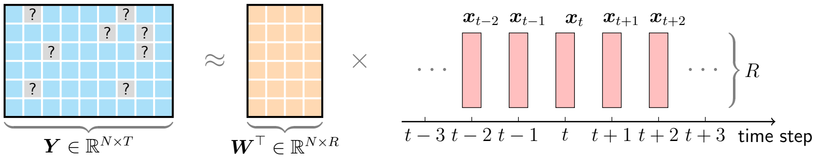
- Disadvantages:
  - Cannot capture temporal correlations.
  - Cannot perform time series forecasting.

Sparse Traffic Flow Forecasting

Temporal matrix factorization (Yu et al.'16; Chen & Sun'22)

Given any partially observed time series data $\boldsymbol{Y} \in \mathbb{R}^{N \times T}$ with observed index set Ω, then temporal matrix factorization assumes a dth-order vector autoregressive (VAR) process on the temporal factor matrix:

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



GitHub repositories:

- **transdim**: Machine learning for spatiotemporal traffic data imputation and forecasting. (970+ stars & 270+ forks)

  https://github.com/xinychen/transdim

- **tracebase**: Multivariate time series forecasting on high-dimensional and sparse Uber movement speed data. (30+ stars)

  https://github.com/xinychen/tracebase

- **awesome-latex-drawing**: Academic drawing examples in LaTeX. (1,000+ stars & 140+ forks)

  https://github.com/xinychen/awesome-latex-drawing

Dynamic Pattern Discovery

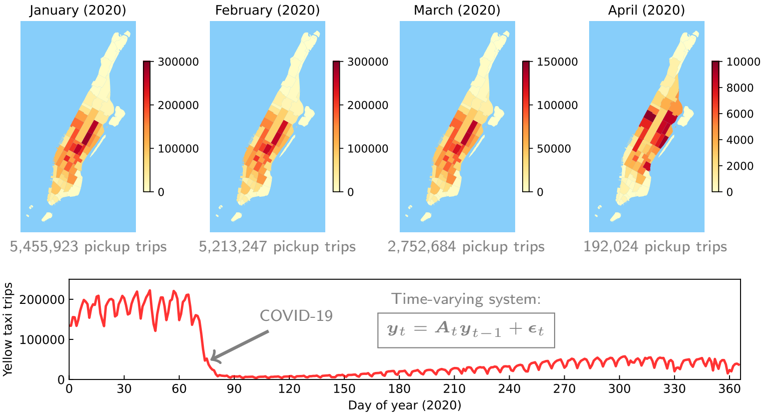
❹ X. Chen, C. Zhang[†], X. Chen, N. Saunier, L. Sun (2022). Discovering dynamic patterns from spatiotemporal data with time-varying low-rank autoregression. arXiv preprint arXiv:2211.15482.

(Under 2nd review at IEEE Transactions on Knowledge and Data Engineering)

Dynamic Pattern Discovery

Motivation:

- NYC (yellow) taxi data[2]



- How to characterize the dynamic patterns?

[2] https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page

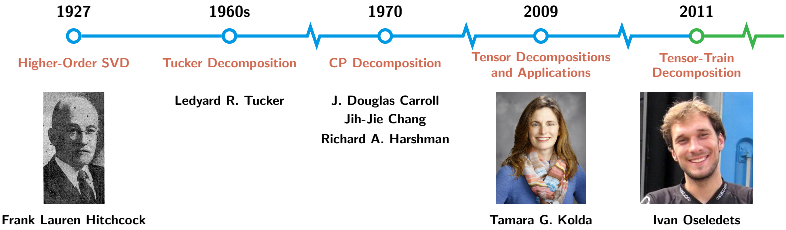
Dynamic Pattern Discovery

- Given a sequence of spatiotemporal measurements $\boldsymbol{y}_t \in \mathbb{R}^N$, $t = 1, 2, \ldots, T$

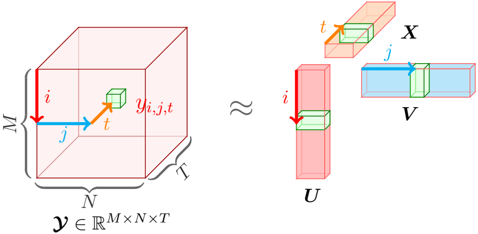
$$\min_{\{\boldsymbol{A}_t\}} \quad \underbrace{\frac{1}{2} \sum_t \left\| \boldsymbol{y}_t - \boldsymbol{A}_t \boldsymbol{y}_{t-1} \right\|_2^2}_{\text{Time-varying autoregression}}$$

- A sequence of coefficient matrices $\{\boldsymbol{A}_t\}$ of size $N \times N$.

[**Over-parameterization**] $\mathcal{O}(N^2(T-1))$ parameters vs. $\mathcal{O}(NT)$ data.


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 decomposition**: Factorize $\boldsymbol{\mathcal{Y}}$ into the combination of rank-R factor matrices, i.e., $\boldsymbol{\mathcal{Y}} \approx \sum_{r=1}^{R} \boldsymbol{u}_r \otimes \boldsymbol{v}_r \otimes \boldsymbol{x}_r$.



Dynamic Pattern Discovery

- **(Ours)** Parameterize coefficients via tensor factorization (TF):

$$\min_{\boldsymbol{W},\boldsymbol{G},\boldsymbol{V},\boldsymbol{X}} \quad \underbrace{\frac{1}{2}\sum_{t}\left\|\boldsymbol{y}_t - \boldsymbol{W}\boldsymbol{G}(\boldsymbol{x}_t^\top \otimes \boldsymbol{V})^\top \boldsymbol{y}_{t-1}\right\|_2^2}_{\text{Let } \boldsymbol{\mathcal{A}}_t = \boldsymbol{\mathcal{G}} \times_1 \boldsymbol{W} \times_2 \boldsymbol{V} \times_3 \boldsymbol{x}_t^\top \text{ be the TF}}$$

- Alternating minimization (Let f be the obj.)

$$\boldsymbol{W} := \{\boldsymbol{W} \mid \frac{\partial f}{\partial \boldsymbol{W}} = \boldsymbol{0}\} \qquad \boldsymbol{G} := \{\boldsymbol{G} \mid \frac{\partial f}{\partial \boldsymbol{G}} = \boldsymbol{0}\}$$

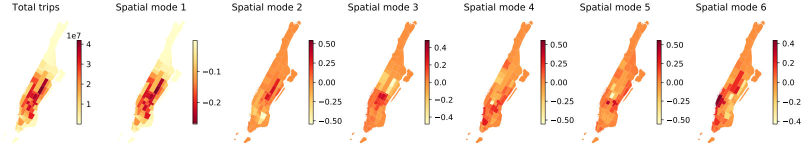
$$\boldsymbol{V} := \{\boldsymbol{V} \mid \frac{\partial f}{\partial \boldsymbol{V}} = \boldsymbol{0}\} \qquad \boldsymbol{x}_t := \{\boldsymbol{x}_t \mid \frac{\partial f}{\partial \boldsymbol{x}_t} = \boldsymbol{0}\}$$

- Solve each subproblem by **conjugate gradient** and **least squares**.

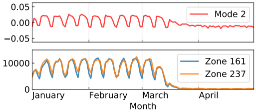
- Time-varying autoregression with TF

$$\min_{\boldsymbol{W},\boldsymbol{G},\boldsymbol{V},\boldsymbol{X}} \; \frac{1}{2}\sum_{t}\left\|\boldsymbol{y}_t - \boldsymbol{W}\boldsymbol{G}(\boldsymbol{x}_t^\top \otimes \boldsymbol{V})^\top \boldsymbol{y}_{t-1}\right\|_2^2$$

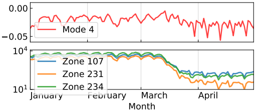
- **NYC taxi dataset** (pickup)



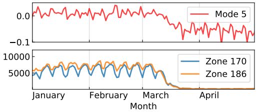
Spatial modes in $\boldsymbol{W}$



Mode #2 & taxi trips (2020)



Mode #4 & taxi trips (2020)



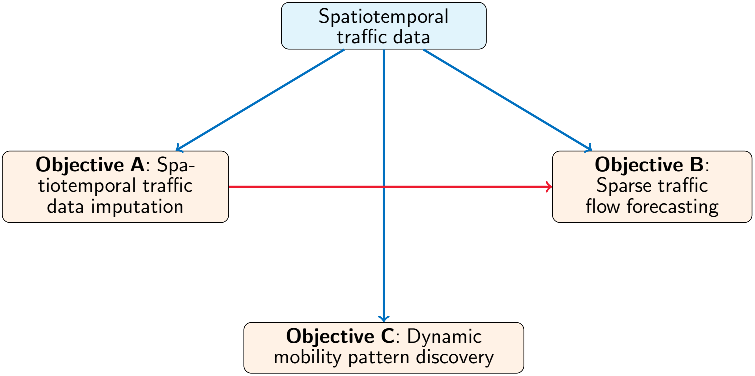
Mode #5 & taxi trips (2020)

Temporal modes in $\boldsymbol{X}$

- Produce interpretable patterns and identify the changing point of system (mainly due to COVID-19).

Conclusion

We give some studies on **spatiotemporal traffic data modeling**.



References

A short list:

- **[Liu & Zhang'22]** G. Liu and W. Zhang (2022). Recovery of future data via convolution nuclear norm minimization. IEEE Transactions on Information Theory, 69(1), 650–665.
- **[Yu et al.'16]** H.-F. Yu, N. Rao, and I. S. Dhillon (2016). Temporal regularized matrix factorization for high-dimensional time series prediction. Advances in neural information processing systems (NIPS).

Thanks for your attention!

Any Questions?

Slides: https://xinychen.github.io/slides/traffic_data_modeling_v1.pdf

About me:

- Homepage: https://xinychen.github.io
- Google Scholar: user=mCrW04wAAAAJhl (**700+ citations**)
- GitHub: https://github.com/xinychen (**3.3k+ stars**)
- Blog: https://medium.com/@xinyu.chen (**70k+ views**)
- How to reach me: chenxy346@gmail.com