Modeling Urban Traffic Data with Matrix and Tensor Approaches

2024 INFORMS Annual Meeting

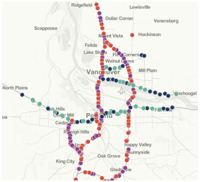
Xinyu Chen

Postdoctoral Associate, MIT

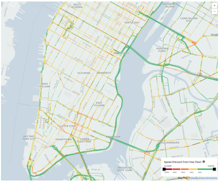
October 21, 2024

Urban Traffic Data

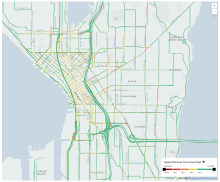
- Transport & mobility application scenarios



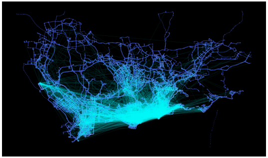
Highway (Portland)



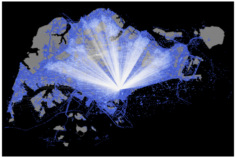
Uber movement (NYC)



Uber movement (Seattle)

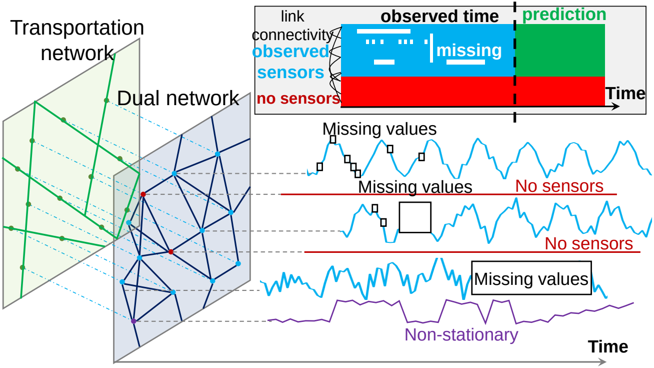


Taxi trajectory (Shenzhen)



Passenger flow (Singapore)

- Challenges: Sparsity, high-dimensionality, and multi-dimensionality, time-varying system



Transportation network
Dual network
link connectivity
observed time
prediction
observed sensors
missing
no sensors
Time
Missing values
Missing values
No sensors
No sensors
Missing values
Non-stationary
Time

Papers:

- X. Chen, Z. Cheng, H.Q. Cai, N. Saunier, L. Sun (2024). "Laplacian Convolutional Representation for Traffic Time Series Imputation". IEEE Transactions on Knowledge and Data Engineering, 36 (11): 6490–6502.
- 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.
- X. Chen, X.L. Zhao, C. Cheng (2024). "Forecasting Urban Traffic States with Sparse Data Using Hankel Temporal Matrix Factorization". INFORMS Journal on Computing. Early access.
- X. Chen, C. Zhang, X. Chen, N. Saunier, L. Sun (2024). "Discovering Dynamic Patterns from Spatiotemporal Data with Time-Varying Low-Rank Autoregression". IEEE Transactions on Knowledge and Data Engineering, 36 (2): 504–517.

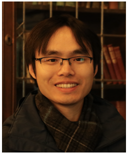
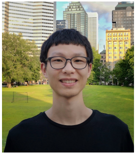
ML $\Rightarrow$ Imputation & Prediction & Pattern Discovery

Laplacian Convolutional Representation for Traffic Time Series Imputation

IEEE Transactions on Knowledge and Data Engineering, 2024

Code: https://github.com/xinychen/transdim

Blog: https://spatiotemporal-data.github.io/posts/ts_conv


Xinyu Chen
UdeM → MIT

Zhanhong Cheng
McGill → UF

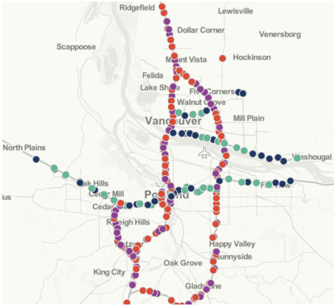
HanQin Cai
UCF

Nicolas Saunier
PolyMtl

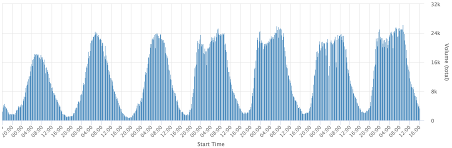
Lijun Sun
McGill

Traffic Flow Data

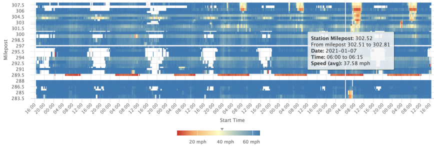
- Portland highway traffic data[1]



Highway network & sensor locations



Traffic volume



Traffic speed

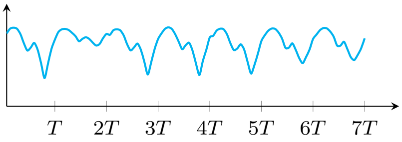
- $\boldsymbol{X} \in \mathbb{R}^{N \times T}$ with N spatial locations $\times$ T time steps
- Traffic volume/speed shows strong spatial/temporal dependencies

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

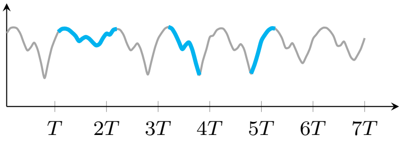
Time Series Imputation

Motivation: Traffic imputation

- Global trends (e.g., daily/weekly periodicity):



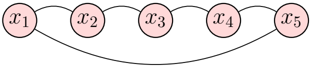
- Local trends (e.g., short-term time series trends):



How to characterize both global and local trends in sparse data?

Local Trend Modeling

- 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:

$$\boldsymbol{\ell} \triangleq \underbrace{(2, -1, 0, 0, -1)^\top}_{\text{first column of } \boldsymbol{L}}$$

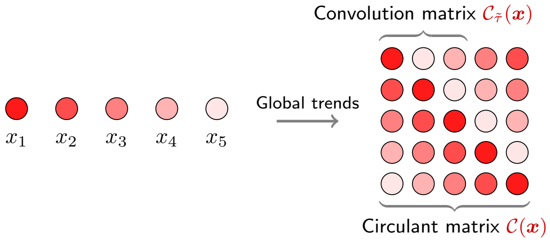
extending to the degree 2τ for $\boldsymbol{x} \in \mathbb{R}^T$.

- Temporal regularization:

$$\mathcal{R}(\boldsymbol{x}) = \underbrace{\frac{1}{2}\|\boldsymbol{L}\boldsymbol{x}\|_2^2}_{\text{mat-vec mul.}} = \underbrace{\frac{1}{2}\|\boldsymbol{\ell} \star \boldsymbol{x}\|_2^2}_{\text{conv. } \star}$$

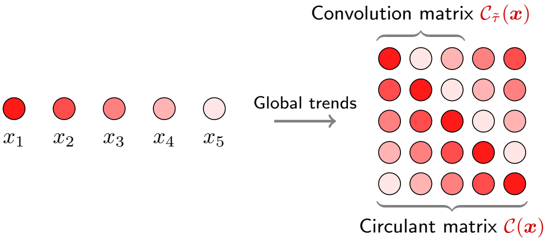
Global Trend Modeling

Circulant matrix $\mathcal{C}(\boldsymbol{x})$ vs. convolution matrix $\mathcal{C}_{\tilde{\tau}}(\boldsymbol{x})$



Global Trend Modeling

Circulant matrix $\mathcal{C}(\boldsymbol{x})$ vs. convolution matrix $\mathcal{C}_{\tilde{\tau}}(\boldsymbol{x})$



- Circulant/Convolution nuclear norm minimization
  - A balance between global and local trends modeling?

CircNNM (Liu'22, Liu & Zhang'23)

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

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

ConvNNM (Liu'22, Liu & Zhang'23)

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

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

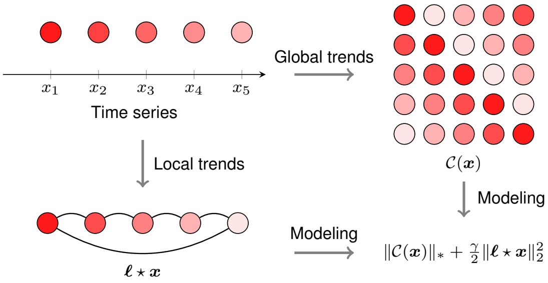
Global + Local Trends?

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/local time series trends:

$$\min_{\boldsymbol{x}} \; \underbrace{\|\mathcal{C}(\boldsymbol{x})\|_*}_{\text{global}} + \frac{\gamma}{2} \underbrace{\|\boldsymbol{\ell} \star \boldsymbol{x}\|_2^2}_{\text{local}}$$

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



Laplacian Convolutional Representation

- Augmented Lagrangian function:[2]

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

- The ADMM scheme:

$$\begin{cases}\boldsymbol{x}:=\arg\min\limits_{\boldsymbol{x}}\;\mathcal{L}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{w}) & \text{(Nuclear norm minimization)}\\ \boldsymbol{z}:=\arg\min\limits_{\boldsymbol{z}}\;\mathcal{L}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{w}) & \text{(Closed-form solution)}\\ \boldsymbol{w}:=\boldsymbol{w}+\lambda(\boldsymbol{x}-\boldsymbol{z}) & \text{(Standard update)}\end{cases}$$

- Optimize $\boldsymbol{x}$?

$$\underbrace{\|\mathcal{C}(\boldsymbol{x})\|_*=\|\mathcal{F}(\boldsymbol{x})\|_1}_{\text{property of circulant matrix}}\qquad\&\qquad\underbrace{\frac{1}{2}\|\boldsymbol{\ell}\star\boldsymbol{x}\|_2^2=\frac{1}{2T}\|\mathcal{F}(\boldsymbol{\ell})\circ\mathcal{F}(\boldsymbol{x})\|_2^2}_{\text{property of circular convolution}}$$

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

[2] $\boldsymbol{w}\in\mathbb{R}^T$ (Lagrange multiplier); $\langle\cdot,\cdot\rangle$ (inner product).

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 $\{\hat{\boldsymbol{\ell}},\hat{\boldsymbol{x}},\hat{\boldsymbol{z}},\hat{\boldsymbol{w}}\}\triangleq\mathcal{F}\{\boldsymbol{\ell},\boldsymbol{x},\boldsymbol{z},\boldsymbol{w}\}$ (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:

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

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

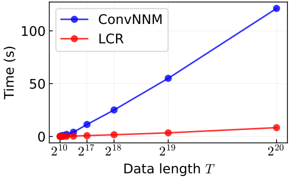
$$\hat{x}_t := \frac{\hat{h}_t}{|\hat{h}_t|}\cdot\underbrace{\max\{0,|\hat{h}_t|-1/\delta\}}_{\text{shrinkage (e.g., ReLU)}},\ t\in[T].$$

Laplacian Convolutional Representation

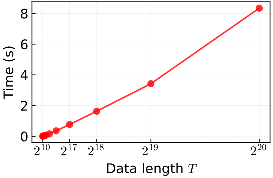
Empirical time complexity

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

- Ours: **LCR**
  - An FFT implementation in $\mathcal{O}(T \log T)$
  - The logarithmic factor $\log T$ makes the FFT highly efficient
- Baseline: **ConvNNM** (Liu'22, Liu & Zhang'23)
  - Convolution matrix $\mathcal{C}_{\tilde{\tau}}(\boldsymbol{y}) \in \mathbb{R}^{T \times \tilde{\tau}}$ with kernel size $\tilde{\tau} = 2^4$
  - Singular value thresholding in $\mathcal{O}(\tilde{\tau}^2 T)$

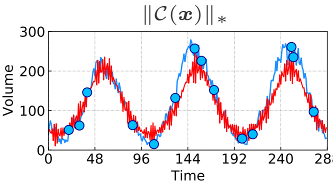


ConvNNM vs. LCR

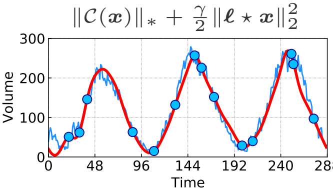


LCR

- Traffic speed imputation[3]



$\Downarrow$ Plus **local** time series trends



Highlights:

- Rethinking the importance of local trend modeling in traffic data imputation tasks.
- Finding a unified global and local trend modeling framework whose optimization can be efficiently solved by FFT:

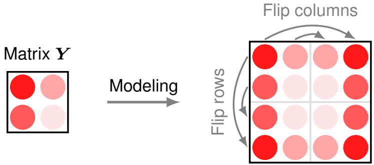
$$\min_{\boldsymbol{x}} \underbrace{\|\mathcal{C}(\boldsymbol{x})\|_*}_{\text{global}} + \frac{\gamma}{2} \underbrace{\|\boldsymbol{\ell} \star \boldsymbol{x}\|_2^2}_{\text{local}}$$

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

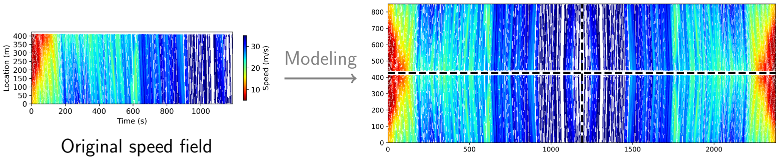
[3] Blue dot: partial observation; red line: imputation.

Speed field reconstruction[4]

- Flipping operation on a matrix:



- Flipping operation on a speed field of vehicular traffic flow:



Original speed field

Flipped speed field

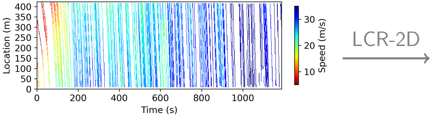
[4]Highway Drone (HighD) dataset at https://www.highd-dataset.com/

Speed field reconstruction

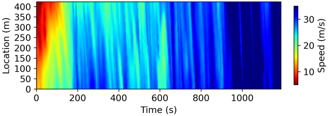
- Scenario: Mask trajectories of 70% vehicles
- LCR-2D on partially observed $\boldsymbol{Y} \in \mathbb{R}^{N \times T}$:

$$\min_{\boldsymbol{X}} \quad \underbrace{\|\mathcal{C}(\boldsymbol{X})\|_*}_{\text{global trend}} + \frac{\gamma}{2} \underbrace{\|(\boldsymbol{\ell}_s \boldsymbol{\ell}^\top) \star \boldsymbol{X}\|_F^2}_{\text{local trend}}$$

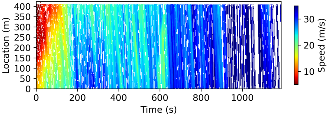
$$\text{s.t. } \|\mathcal{P}_\Omega(\boldsymbol{X} - \boldsymbol{Y})\|_F \leq \epsilon$$



Sparse speed field



Reconstructed speed field

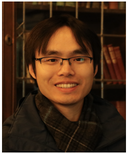


Ground-truth speed field

Bayesian Temporal Factorization for Multidimensional Time Series Prediction

IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022

Code: https://github.com/xinychen/transdim



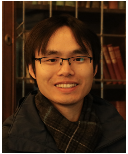
Xinyu Chen
UdeM → MIT

Lijun Sun
McGill

Forecasting Urban Traffic States with Sparse Data Using Hankel Temporal Matrix Factorization

INFORMS Journal on Computing, 2024

Code: https://github.com/xinychen/tracebase



Xinyu Chen
UdeM → MIT

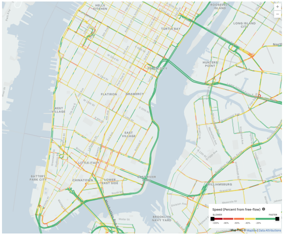
Xi-Le Zhao
UESTC

Chun Cheng
DUT

Revisit Traffic Prediction

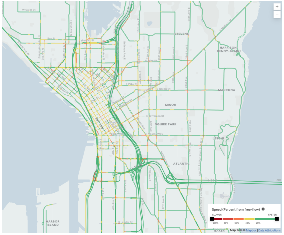
A classical problem w/ new ideas?

- Uber (hourly) movement speed data[5]



NYC movement

64.43% (missing)



Seattle movement

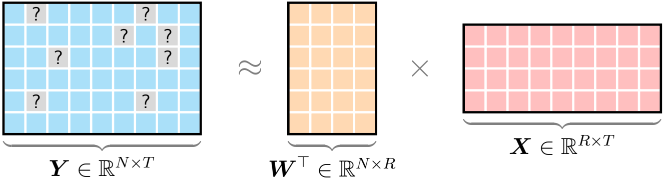
84.95% (missing)

  - {road segment, time step (hour), average speed}
  - Computing hourly speed: Road segments have 5+ unique trips.
- Challenge: Forecasting network-wide traffic states with sparse data.

[5] https://movement.uber.com/ (not available now)

Matrix Factorization

A simple approach to reconstruct missing values.



MF (Koren et al.'09)

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

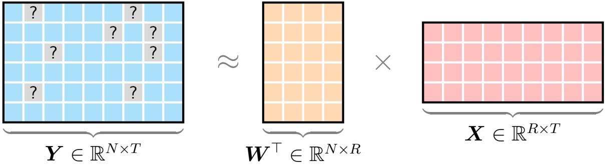
$$\min_{\boldsymbol{W},\boldsymbol{X}} \; \frac{1}{2}\|\mathcal{P}_\Omega(\boldsymbol{Y} - \boldsymbol{W}^\top \boldsymbol{X})\|_F^2$$

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

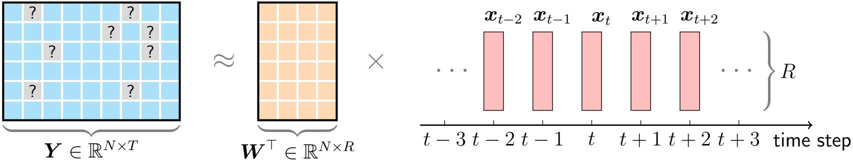
- ✓ Learn from sparse data
- ✓ Spatial factor matrix $\boldsymbol{W}$
- ✓ Temporal factor matrix $\boldsymbol{X}$
- ✗ Temporal correlations?

Temporal Matrix Factorization

Vector autoregression (VAR) on the temporal factor matrix.



$\Downarrow$ $\boldsymbol{X}$ is time series?



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

Vector autoregression (VAR) on the temporal factor matrix.

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

$$\boldsymbol{x}_t = \sum_{k=1}^{d}\boldsymbol{A}_k\boldsymbol{x}_{t-k} + \underbrace{\boldsymbol{\epsilon}_t}_{\mathcal{N}(\boldsymbol{0},\boldsymbol{I})}$$

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

$\Downarrow$ Yu et al.'16 **Chen & Sun'22**

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

Bayesian Temporal Matrix Factorization

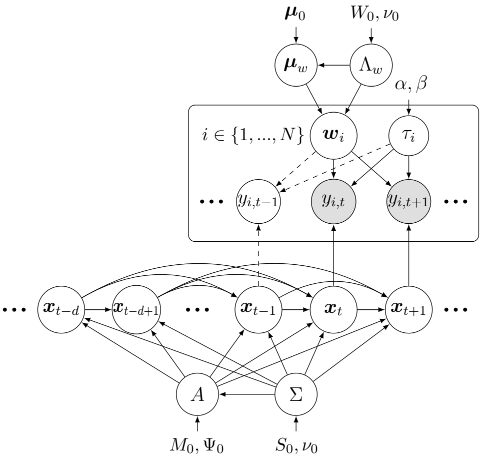
- Bayesian network (Chen & Sun'22)
  - Observations $(i,t)\in\Omega$:

    $$y_{i,t}\sim\mathcal{N}(\underbrace{\boldsymbol{w}_i^\top\boldsymbol{x}_t}_{\text{MF}},\tau_i^{-1})$$

  - Prior of parameters:

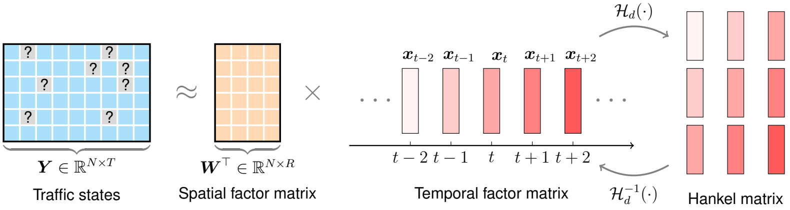
    $$\begin{cases}\boldsymbol{w}_i\sim\mathcal{N}(\boldsymbol{\mu}_w,\Lambda_w^{-1})\\ \boldsymbol{x}_t\sim\mathcal{N}(\underbrace{\boldsymbol{A}\boldsymbol{x}_{t-1}}_{\text{VAR}},\Sigma)\\ \tau_i\sim\text{Gamma}(\alpha,\beta)\end{cases}$$

  - Conjugate prior of hyperparameters.



Hankel Temporal Matrix Factorization

- HTMF (Chen, Zhao, & Cheng'24)
  - Automatic temporal modeling via Hankel matrices.



- Optimization problem

$$\min_{\boldsymbol{W}, \boldsymbol{X}, \boldsymbol{F}} \underbrace{\frac{1}{2}\left\|\mathcal{P}_\Omega(\boldsymbol{Y} - \boldsymbol{W}^\top \boldsymbol{X})\right\|_F^2}_{\text{MF on observations } \boldsymbol{Y}} + \underbrace{\frac{\gamma}{2}\|\boldsymbol{F} - \boldsymbol{X}\|_F^2}_{\text{bias mitigation}}$$

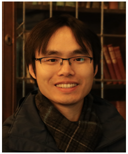
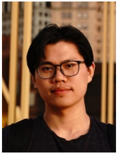
$$\text{s.t.} \quad \underbrace{\operatorname{rank}(\mathcal{H}_d(\boldsymbol{F})) = R}_{\text{rank-}R \text{ Hankel matrix}}$$

Discovering Dynamic Patterns from Spatiotemporal Data with Time-Varying Low-Rank Autoregression

IEEE Transactions on Knowledge and Data Engineering, 2024

Code: https://github.com/xinychen/vars

Blog: https://spatiotemporal-data.github.io/posts/time_varying_model


Xinyu Chen
UdeM → MIT

Chengyuan Zhang⋆
McGill

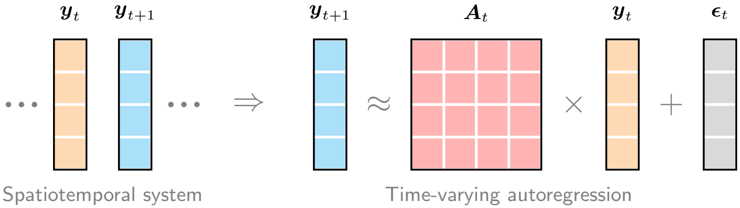
Xiaoxu Chen
McGill

Nicolas Saunier
PolyMtl

Lijun Sun
McGill

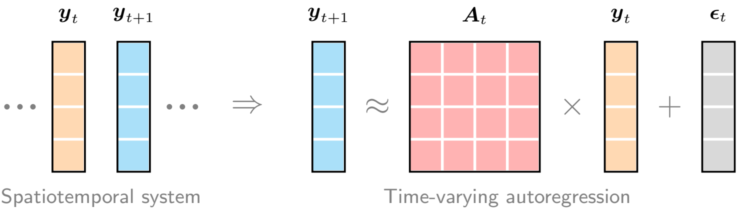
Autoregression

- How to characterize dynamical systems?



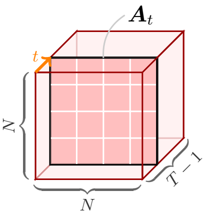
Autoregression

- How to characterize dynamical systems?

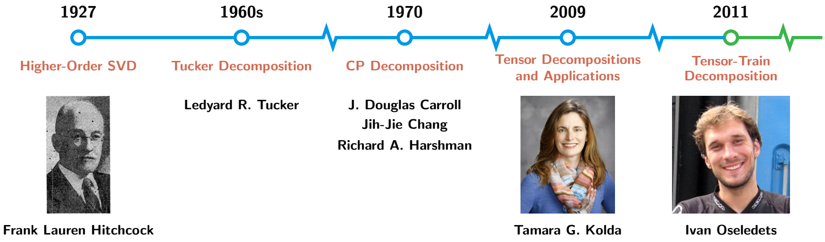


- On spatiotemporal systems $\boldsymbol{Y} \in \mathbb{R}^{N \times T}$:

$$\underbrace{\boldsymbol{y}_{t+1} = \boldsymbol{A}\boldsymbol{y}_t + \boldsymbol{\epsilon}_t}_{\text{time-invariant (e.g., DMD)}} \quad \text{v.s.} \quad \underbrace{\boldsymbol{y}_{t+1} = \boldsymbol{A}_t\boldsymbol{y}_t + \boldsymbol{\epsilon}_t}_{\text{time-varying, but over-para.}}$$



- How to discover spatial/temporal modes (patterns) from the tensor $\boldsymbol{\mathcal{A}} \triangleq \{\boldsymbol{A}_t\}_{t \in [T-1]}$?


1927
1960s
1970
2009
2011
Higher-Order SVD
Tucker Decomposition
CP Decomposition
Tensor Decompositions and Applications
Tensor-Train Decomposition
Frank Lauren Hitchcock
Ledyard R. Tucker
J. Douglas Carroll
Jih-Jie Chang
Richard A. Harshman
Tamara G. Kolda
Ivan Oseledets

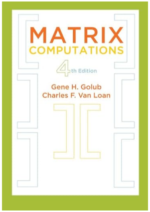
Time-Varying Autoregression

- Tensor factorization[6]:

$$\boldsymbol{\mathcal{A}}=\underbrace{\boldsymbol{\mathcal{G}}\times_1\boldsymbol{W}\times_2\boldsymbol{V}\times_3\boldsymbol{X}}_{\text{Tucker decomposition}}$$

$$\Updownarrow$$

$$\boldsymbol{A}_t=\boldsymbol{\mathcal{G}}\times_1\underbrace{\boldsymbol{W}}_{\text{spatial modes}}\times_2\boldsymbol{V}\times_3\underbrace{\boldsymbol{x}_t^\top}_{\text{temporal modes}}$$



- **(Ours)** Time-varying low-rank autoregression:

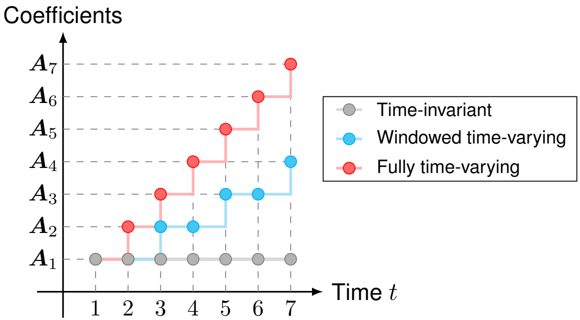
$$\min_{\boldsymbol{\mathcal{G}},\boldsymbol{W},\boldsymbol{V},\boldsymbol{X}}\ \frac{1}{2}\sum_{t\in[T-1]}\left\|\boldsymbol{y}_{t+1}-\underbrace{(\boldsymbol{\mathcal{G}}\times_1\boldsymbol{W}\times_2\boldsymbol{V}\times_3\boldsymbol{x}_t^\top)}_{\text{Tucker decomposition}}\boldsymbol{y}_t\right\|_2^2$$

- Dominant spatial/temporal patterns in e.g., $\{\boldsymbol{W},\boldsymbol{X}\}$.

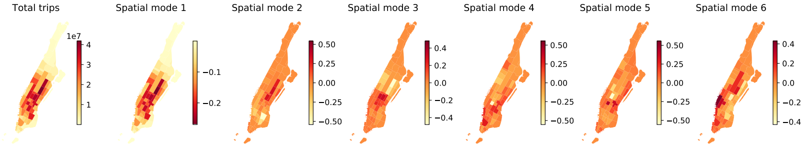
[6] $\times_k$, $\forall k$ is the mode-k product between tensor and matrix/vector.

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

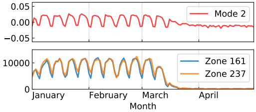
$$\underbrace{\boldsymbol{y}_{t+1} = \boldsymbol{A}\boldsymbol{y}_t + \boldsymbol{\epsilon}_t}_{\text{time-invariant (e.g., DMD)}} \quad \text{v.s.} \quad \underbrace{\boldsymbol{y}_{t+1} = \boldsymbol{A}_t\boldsymbol{y}_t + \boldsymbol{\epsilon}_t}_{\text{fully time-varying (ours)}}$$



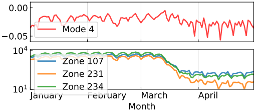
- **NYC taxi dataset** (pickup)



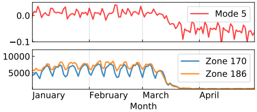
Spatial patterns in $\boldsymbol{W}$



Pattern #2 & taxi trips (2020)



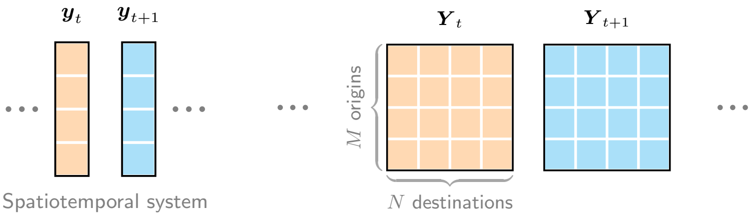
Pattern #4 & taxi trips (2020)



Pattern #5 & taxi trips (2020)

Vision & Insight

- Discovering spatial/temporal patterns from 2D and 3D spatiotemporal systems with unsupervised learning:
  - Time-varying autoregression on the data
  - Tensor factorization on the coefficients



Thanks for your attention!

Any Questions?

Slides: https://xinychen.github.io/slides/informs24.pdf

Reproducible research: https://spatiotemporal-data.github.io

About me:

- Homepage: https://xinychen.github.io
- How to reach me: chenxy346@gmail.com
- Or send to: xinychen@mit.edu