Spatiotemporal Data Modeling

Dynamic Patterns and Long-Range Correlations

Xinyu Chen

University of Montreal, Canada

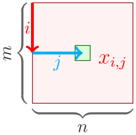
December 4, 2023

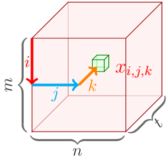
Past work: (Pattern discovery, traffic data imputation/forecasting)

1. X. Chen, C. Zhang, X. Chen, N. Saunier, L. Sun (2023). Discovering dynamic patterns from spatiotemporal data with time-varying low-rank autoregression. IEEE Transactions on Knowledge and Data Engineering. Early access.
2. X. Chen, Z. Cheng, N. Saunier, L. Sun (2022). Laplacian convolutional representation for traffic time series imputation. arXiv preprint arXiv:2212.01529.
3. 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.

Tensors

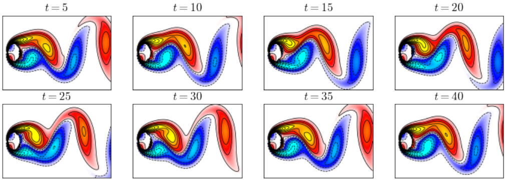
- What is tensor? $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ vs. $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m \times n \times t}$





- Tensors are everywhere!

Color image with RGB channels



Dynamical system (fluid flow)

Tensor Factorization

- Revisit tensor factorization (TF)

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

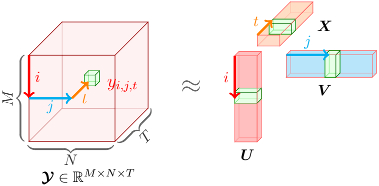
2009 — Tensor Decompositions and Applications

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).



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

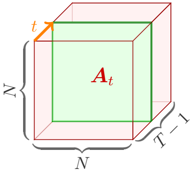
Time-Varying Autoregression

- **Motivation: Identify the dynamic patterns of time-varying systems.**
  (e.g., fluid flow, climate variables, and human mobility)

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

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

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



Coefficient tensor $\boldsymbol{\mathcal{A}}$

Parameterize coefficients via TF:

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{G}} \times_1 \boldsymbol{W} \times_2 \boldsymbol{V} \times_3 \boldsymbol{X}$$

$$\Rightarrow \quad \boldsymbol{A}_t = \boldsymbol{\mathcal{G}} \times_1 \boldsymbol{W} \times_2 \boldsymbol{V} \times_3 \boldsymbol{x}_t^\top$$

Significance: Spatial models/patterns $\boldsymbol{W}$ & temporal modes/patterns $\boldsymbol{X}$

Time-Varying Autoregression

- Optimization problem:

$$\min_{\boldsymbol{W},\boldsymbol{\mathcal{G}},\boldsymbol{V},\boldsymbol{X}} \frac{1}{2}\sum_{t=2}^{T}\left\|\boldsymbol{y}_t-(\boldsymbol{\mathcal{G}}\times_1\boldsymbol{W}\times_2\boldsymbol{V}\times_3\boldsymbol{x}_t^\top)\boldsymbol{y}_{t-1}\right\|_2^2$$

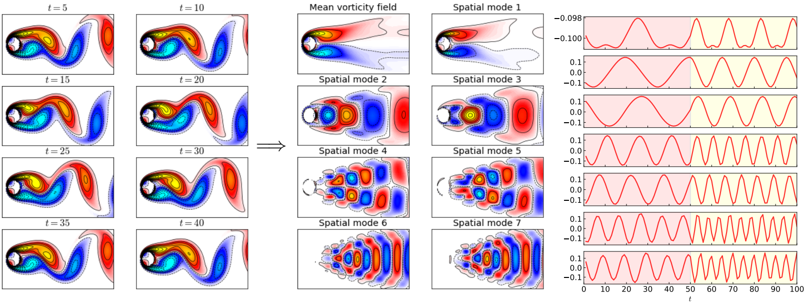
- Alternating minimization:

Algorithm (Chen et al.'23)

Repeat

- Estimate $\boldsymbol{\mathcal{G}}$ by least squares (LS)
- Estimate $\boldsymbol{W}$ by LS
- Estimate $\boldsymbol{V}$ by conjugate gradient (CG, Golub & Van Loan'13)
- **For** $t\in\{2,3,\dots,T\}$ **do**
  - Estimate $\boldsymbol{x}_t$ by LS

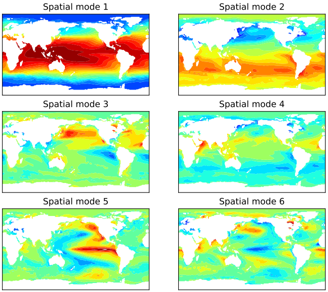
- **Multi-resolution fluid flow dataset** (the first 50 snapshots + 50 snapshots randomly selected from the last 100 snapshots)



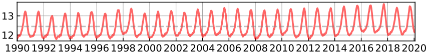
Vectorize snapshots as $\{\boldsymbol{y}_t\}$ Spatial modes in $\boldsymbol{W}$ Temporal modes in $\boldsymbol{X}$

- Produce interpretable patterns
- Identify the system of different frequencies

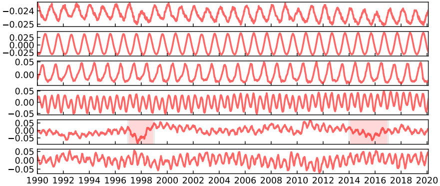
- Sea surface temperature dataset



Spatial modes in $\boldsymbol{W}$



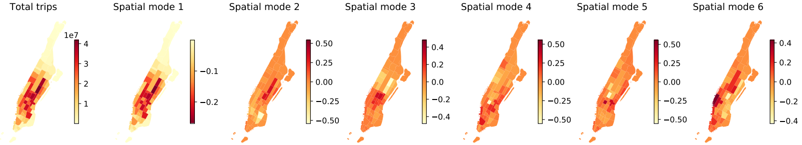
Time series of mean temperature



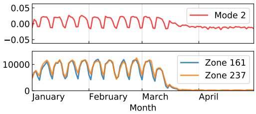
Temporal modes in $\boldsymbol{X}$

Identify two strongest El Nino events (on 1997-98 & 2014-16).

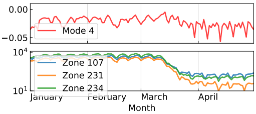
- **NYC taxi dataset** (pickup)



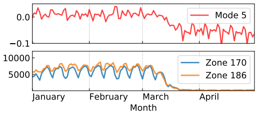
Spatial modes in $\boldsymbol{W}$



Mode #2 & taxi trips (2020)



Mode #4 & taxi trips (2020)



Mode #5 & taxi trips (2020)

Temporal modes in $\boldsymbol{X}$

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

Orthogonal Spatial/Temporal Patterns?

- Optimization problem with orthogonal constraints:[1]

$$\min_{\boldsymbol{W},\boldsymbol{\mathcal{G}},\boldsymbol{V},\boldsymbol{X}} \ \frac{1}{2}\sum_{t=2}^{T}\left\|\boldsymbol{y}_t-(\boldsymbol{\mathcal{G}}\times_1\boldsymbol{W}\times_2\boldsymbol{V}\times_3\boldsymbol{x}_t^\top)\boldsymbol{y}_{t-1}\right\|_2^2$$

$$\text{s.t.}\quad \begin{cases}\boldsymbol{W}^\top\boldsymbol{W}=\boldsymbol{I}_R\\ \boldsymbol{V}^\top\boldsymbol{V}=\boldsymbol{I}_R\\ \boldsymbol{X}^\top\boldsymbol{X}=\boldsymbol{I}_R\end{cases}$$

Algorithm (Chen et al.'23)

Repeat

- Estimate $\boldsymbol{\mathcal{G}}$ by LS
- Estimate $\boldsymbol{W}$ by LS
- Estimate $\boldsymbol{V}$ by conjugate gradient
- **For** $t\in\{2,3,\dots,T\}$ **do**
  - Estimate $\boldsymbol{x}_t$ by LS

$\Rightarrow$

Algorithm

Repeat

- Estimate $\boldsymbol{\mathcal{G}}$ by least squares
- Estimate $\boldsymbol{W}$ by orthogonal projection (Cai et al.'14)
- Estimate $\boldsymbol{V}$ by projected conjugate gradient (Calamai & More'87)
- Estimate $\boldsymbol{X}$ by projected conjugate gradient (Calamai & More'87)

[1] https://spatiotemporal-data.github.io/probs/orth-var

Spatiotemporal Mobility Networks

State-Space Model (SSM)

- **State transition equation:**

$$\underbrace{\boldsymbol{x}_{t+1}}_{\text{state}} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\underbrace{\boldsymbol{u}_t}_{\text{input}} + \underbrace{\boldsymbol{w}_t}_{\mathcal{N}(\boldsymbol{0},\boldsymbol{I})}$$

- **Observation equation:**

$$\underbrace{\boldsymbol{y}_t}_{\text{output}} = \boldsymbol{C}\boldsymbol{x}_t + \boldsymbol{D}\underbrace{\boldsymbol{u}_t}_{\text{input}} + \underbrace{\boldsymbol{v}_t}_{\mathcal{N}(\boldsymbol{0},\boldsymbol{I})}$$

vs.

Recurrent Neural Network (RNN)

- **Hidden state update:**

$$\boldsymbol{h}_t = \text{activation}(\boldsymbol{W}_{hh}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{hx}\boldsymbol{x}_t + \underbrace{\boldsymbol{b}_h}_{\text{bias}})$$

- **Output:**

$$\underbrace{\boldsymbol{y}_t}_{\text{output}} = \text{activation}(\boldsymbol{W}_{yh}\underbrace{\boldsymbol{h}_t}_{\text{state}} + \underbrace{\boldsymbol{b}_y}_{\text{bias}})$$

- Long-range spatiotemporal modeling, e.g.,
  - State-space layers (Smith et al.'22)
  - Convolutional SSM (Smith et al.'23)

References

A short list:

- [Cai et al.'14] J.-F. Cai, H. Ji, Z. Shen, and G.-B. Ye, "Data-driven tight frame construction and image denoising," Applied and Computational Harmonic Analysis, vol. 37, no. 1, pp. 89–105, 2014.
- [Calamai & More'87] P. H. Calamai and J. J. More, "Projected gradient methods for linearly constrained problems," Mathematical programming, vol. 39, no. 1, pp. 93–116, 1987.
- [Golub & Van Loan'13] G. H. Golub and C. F. Van Loan, Matrix computations. JHU press, 2013.
- [Smith et al.'22] J. T. H. Smith, A. Warrington, and S. W. Linderman, "Simplified state space layers for sequence modeling." arXiv preprint arXiv:2208.04933. (ICLR'23)
- [Smith et al.'23] J. T. H. Smith, S. De Mello, J. Kautz, S. W. Linderman, and W. Byeon, "Convolutional State Space Models for Long-Range Spatiotemporal Modeling." arXiv preprint arXiv:2310.19694. ()

POLYTECHNIQUE MONTRÉAL
UNIVERSITÉ D'INGÉNIERIE

IVADO

Thanks for your attention!

Any Questions?

About me:

- Homepage: https://xinychen.github.io
- GitHub: https://github.com/xinychen (**4k+ stars**)
- How to reach me: chenxy346@gmail.com