



**POLYTECHNIQUE
MONTRÉAL**

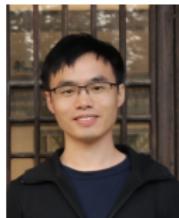
UNIVERSITÉ
D'INGÉNIERIE



Laplacian Convolutional Representation for Traffic Time Series Imputation

Xinyu Chen

July 19, 2023



Ph.D. Candidate

Xinyu Chen
Polytechnique Montréal



Postdoc

Dr. Zhanhong Cheng
McGill University



Supervisor

Prof. Nicolas Saunier
Polytechnique Montréal



Co-supervisor

Prof. Lijun Sun
McGill University

Current work:

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

GitHub repository:

- **transdim**: Machine learning for spatiotemporal traffic data imputation and forecasting. (1,000+ stars & 270+ forks on GitHub)
<https://github.com/xinychen/transdim>

Slides:

- <https://xinychen.github.io/slides/LCR.pdf>

Outline

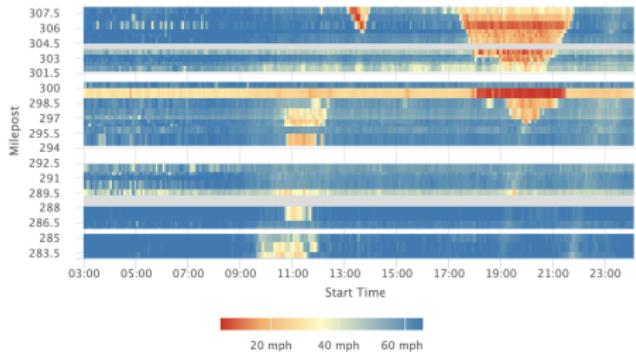
- **Motivation**
- **Low-Rank Laplacian Convolutional Model**
 - Reformulate Laplacian Regularization
 - Traffic Time Series Imputation
 - Multivariate Model for Speed Field Reconstruction
- **Conclusion**

Motivation

- Portland highway traffic speed data¹



Highway network & sensor locations



Traffic speed field

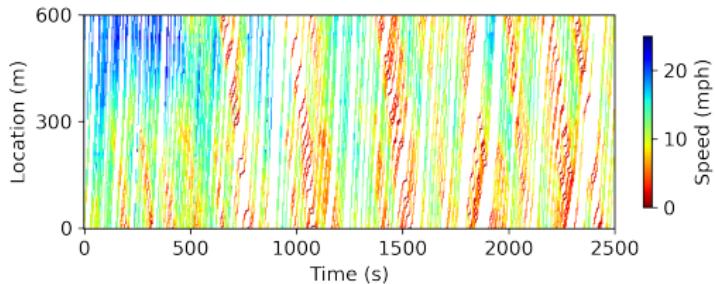
- Speed field $\mathbf{Y} \in \mathbb{R}^{N \times T}$ (N locations & T time steps)
- Speed field shows strong spatial/temporal dependencies

¹<https://portal.its.pdx.edu/home>

Motivation

- How to reconstruct missing values from partial observations?

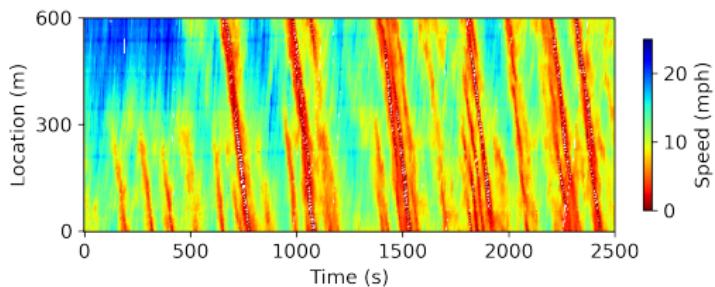
Motivation



200-by-500 matrix
(NGSIM)



Reconstruct speed field from
20% sparse trajectories?



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

Low-Rank Laplacian Convolutional Model

Reformulate Laplacian regularization with circular convolution.

- Intuition of (circulant) Laplacian matrix.

The diagram illustrates the mapping from a graph structure to a matrix. On the left, an "Undirected and circulant graph" is shown with five nodes labeled x_1 , x_2 , x_3 , x_4 , and x_5 . Node x_1 is connected to x_2 and x_5 . Node x_2 is connected to x_1 and x_3 . Node x_3 is connected to x_2 and x_4 . Node x_4 is connected to x_3 and x_5 . Node x_5 is connected to x_1 and x_4 . This represents a circulant graph where each node has the same neighborhood pattern as every other node due to the underlying cyclic structure. An arrow labeled "Modeling" points from the graph to the right, where the (Circulant) Laplacian matrix is given by:

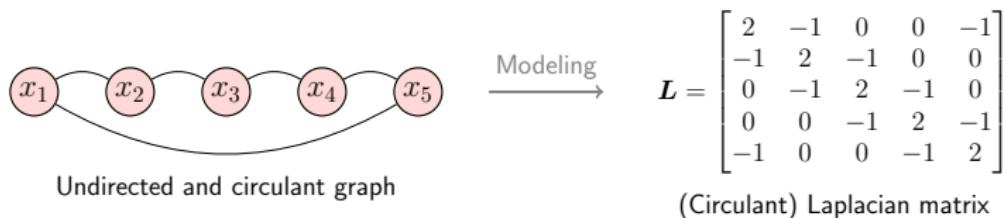
$$\mathbf{L} = \begin{bmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ -1 & 0 & 0 & -1 & 2 \end{bmatrix}$$

(Circulant) Laplacian matrix

Low-Rank Laplacian Convolutional Model

Reformulate Laplacian regularization with circular convolution.

- Intuition of (circulant) Laplacian matrix.



Low-Rank Laplacian Convolutional Model

Reformulate Laplacian regularization with circular convolution.

- Intuition of (circulant) Laplacian matrix.

Undirected and circulant graph

Modeling →

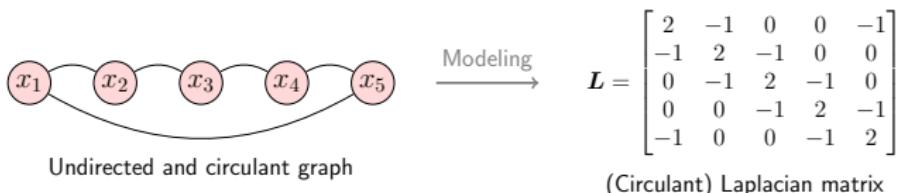
$$\mathbf{L} = \begin{bmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ -1 & 0 & 0 & -1 & 2 \end{bmatrix}$$

(Circulant) Laplacian matrix

Low-Rank Laplacian Convolutional Model

Reformulate Laplacian regularization with circular convolution.

- Intuition of (circulant) Laplacian matrix.



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

- Define Laplacian kernel:

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

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

- Local trend modeling via (Laplacian) temporal regularization:

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

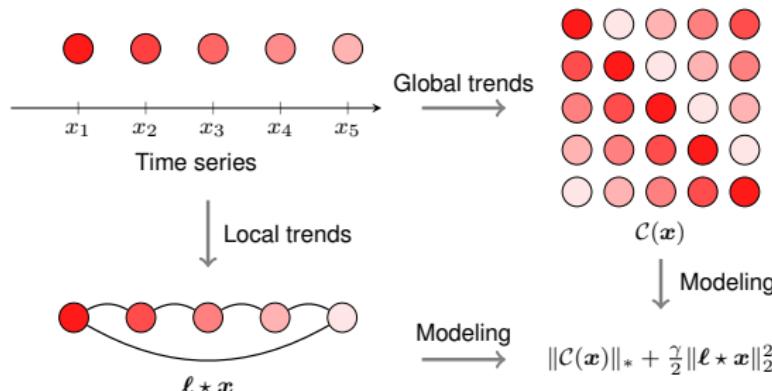
Low-Rank Laplacian Convolutional Model

Laplacian Convolutional Representation (LCR)

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

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

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



Low-Rank Laplacian Convolutional Model

- LCR model:

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

- Augmented Lagrangian function:

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

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

- The ADMM scheme:

$$\left\{ \begin{array}{l} \mathbf{x} := \arg \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mathbf{z}, \mathbf{w}) \\ \mathbf{z} := \arg \min_{\mathbf{z}} \mathcal{L}(\mathbf{x}, \mathbf{z}, \mathbf{w}) \\ = \frac{1}{\lambda + \eta} \mathcal{P}_\Omega(\lambda \mathbf{x} + \mathbf{w} + \eta \mathbf{y}) + \mathcal{P}_\Omega^\perp(\mathbf{x} + \mathbf{w}/\lambda) \\ \mathbf{w} := \mathbf{w} + \lambda(\mathbf{x} - \mathbf{z}) \end{array} \right.$$

Low-Rank Laplacian Convolutional Model

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

$$\begin{aligned}\mathbf{x} &:= \arg \min_{\mathbf{x}} \|\mathcal{C}(\mathbf{x})\|_* + \frac{\gamma}{2} \|\ell * \mathbf{x}\|_2^2 + \frac{\lambda}{2} \|\mathbf{x} - \mathbf{z} + \mathbf{w}/\lambda\|_2^2 \\ \implies \hat{\mathbf{x}} &:= \arg \min_{\hat{\mathbf{x}}} \|\hat{\mathbf{x}}\|_1 + \frac{\gamma}{2T} \|\hat{\ell} \circ \hat{\mathbf{x}}\|_2^2 + \frac{\lambda}{2T} \|\hat{\mathbf{x}} - \hat{\mathbf{z}} + \hat{\mathbf{w}}/\lambda\|_2^2 \\ &= \arg \min_{\hat{\mathbf{x}}} \|\hat{\mathbf{x}}\|_1 + \frac{\lambda}{2T} \|\hat{\mathbf{x}} - (\lambda \hat{\mathbf{z}} - \hat{\mathbf{w}}) \oslash (\lambda \hat{\ell}^* \circ \hat{\ell} + \lambda \mathbb{1}_T)\|_2^2\end{aligned}$$

where we introduce $\{\hat{\ell}, \hat{\mathbf{x}}, \hat{\mathbf{z}}, \hat{\mathbf{w}}\}$ referring to $\{\ell, \mathbf{x}, \mathbf{z}, \mathbf{w}\}$ in the frequency domain.

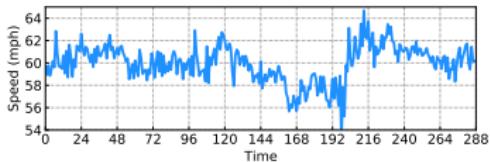
ℓ_1 -norm Minimization in Complex Space (Liu & Zhang'22)

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

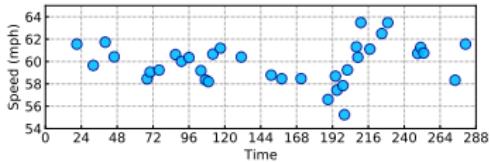
$$\min_{\hat{\mathbf{x}}} \|\hat{\mathbf{x}}\|_1 + \frac{\lambda}{2T} \|\hat{\mathbf{x}} - \hat{\mathbf{h}}\|_2^2$$

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

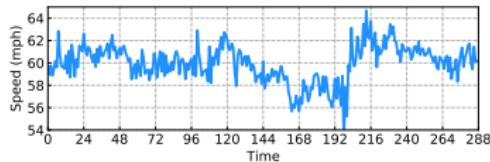
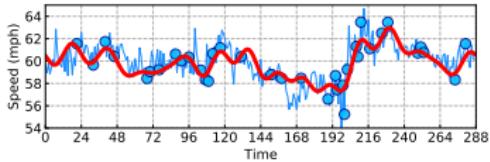
$$\hat{x}_t := \frac{\hat{h}_t}{|\hat{h}_t|} \cdot \max\{0, |\hat{h}_t| - T/\lambda\}, t = 1, \dots, T.$$



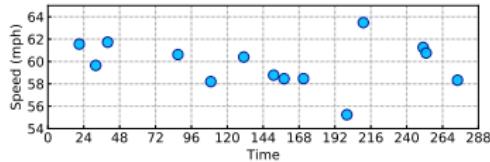
↓ Mask 90% observations



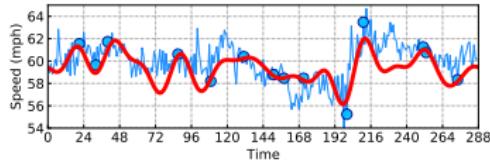
↓ Reconstruct time series



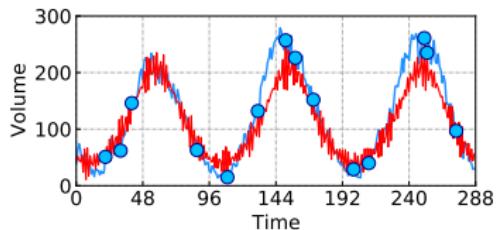
↓ Mask 95% observations



↓ Reconstruct time series

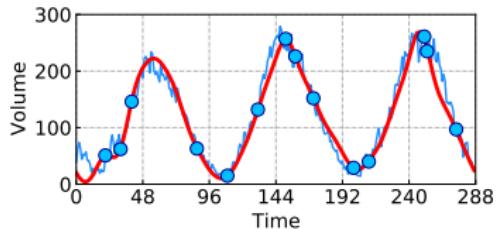


Circulant matrix nuclear norm minimization



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

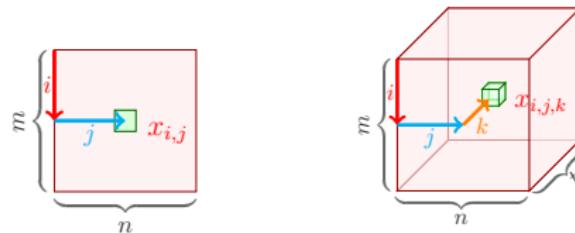
↓ Plus temporal regularization



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

Low-Rank Laplacian Convolutional Model

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

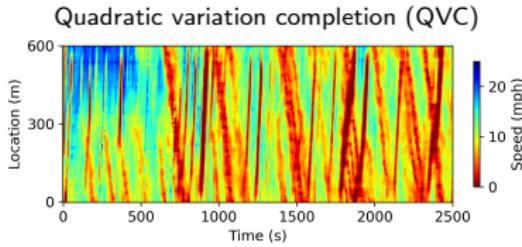


Multivariate LCR (LCR-2D)

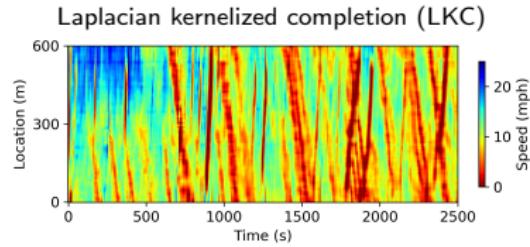
For any partially observed time series $\mathbf{Y} \in \mathbb{R}^{N \times T}$ with observed index set Ω , LCR can be formulated as follows,

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

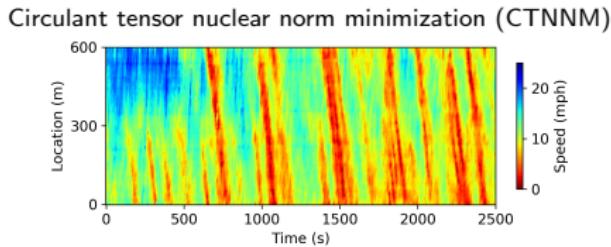
where $\mathcal{C} : \mathbb{R}^{N \times T} \rightarrow \mathbb{R}^{N \times N \times T \times T}$ denotes the circulant operator.



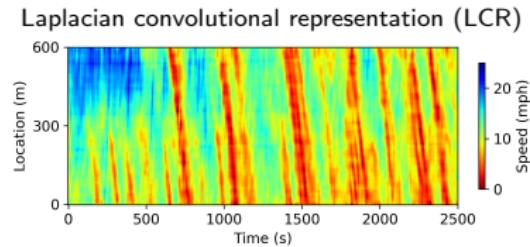
MAPE = 51.50% & RMSE = 4.86mph



MAPE = 46.94% & RMSE = 4.34mph



MAPE = 43.51% & RMSE = 1.65mph



MAPE = 41.29% & RMSE = 1.55mph

- QVC & LKC:

$$\begin{aligned} \min_{\mathbf{X}} \quad & \frac{\gamma}{2} \|(\ell_s \ell^\top) \star \mathbf{X}\|_F \\ \text{s.t. } & \|\mathcal{P}_\Omega(\mathbf{X} - \mathbf{Y})\|_F \leq \epsilon \end{aligned}$$

- CTNNM:

$$\begin{aligned} \min_{\mathbf{X}} \quad & \|\mathcal{C}(\mathbf{X})\|_* \\ \text{s.t. } & \|\mathcal{P}_\Omega(\mathbf{X} - \mathbf{Y})\|_F \leq \epsilon \end{aligned}$$

Conclusion

- **(Starting point)** How to reconstruct sparse speed field?
 - ✓ Matrix factorization (**MF**) ✓ Tensor factorization (**TF**)
- **(Highlight)** The importance of spatiotemporal modeling in low-rank methods?
 - Spatial/temporal **smoothing** regularization:

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{X}} \quad & \frac{1}{2} \left\| \mathcal{P}_{\Omega}(\mathbf{Y} - \mathbf{W}^{\top} \mathbf{X}) \right\|_F^2 + \frac{\rho}{2} (\|\mathbf{W}\|_F^2 + \|\mathbf{X}\|_F^2) \\ & + \frac{\lambda}{2} (\|\mathbf{W} \Psi_1^{\top}\|_F^2 + \|\mathbf{X} \Psi_2^{\top}\|_F^2) \end{aligned}$$

- Automatic temporal modeling via **Hankelization**:

$$\min_{\mathbf{U}, \mathbf{V}, \mathbf{X}} \quad \frac{1}{2} \left\| \mathcal{P}_{\tilde{\Omega}} \left(\mathcal{H}_{\tau}(\mathbf{Y}) - \sum_{r=1}^R \mathbf{u}_r \otimes \mathbf{v}_r \otimes \mathbf{x}_r \right) \right\|_F^2$$

vs.

$$\min_{\mathbf{Q}, \mathbf{S}, \mathbf{U}, \mathbf{V}} \quad \frac{1}{2} \left\| \mathcal{P}_{\dot{\Omega}} \left(\mathcal{H}_{\tau_1, \tau_2}(\mathbf{Y}) - \sum_{r=1}^R \mathbf{q}_r \otimes \mathbf{s}_r \otimes \mathbf{u}_r \otimes \mathbf{v}_r \right) \right\|_F^2$$

- **(Challenge)** How to solve the **high memory consumption** issue in the Hankelization process?



POLYTECHNIQUE
MONTRÉAL

UNIVERSITÉ
D'INGÉNIERIE



Thanks for your attention!

Any Questions?

About me:

- 🏠 Homepage: <https://xinychen.github.io>
- 👤 GitHub: <https://github.com/xinychen> (3.4k+ stars)
- 💻 Blog: <https://medium.com/@xinyu.chen> (70k+ views)
- ✉️ How to reach me: chenxy346@gmail.com