

# Understanding Mobility Periodicity and Speed Field Reconstruction with Machine Learning

**Xinyu Chen**

Assistant Professor

University of Central Florida

Orlando, USA

September 25, 2026

# Data-Driven Transportation Analytics

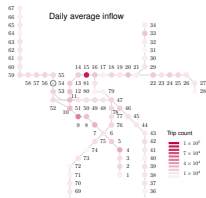
- Spatiotemporal transportation & mobility data



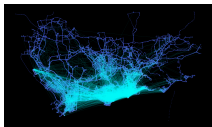
Highway traffic



Uber movement



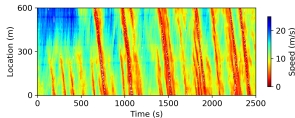
Metro passenger flow



Taxi trajectory



Subway ridership

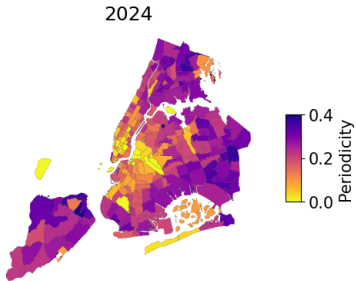


Vehicular traffic flow

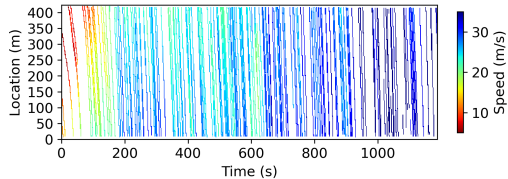
- Challenges: Sparsity, high-dimensionality, multi-dimensionality, heavy tails, irregular sampling, and time-varying systems

# Today's Content

## Part I: Human mobility periodicity



## Part II: Speed field reconstruction



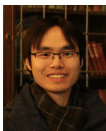
# Human Mobility Periodicity

(arXiv:2506.22895)



<https://github.com/xinychen/integers>

- ☐ Daily & weekly periodicity
- ☐ NYC & Chicago ridehailing
- ☐ Multi-modal mobility
- ☐ Post-pandemic recovery



**Xinyu Chen**  
MIT → UCF



**Vassilis Digalakis Jr**  
BU



**Lijun Ding**  
UCSD



**Dingyi Zhuang**  
MIT

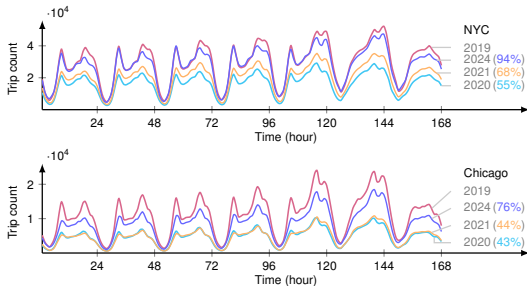
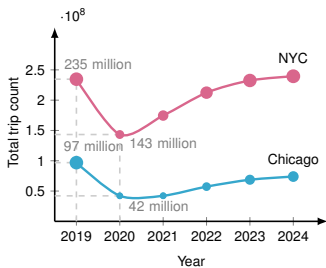


**Jinhua Zhao**  
MIT



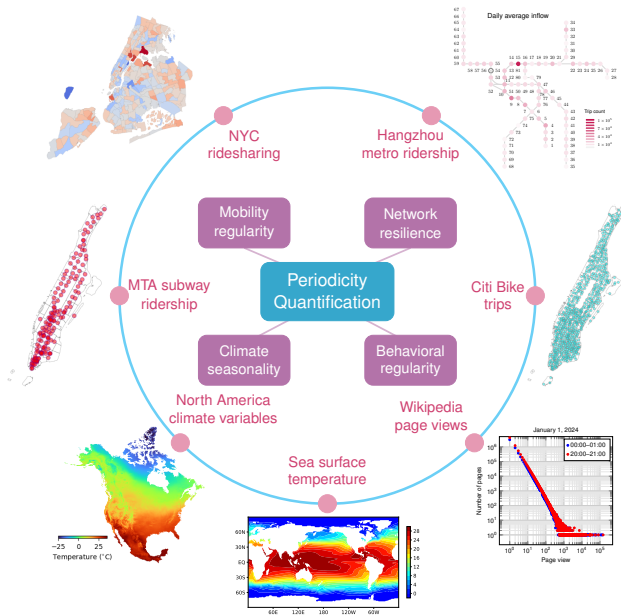
# Motivation

## Ridehailing trip data



## What motivate us most about periodicity?

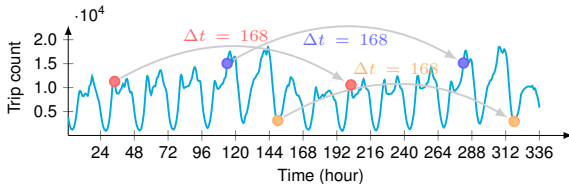
- 1 **Resilience and stability of systems:** Empirically measure the periodicity and predictability of urban systems.
- 2 **Optimization of transport systems:** Optimize resources (e.g., public transit, taxi, ridehailing, micromobility) to meet transport demand efficiently.
- 3 **Demand prediction:** Model long-term travel demand with spatial and temporal mobility patterns.



# Autoregression Reformulation

- Time series autoregression on  $\mathbf{x} \in \mathbb{R}^T$  with order  $d \in \mathbb{Z}^+$

$$\mathbf{w} := \arg \min_{\mathbf{w}} \sum_{t=d+1}^T \left( x_t - \sum_{k=1}^d w_k x_{t-k} \right)^2$$



Periodicity of hourly ridehailing trip time series

- Sparse** coefficient vector  $\mapsto$  **Interpretability**?

$$\underbrace{\mathbf{w}}_{\text{sparsity } \|\mathbf{w}\|_0 \triangleq 3} = \underbrace{(0.33, 0, \dots, 0)}_{k=1}, \underbrace{(0.20)}_{k=167}, \underbrace{(0.46)}_{k=168}^\top \in \mathbb{R}^{168}$$

# Sparse Autoregression

- Identify the dominant auto-correlations

◦  $\tau \in \mathbb{Z}^+$ : Upper bound of the number of nonzero entries in  $\mathbf{w} \in \mathbb{R}^d$

$$\begin{array}{c} \tilde{\mathbf{x}} \\ \begin{array}{|c|} \hline x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ \hline \end{array} \end{array} \approx \begin{array}{c} \mathbf{A} \\ \begin{array}{|c|c|c|c|} \hline x_4 & x_3 & x_2 & x_1 \\ x_5 & x_4 & x_3 & x_2 \\ x_6 & x_5 & x_4 & x_3 \\ x_7 & x_6 & x_5 & x_4 \\ x_8 & x_7 & x_6 & x_5 \\ \hline \end{array} \end{array} \times \begin{array}{c} \mathbf{w} \\ \begin{array}{|c|} \hline w_1 \\ w_2 \\ w_3 \\ w_4 \\ \hline \end{array} \end{array} \left. \vphantom{\begin{array}{c} \mathbf{w} \\ \begin{array}{|c|} \hline w_1 \\ w_2 \\ w_3 \\ w_4 \\ \hline \end{array} \right\} \|\mathbf{w}\|_0 \leq \tau} \right\} \|\mathbf{w}\|_0 \leq \tau$$

$$\begin{aligned} \mathbf{w} &:= \arg \min_{\|\mathbf{w}\|_0 \leq \tau} \sum_{t=d+1}^T \left( x_t - \sum_{k=1}^d w_k x_{t-k} \right)^2 \\ &= \arg \min_{\|\mathbf{w}\|_0 \leq \tau} \|\tilde{\mathbf{x}} - \mathbf{A}\mathbf{w}\|_2^2 \end{aligned}$$

- $\ell_0$ -norm optimization is NP-hard
- Formulate it as a mixed-integer programming

◦ Introduce binary decision variables  $\beta \in \{0, 1\}^d$

$$\begin{array}{ll} \min_{\mathbf{w}} \|\tilde{\mathbf{x}} - \mathbf{A}\mathbf{w}\|_2^2 & \min_{\mathbf{w}, \beta} \|\tilde{\mathbf{x}} - \mathbf{A}\mathbf{w}\|_2^2 \\ \text{s.t. } \underbrace{\|\mathbf{w}\|_0 \leq \tau}_{\clubsuit \text{ sparsity of } \mathbf{w}} & \iff \text{s.t. } \underbrace{-\beta \leq \mathbf{w} \leq \beta}_{\text{bounds being either 0 or } \pm 1} \quad \underbrace{\|\beta\|_1 \leq \tau}_{\clubsuit \text{ sparsity of } \beta} \end{array}$$

# $\ell_0$ -Norm Optimization

$$\min_w \|y - Aw\|_2^2 \quad \text{s. t.} \quad \|w\|_0 \leq \tau$$

Subspace Pursuit (SP)

Wei Dai  
Olgica Milenkovic

Greedy

1993

Orthogonal Matching  
Pursuit (OMP)

Y. C. Pati  
R. Rezaeiifar  
P. S. Krishnaprasad

2008

Compressive Sampling Matching  
Pursuit (CoSaMP)

Deanna Needell  
Joel A. Tropp

2015

Modern Optimization (i.e., MIP)

Dimitris Bertsimas  
Angela King  
Rahul Mazumder

# $\ell_0$ -Norm Optimization

$$\min_w \|y - Aw\|_2^2 \quad \text{s. t.} \quad \|w\|_0 \leq \tau$$

Subspace Pursuit (SP)

Wei Dai  
Olga Milenkovic

Greedy

1993

Orthogonal Matching  
Pursuit (OMP)

Y. C. Pati  
R. Rezaifar  
P. S. Krishnaprasad

2008

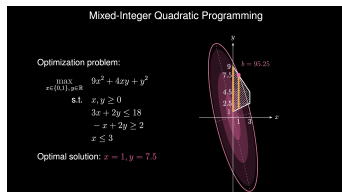
Compressive Sampling Matching  
Pursuit (CoSaMP)

Deanna Needell  
Joel A. Tropp

2015

Modern Optimization (i.e., MIP)

Dimitris Bertsimas  
Angela King  
Rahul Mazumder



[youtube.com/@data-worker](https://youtube.com/@data-worker)

# Sparse Autoregression Done Right

$$\min_{\mathbf{w}, \boldsymbol{\beta}} \underbrace{\sum_{t=d+1}^T \left( x_t - \sum_{k=1}^d w_k x_{t-k} \right)^2}_{\text{Time series autoregression}} \quad \text{s.t.} \quad \underbrace{-\beta_k \leq w_k \leq \beta_k}_{\text{Lower and upper bounds}}, \quad \underbrace{\sum_{k=1}^d \beta_k \leq \tau}_{\clubsuit \text{ Sparsity}}, \quad \underbrace{\beta_k \in \{0, 1\}}_{\text{Binary variable}}$$

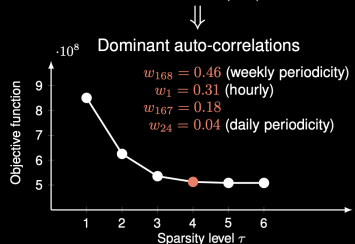
- $\mathbf{w} \in \mathbb{R}^d$ : Auto-correlations
- $\boldsymbol{\beta} \in \{0, 1\}^d$ : Sparsity pattern
- $d = 168$ : Autoregression order

```

1 import numpy as np
2 from docplex.mp.model import Model
3
4 def sparse_ar(x, d, tau):
5     model = Model('Sparse Autoregression')
6     T = x.shape[0]
7     w = [model.continuous_var(name = f'w_{k}') for k in range(d)]
8     beta = [model.binary_var(name = f'beta_{k}') for k in range(d)]
9     model.minimize(model.sum([x[t] - model.sum(w[k] * x[t - k - 1]
10                                     for k in range(d))) ** 2
11                             for t in range(d, T)])
12     model.add_constraint(model.sum(beta[k] for k in range(d)) <= tau)
13     for k in range(d):
14         model.add_constraint(w[k] <= beta[k])
15         model.add_constraint(w[k] >= - beta[k])
16     solution = model.solve()
17     return np.array(solution.get_values(w))

```

<https://github.com/xinychen/integers>

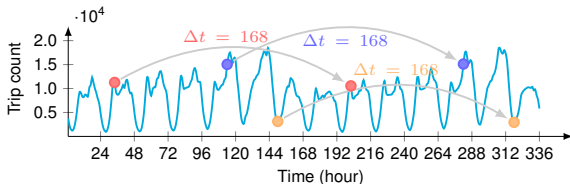


## Solution Quality $\rightarrow$ Better Interpretability?

- Sparse autoregression

$$\min_{w \geq 0} \|\tilde{x} - Aw\|_2^2 \quad \text{s.t.} \quad \|w\|_0 \leq \tau$$

- Subspace pursuit (SP) sometimes fails



Periodicity of ridehailing trip time series

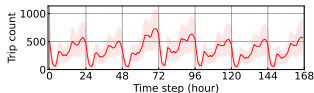
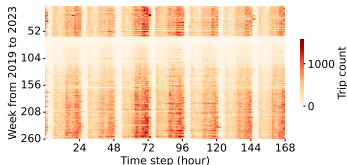
- Exact solution w/ mixed-integer programming (MIP)
- An intuitive example (sparsity  $\tau = 2$ ):

$$\underbrace{w = (\dots, \underbrace{0.02}_{k=53}, \dots, \underbrace{0.96}_{k=168})^\top}_{\text{obj.} = 8.32 \times 10^7 \text{ (SP)}} \quad \text{vs.} \quad \underbrace{w = (\underbrace{0.22}_{k=1}, \dots, \underbrace{0.77}_{k=168})^\top}_{\clubsuit \text{ obj.} = 6.25 \times 10^7 \text{ (MIP)}}$$

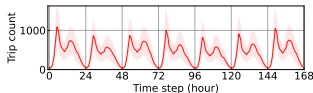
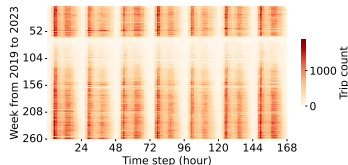


# John F. Kennedy International Airport

- Daily & weekly periodicity: **dropoff** > **pickup** trips at JFK Airport
  - Pickup trips are relevant to flight delay, baggage claim, and other factors.
  - Dropoff trips to airport are highly related to flight schedules.



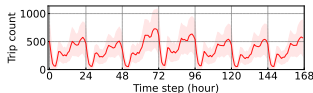
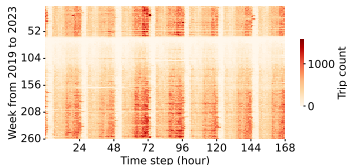
Pickup trips from JFK Airport



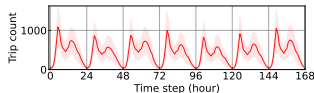
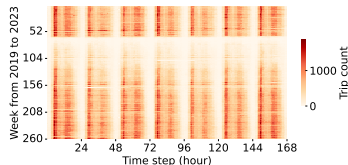
Dropoff trips to JFK Airport

# John F. Kennedy International Airport

- Daily & weekly periodicity: **dropoff** > **pickup** trips at JFK Airport
  - Pickup trips are relevant to flight delay, baggage claim, and other factors.
  - Dropoff trips to airport are highly related to flight schedules.



Pickup trips from JFK Airport



Dropoff trips to JFK Airport

- Sparse coefficient vectors (sparsity  $\tau = 3$ ):

$$\mathbf{w} = (\underbrace{0.31}_{k=1}, \dots, \underbrace{0.28}_{k=24}, \dots, \underbrace{0.41}_{k=168})^\top \quad \text{vs.} \quad \mathbf{w} = (\underbrace{0.18}_{k=1}, \dots, \underbrace{0.35}_{k=24}, \dots, \underbrace{0.47}_{k=168})^\top$$

## High-Dimensional Sparse Autoregression

- On high-dimensional time series with a large  $N$ :

$$\begin{aligned} & \min_{\underbrace{\{\mathbf{w}_n\}_{n=1}^N, \boldsymbol{\beta}}_{(N+1)d \text{ decision var.}}} \underbrace{\sum_{n=1}^N \|\tilde{\mathbf{x}}_n - \mathbf{A}_n \mathbf{w}_n\|_2^2}_{\text{multivariate time series}} \\ & \text{s.t.} \quad \underbrace{0 \leq \mathbf{w}_n \leq \boldsymbol{\beta}}_{\text{bounds being either 0 or 1}}, \quad \underbrace{\|\boldsymbol{\beta}\|_1 \leq \tau}_{\text{sparsity of } \boldsymbol{\beta}}, \quad \boldsymbol{\beta} \in \{0, 1\}^d \end{aligned}$$

- How to handle **millions of** time series (e.g.,  $N \geq 10^6$ )?

# High-Dimensional Sparse Autoregression

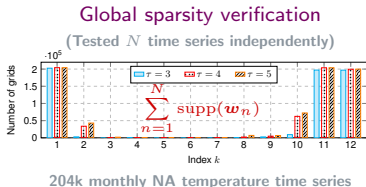
- On high-dimensional time series with a large  $N$ :

$$\begin{aligned}
 & \min_{\underbrace{\{\mathbf{w}_n\}_{n=1}^N, \beta}_{(N+1)d \text{ decision var.}}} \underbrace{\sum_{n=1}^N \|\tilde{\mathbf{x}}_n - \mathbf{A}_n \mathbf{w}_n\|_2^2}_{\text{multivariate time series}} \\
 & \text{s.t.} \quad \underbrace{0 \leq \mathbf{w}_n \leq \beta}_{\text{bounds being either 0 or 1}}, \quad \underbrace{\|\beta\|_1 \leq \tau}_{\text{sparsity of } \beta}, \quad \beta \in \{0, 1\}^d
 \end{aligned}$$

- How to handle **millions of** time series (e.g.,  $N \geq 10^6$ )?
- Two-stage optimization (♣):

- Learn sparsity patterns in  $\beta \in \{0, 1\}^d$

$$\begin{aligned}
 & \min_{\mathbf{w}, \beta} \underbrace{\text{tr}(\mathbf{w} \mathbf{w}^\top \mathbf{P})}_{\text{quadratic}} - \underbrace{2 \mathbf{w}^\top \mathbf{q}}_{\text{linear}} \\
 & \text{s.t. } 0 \leq \mathbf{w} \leq \beta, \quad \|\beta\|_1 \leq \tau
 \end{aligned}$$

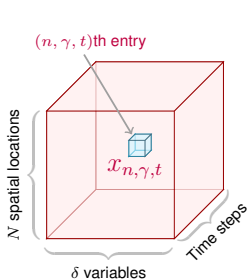


- Quadratic programming with index set  $\Omega = \text{supp}(\beta)$

$$\mathbf{w}_n := \arg \min_{\mathbf{w} \geq 0} \|\tilde{\mathbf{x}}_n - \mathbf{A}_n \mathbf{w}\|_2^2 \quad \text{s.t. } w_k = 0, \forall k \notin \Omega$$

# Human Mobility Periodicity

- Human mobility data



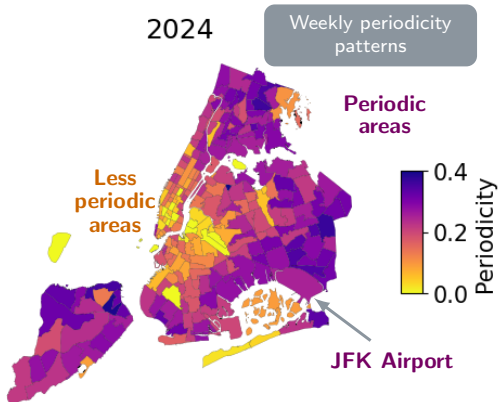
$$\begin{aligned}
 & \min_{\{w_{n,\gamma}\}, \beta} \sum_{n=1}^N \sum_{\gamma=1}^{\delta} \|\tilde{x}_{n,\gamma} - A_{n,\gamma} w_{n,\gamma}\|_2^2 \\
 & \text{s.t. } 0 \leq w_{n,\gamma} \leq \beta \\
 & \quad \|\beta\|_1 \leq \tau \quad \leftarrow \text{sparse} \\
 & \quad \beta \in \{0, 1\}^d \quad \leftarrow \text{binary}
 \end{aligned}$$

spatial location variable  
 hourly trip data

- Quantify **weekly periodicity** by  $\{w_{n,\gamma,k}\}$  at index  $k = 168$

$$\begin{array}{ccc}
 \text{Trip data at } t & & \text{Coefficients} \quad \text{Trip data at } t - k \\
 \begin{array}{c} N \text{ locations} \\ \left\{ \begin{array}{|c|} \hline \text{3x3 grid of red circles} \\ \hline \end{array} \right. \\ \delta \text{ variables} \end{array} & \approx \sum_{k \in [d]} & \begin{array}{c} \left\{ \begin{array}{|c|} \hline \text{3x3 grid of blue circles} \\ \hline \end{array} \right\} \times \begin{array}{c} \left\{ \begin{array}{|c|} \hline \text{3x3 grid of pink circles} \\ \hline \end{array} \right\} \end{array}
 \end{array}$$

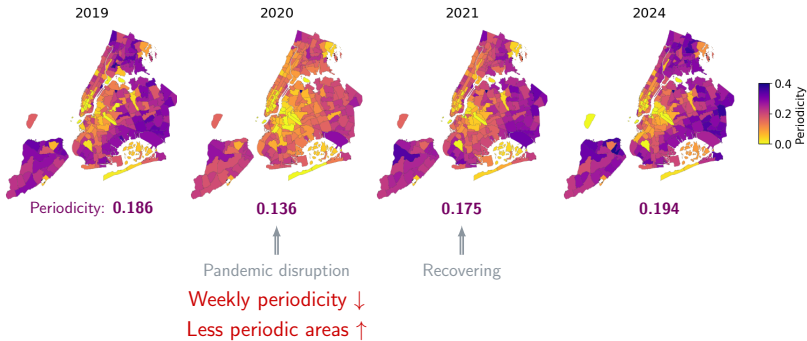
# NYC Ridehailing



Hourly trip time series of 265 pickup areas

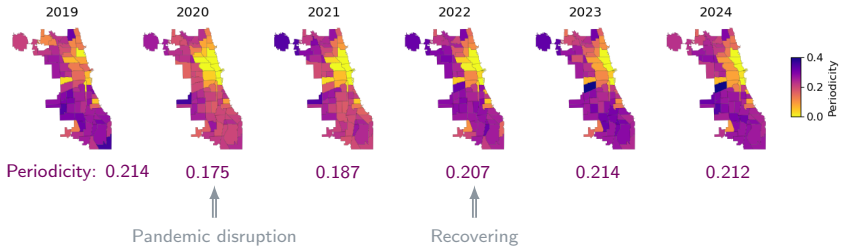
# NYC Ridehailing

## Weekly periodicity patterns



# Chicago Ridehailing

## Weekly periodicity patterns



Weekly periodicity ↓  
Less periodic areas ↑

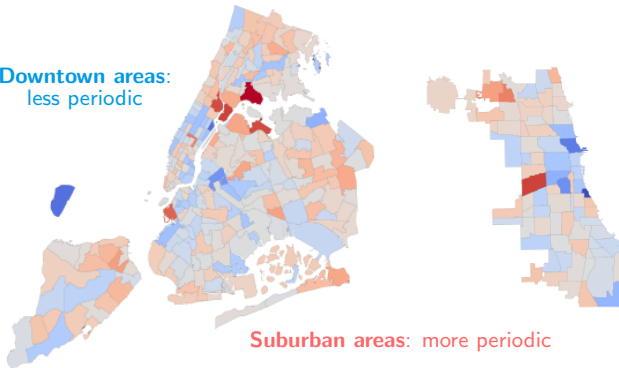
Hourly trip time series of 77 pickup areas



# Post-Pandemic Recovery

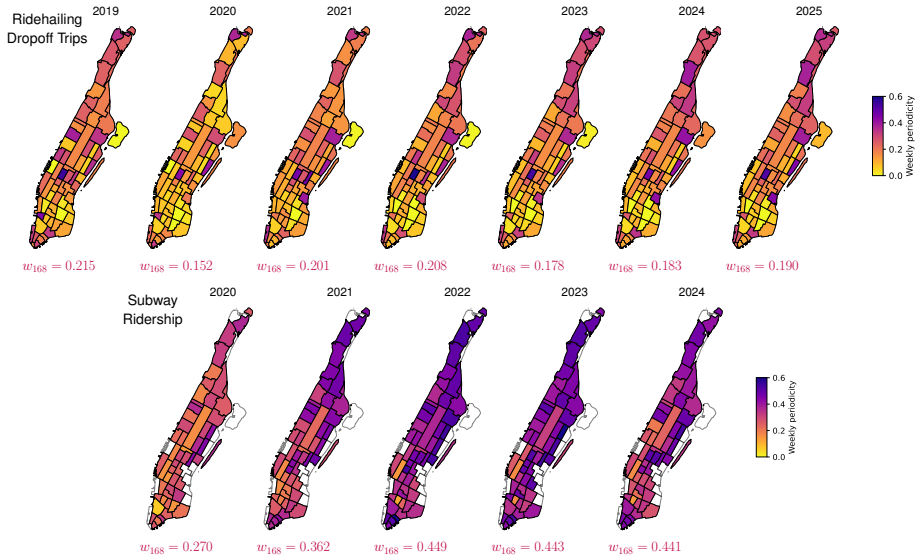
2024's periodicity **minus** 2019's periodicity

**Downtown areas:**  
less periodic



**Suburban areas:** more periodic

# Post-Pandemic Recovery



# Weekly Periodicity of Manhattan Mobility

Mobility data of 2024

**North areas:**

Ridehailing > Yellow taxi

**Overall:**

Subway > Others

Ridehailing

Yellow taxi

Subway

Bikesharing  
(member)

Bikesharing  
(all)

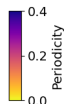
Periodicity: 0.122

0.120

0.363

0.191

0.151



**South areas:**

Yellow taxi > Ridehailing

**Remarkable** weekly periodicity difference  
between membership and all travels

# Daily Periodicity of Manhattan Mobility (Weekdays)

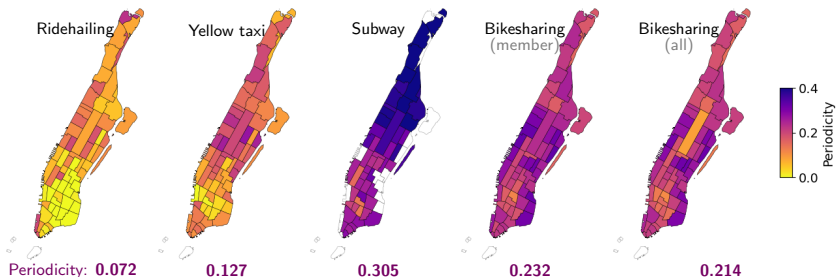
Mobility data of 2024

**Overall:**

Yellow taxi > Ridehailing

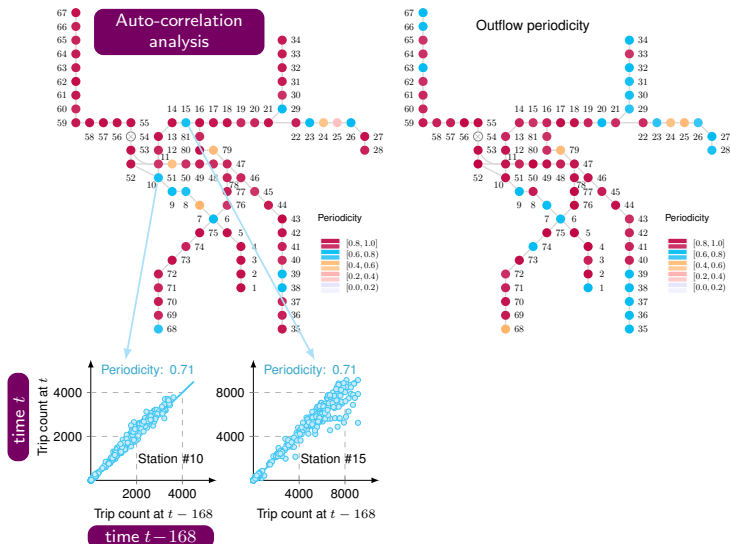
**Subway:**

North areas > South areas



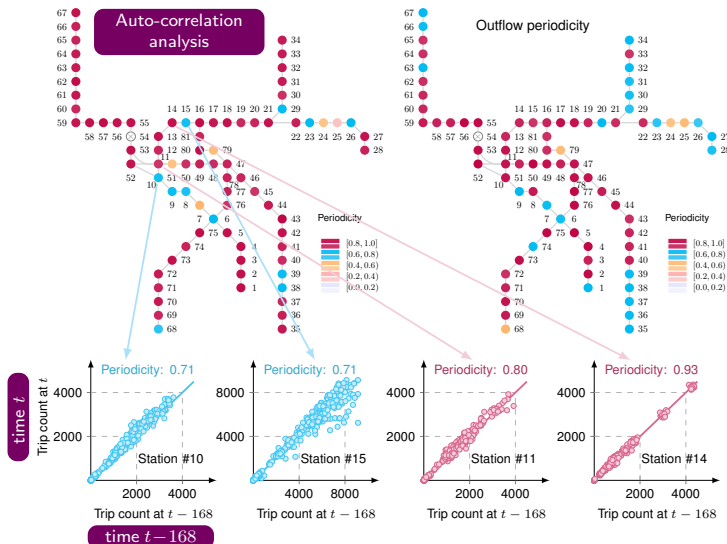
**Marginal** daily periodicity difference  
between membership and all travels

# Hangzhou Metro Passenger Flow



"Closeness" to the diagonal curve  $y = x$

# Hangzhou Metro Passenger Flow

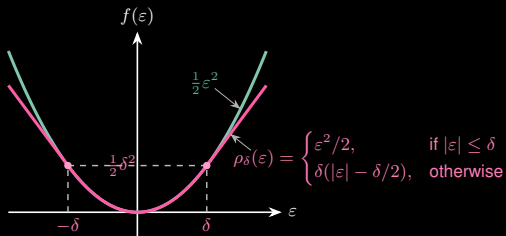


"Closeness" to the diagonal curve  $y = x$

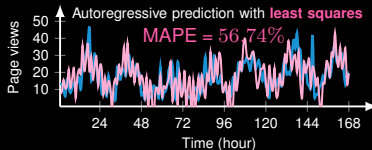
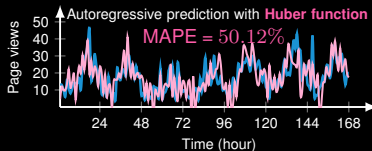
# How Does Huber Regression Work?

Optimization:  $\min_{\mathbf{w}} \underbrace{\frac{1}{2} \sum_{i=1}^n (y_i - \mathbf{w}^\top \mathbf{x}_i)^2}_{\text{Least squares}} \xrightarrow{\text{Reformulate}} \min_{\mathbf{w}} \sum_{i=1}^n \underbrace{\rho_\delta(y_i - \mathbf{w}^\top \mathbf{x}_i)}_{\text{Huber loss function}}$

- Decision variables  $\mathbf{w} \in \mathbb{R}^d$ : Coefficients
- Data  $y_i \in \mathbb{R}$ ,  $i = 1, \dots, n$ : Target values
- Data  $\mathbf{x}_i \in \mathbb{R}^d$ ,  $i = 1, \dots, n$ : Design vectors



Wikipedia page view time series prediction  
("Bayesian\_network")



# Huber Sparse Autoregression

- In contrast to  $\ell_2$ -norm, Huber loss function  $\rho_\delta(\varepsilon_t)$  is more robust:

$$\begin{aligned} \min_{w, \beta, \alpha, \gamma} \quad & \sum_{t=d+1}^T \underbrace{\gamma_t^2}_{\text{quadratic region}} + \underbrace{2\delta\alpha_t}_{\text{linear region}} \\ \text{s.t.} \quad & \varepsilon_t = x_t - \sum_{k=1}^d w_k x_{t-k}, \\ & -\alpha_t \leq \varepsilon_t - \gamma_t \leq \alpha_t, \quad \alpha_t \geq 0 \\ & -\beta_k \leq w_k \leq \beta_k, \\ & \sum_{k=1}^d \beta_k \leq \tau, \quad \beta_k \in \{0, 1\} \end{aligned}$$

with slack vectors  $\alpha, \gamma$  (modeling residual  $\varepsilon_t$ ).

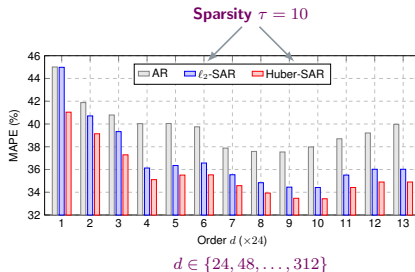
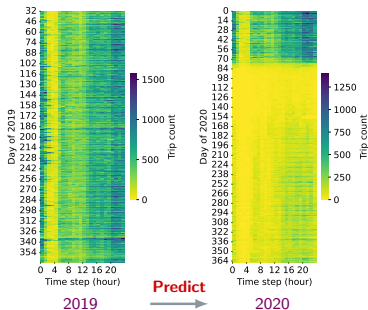
- The optimization is a Mixed-Integer Quadratic Programming (**MIQP**)



# Huber Sparse Autoregression

How does Huber loss help?

- Predict pickup trips at JFK Airport (one-step ahead)

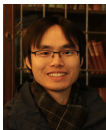


# Speed Field Reconstruction

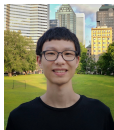
Laplacian Convolutional Representation

(IEEE TKDE'24)

- Vehicular traffic flow
- Speed field reconstruction
- Laplacian regularization
- Fast Fourier transform



**Xinyu Chen**  
PolyMtl → UCF



**Zhanhong Cheng**  
McGill → ZJU



**HanQin Cai**  
UCF

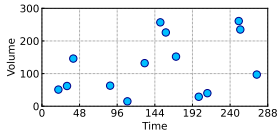


**Nicolas Saunier**  
PolyMtl

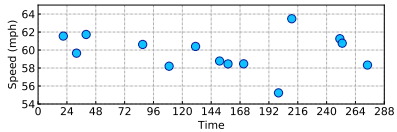
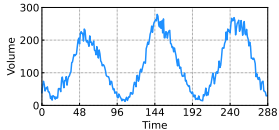


**Lijun Sun**  
McGill

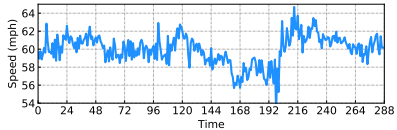
# Motivation



Reconstruct  
traffic volume?

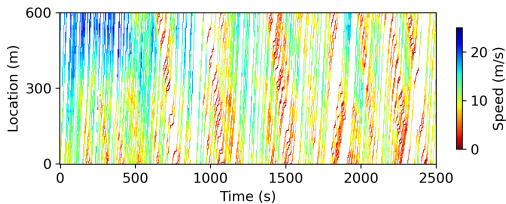


Reconstruct  
traffic speed?

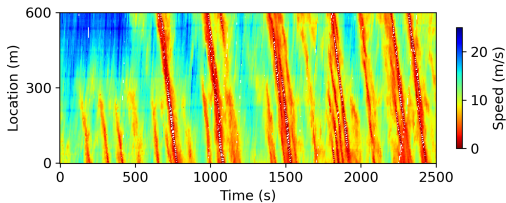


- How to utilize the global trends of traffic time series?
- How to produce local consistency of traffic data?

# Motivation



200-by-500 matrix  
(NGSIM)  $\Downarrow$  Reconstruct speed field from  
20% sparse trajectories?

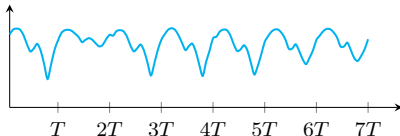


- How to learn from sparse spatiotemporal data?
- How to characterize spatial/temporal local dependencies?

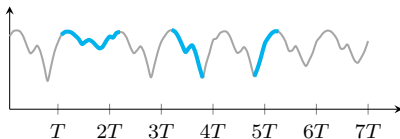
# Motivation

## Sparse time series imputation

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



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

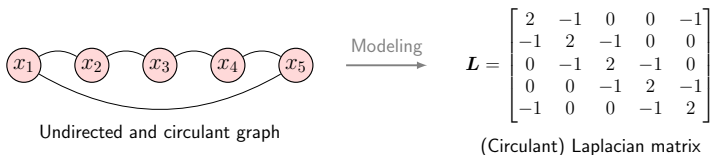


**Modeling global & local trends simultaneously?**

## Circular Convolution

Reformulate Laplacian regularization with circular convolution.

- Intuition of (circulant) Laplacian matrix.



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

$$Lx = \begin{bmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ -1 & 0 & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 0 \\ 0 \\ -1 \end{bmatrix} \star \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \ell \star x$$

where  $\star$  denotes the circular convolution.

- Local trend modeling via (Laplacian) temporal regularization:

$$\mathcal{R}(x) = \frac{1}{2} \|Lx\|_2^2 = \frac{1}{2} \|\ell \star x\|_2^2$$

# Laplacian Regularization

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

---

<sup>1</sup>It refers to the Convolution theorem.

# Laplacian Regularization

- Define Laplacian kernel:

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

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

- Local trend modeling via (Laplacian) temporal regularization:

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

- Property with discrete Fourier transform (denoted by  $\mathcal{F}(\cdot)$ )<sup>1</sup>:

$$\mathcal{R}_\tau(\boldsymbol{x}) = \frac{1}{2} \|\boldsymbol{\ell} \star \boldsymbol{x}\|_2^2 = \underbrace{\frac{1}{2T} \|\mathcal{F}(\boldsymbol{\ell}) \circ \mathcal{F}(\boldsymbol{x})\|_2^2}_{\text{w/ FFT in } \mathcal{O}(T \log T) \text{ time}}$$

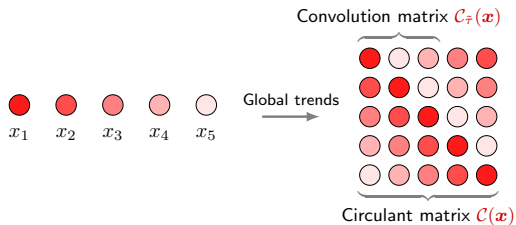
---

<sup>1</sup>It refers to the Convolution theorem.



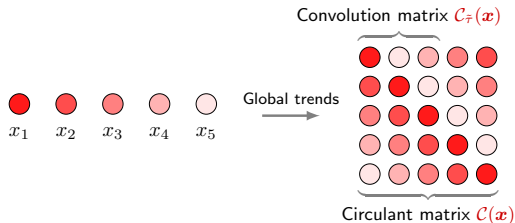
# Global Trend Modeling

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



# Global Trend Modeling

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



- Circulant/Convolution nuclear norm minimization (w/  $\|\mathcal{C}(\mathbf{x})\|_* = \|\mathcal{F}(\mathbf{x})\|_1$ )
  - A balance between global and local trends modeling?

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

Estimating  $\mathbf{x}$ :

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

on data  $\mathbf{y}$  w/ observed index set  $\Omega$ .

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

Estimating  $\mathbf{x}$ :

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

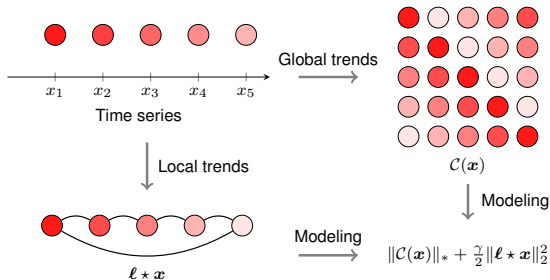
on data  $\mathbf{y}$  w/ observed index set  $\Omega$ .

# Global + Local Trends?

## Laplacian Convolutional Representation (LCR)

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

$$\begin{aligned} \min_{\mathbf{x}} \quad & \underbrace{\|\mathcal{C}(\mathbf{x})\|_*}_{\text{global}} + \frac{\gamma}{2} \underbrace{\|\ell \star \mathbf{x}\|_2^2}_{\text{local}} \\ \text{s.t.} \quad & \|\mathcal{P}_{\Omega}(\mathbf{x} - \mathbf{y})\|_2 \leq \epsilon \quad (\text{data-fidelity constraint}) \end{aligned}$$



# Laplacian Convolutional Representation

- LCR model:

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

$$\begin{aligned} \Rightarrow \quad & \min_{\mathbf{x}, \mathbf{z}} \quad \underbrace{\|\mathcal{C}(\mathbf{x})\|_* + \frac{\gamma}{2} \|\ell \star \mathbf{x}\|_2^2}_{\text{global + local modeling}} + \underbrace{\frac{\eta}{2} \|\mathcal{P}_\Omega(\mathbf{z} - \mathbf{y})\|_2^2}_{\text{constraint to regularization}} \\ & \text{s.t. } \mathbf{z} = \mathbf{x} \end{aligned}$$

“The alternating direction method of multipliers (ADMM) is an algorithm that solves convex optimization problems by breaking them into smaller pieces, each of which are then easier to handle.”

— Source: <https://stanford.edu/~boyd/admm.html>

# Laplacian Convolutional Representation

- LCR model:

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{z}} \quad & \|\mathcal{C}(\mathbf{x})\|_* + \frac{\gamma}{2} \|\ell \star \mathbf{x}\|_2^2 + \frac{\eta}{2} \|\mathcal{P}_\Omega(\mathbf{x} - \mathbf{y})\|_2^2 \\ \text{s.t.} \quad & \mathbf{z} = \mathbf{x} \end{aligned}$$

- Augmented Lagrangian function<sup>2</sup>:

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

- The ADMM scheme:

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

---

<sup>2</sup>w/ Lagrange multiplier  $\mathbf{w} \in \mathbb{R}^T$  and inner product  $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y}$ .

# Laplacian Convolutional Representation

- LCR model:

$$\min_{\mathbf{x}, \mathbf{z}} \|\mathcal{C}(\mathbf{x})\|_* + \frac{\gamma}{2} \|\ell \star \mathbf{x}\|_2^2 + \frac{\eta}{2} \|\mathcal{P}_\Omega(\mathbf{x} - \mathbf{y})\|_2^2$$

$$\text{s.t. } \mathbf{z} = \mathbf{x}$$

- Augmented Lagrangian function<sup>2</sup>:

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

- The ADMM scheme:

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

- Optimize  $\mathbf{x}$ ?

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

---

<sup>2</sup>w/ Lagrange multiplier  $\mathbf{w} \in \mathbb{R}^T$  and inner product  $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y}$ .

# Laplacian Convolutional Representation

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

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

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

# Laplacian Convolutional Representation

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

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

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

## $\ell_1$ -norm Minimization in Complex Space (Liu & Zhang'23)

For any optimization problem in the form of  $\ell_1$ -norm minimization in complex space:

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

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

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

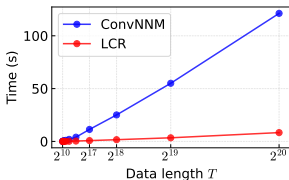


# Laplacian Convolutional Representation

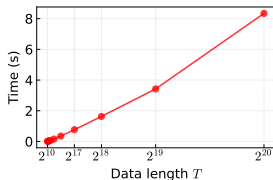
## Empirical time complexity

On the synthetic data  $\mathbf{y} \in \mathbb{R}^T$  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  $\mathcal{C}_{\tilde{\tau}}(\mathbf{y}) \in \mathbb{R}^{T \times \tilde{\tau}}$  with kernel size  $\tilde{\tau} = 2^4$
  - Singular value thresholding in  $\mathcal{O}(\tilde{\tau}^2 T)$

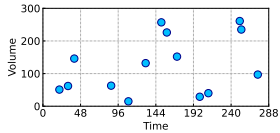


ConvNNM vs. LCR

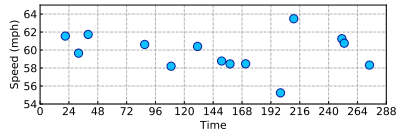
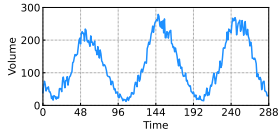


LCR

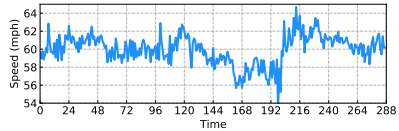
# Experiments



⇓ Reconstruct  
traffic volume?



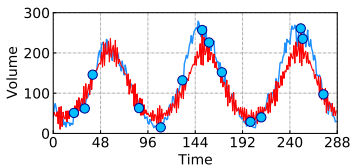
⇓ Reconstruct  
traffic speed?



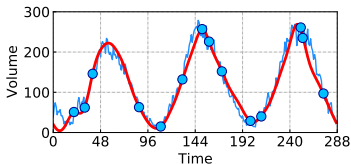
- How to utilize the global trends of traffic time series?
- How to produce local consistency of traffic data?

# Experiments

- Substantial performance gains?



Plus **local** time series trends



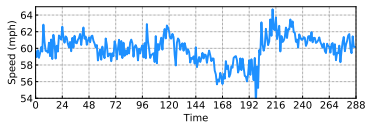
**CircNNM:**

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

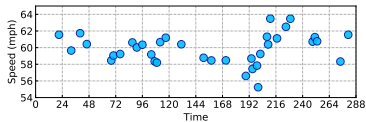
**LCR:**

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

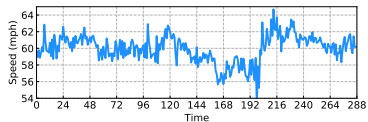
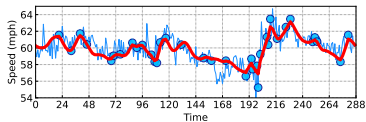
# Experiments



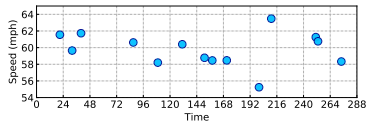
⇓ Mask 90% observations



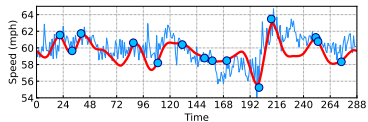
⇓ Reconstruct time series



⇓ Mask 95% observations

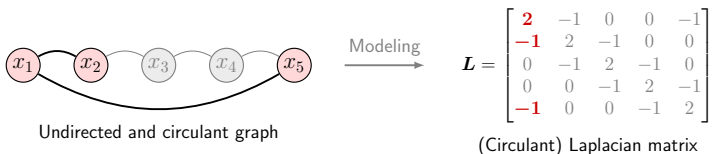


⇓ Reconstruct time series



## Experiments

- The start data points and end data points are connected?



- Flipping operation on  $\mathbf{x} \in \mathbb{R}^5$ :

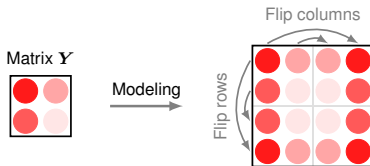
$$\mathbf{x}_{\text{new}} = \begin{bmatrix} \mathbf{x} \\ \mathbf{Jx} \end{bmatrix} = \underbrace{(\mathbf{x}_1, x_2, x_3, x_4, \mathbf{x}_5)}_{\text{original time series}}, \underbrace{(\mathbf{x}_5, x_4, x_3, x_2, \mathbf{x}_1)}_{\text{flipped time series}}^\top \in \mathbb{R}^{10}$$

where  $\mathbf{J} \in \mathbb{R}^{5 \times 5}$  is the exchange matrix.

# Experiments

## Speed field reconstruction<sup>3</sup>

- Flipping operation on a matrix:



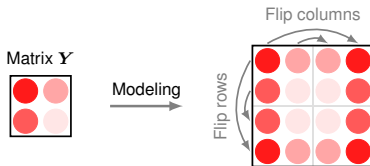
---

<sup>3</sup>Highway Drone (HighD) dataset at <https://www.highd-dataset.com/>

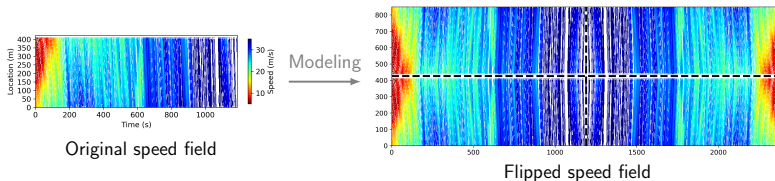
# Experiments

## Speed field reconstruction<sup>3</sup>

- Flipping operation on a matrix:



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



<sup>3</sup>Highway Drone (HighD) dataset at <https://www.highd-dataset.com/>

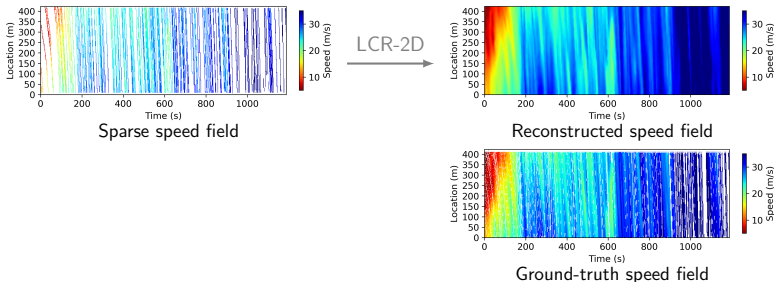
# Experiments

## Speed field reconstruction<sup>4</sup>

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

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

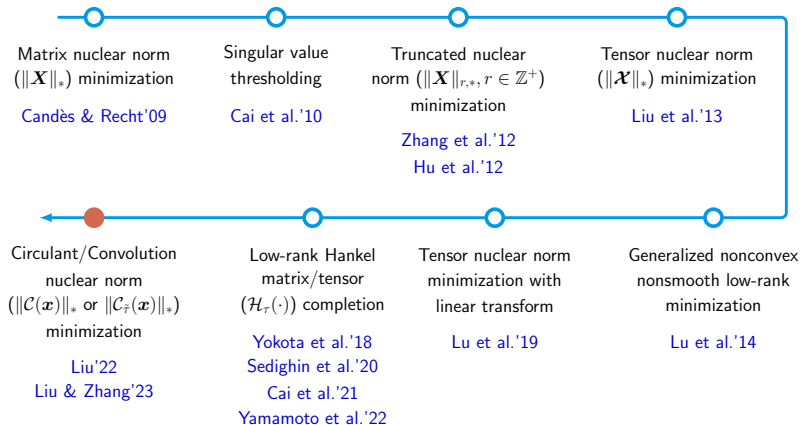
s.t.  $\|\mathcal{P}_\Omega(\mathbf{X} - \mathbf{Y})\|_F \leq \epsilon$



<sup>4</sup>Highway Drone (HighD) dataset at <https://www.highd-dataset.com/>



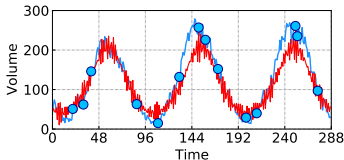
# Contributions



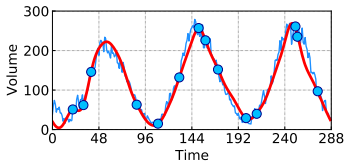
(Ours) **LCR**:

- ✓ Local trend modeling
- ✓ An FFT implementation

# Vision & Insight



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**:

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

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

## Concluding Remarks

- **(Starting point)** How to impute traffic time series?
  - ✓ Low-rank models    ✓ Temporal regularization
- **(Solution)** Time series trend modeling in the low-rank framework?

- Global time series trend modeling (low-rank model):

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

- Local time series trend modeling (temporal regularization):

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

- **(Highlight)** A unified framework with the FFT implementation.

# Thanks for your attention!

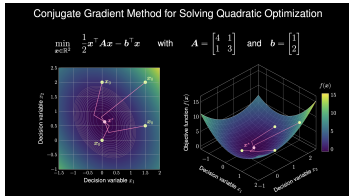
## Any Questions?

Slides: <https://xinychen.github.io/slides/vandy26.pdf>

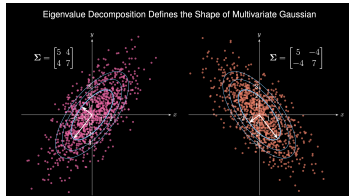
### About me:

- 🏠 Homepage: <https://xinychen.github.io>
- 🐙 GitHub: <https://github.com/xinychen>
- 📺 YouTube: <https://www.youtube.com/@data-worker>
- ✉️ How to reach me: [chenxy346@gmail.com](mailto:chenxy346@gmail.com)

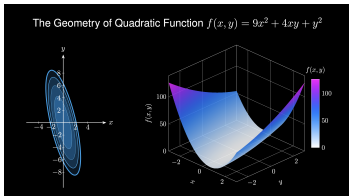
# Matrix Computations and Optimization



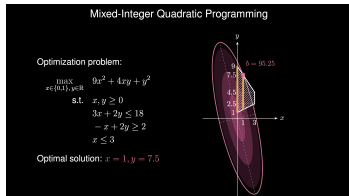
Conjugate Gradient



Covariance Matrix



Quadratic Function



Mixed-Integer Quadratic Programming

<https://www.youtube.com/@data-worker>