Matrix and Tensor Models for Spatiotemporal Traffic Data Imputation and Forecasting

Xinyu Chen

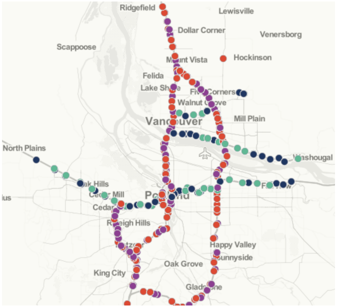
Ph.D., University of Montreal, Canada

December 28, 2023

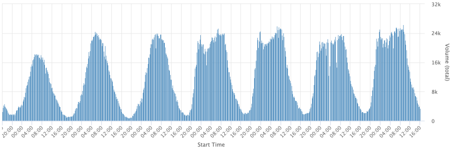
Outline

Spatiotemporal Traffic Data

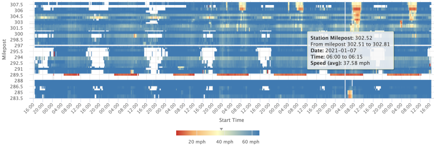
- Many spatiotemporal tra c time series data are in the form of **matrix**.
  - Portland highway tra c data[1]



Highway network & sensor locations



Tra c volume



Tra c speed

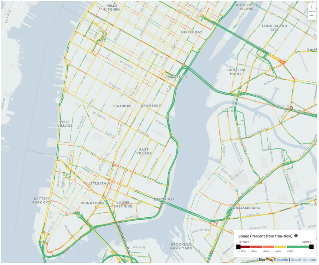
- $\boldsymbol{X} \in \mathbb{R}^{N \times T}$ with N spatial locations T time steps
- Tra c volume/speed shows strong spatial/temporal dependencies

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

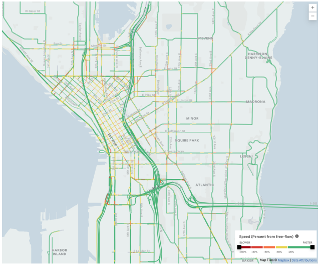
Urban Movement Data

High-dimensional & sparse

Uber (hourly) movement speed data



NYC movement



Seattle movement

*f*road segment, time slot (hour), average speed*g*

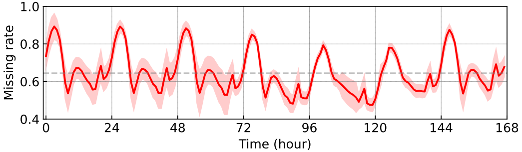
Computing hourly speed: Road segments have 5+ unique trips.

Issue: Insu cient sampling of ridesharing vehicles on the road network!

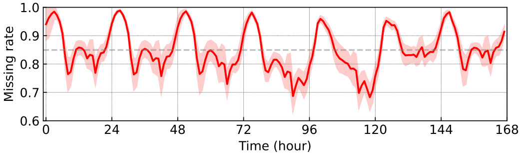
Urban Movement Data

High-dimensional & sparse

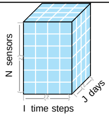
- **NYC** movement speed data (2019)
  - **98,210** road segments & 8,760 time steps (hours)
  - Overall missing rate: **64.43%**



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



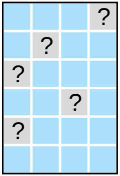
Spatiotemporal Tra c Data

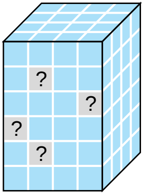
Sparse	
High-dimensional	
Multidimensional	
Noises & outliers	
Nonstationary	

[redacted]

Problem Formulation

Objective A: Impute missing values in the data matrix $Y \in \mathbb{R}^{N\times T}$ (or tensor $\mathcal{Y} \in \mathbb{R}^{M\times N\times T}$).





- Matrix completion (Observed index set Ω)

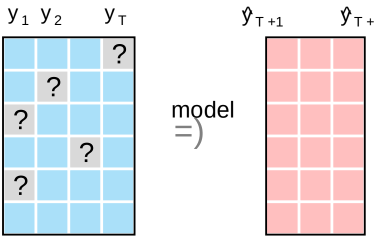
$$\underbrace{\mathcal{P}_{\Omega}(Y)}_{\text{Partially observed}} \xrightarrow{\text{Estimate}} \underbrace{\mathcal{P}_{\Omega^{\perp}}(Y)}_{\text{Unobserved}}$$

Modeling process:

- How to make use of spatiotemporal traffic patterns?
- How to make use of traffic time series dynamics?

Problem Formulation

Objective B: Given a partially observed data $Y \in \mathbb{R}^{N \times T}$ consisting of time series $y_1, \ldots, y_T \in \mathbb{R}^N$, forecast data points $y_{T+\tau}$, $\tau \in \mathbb{N}^+$.



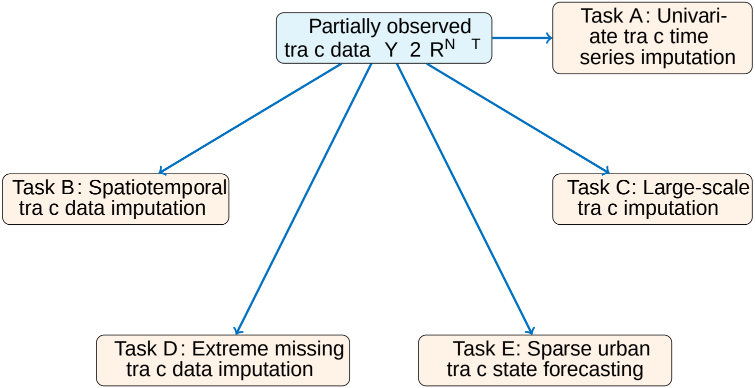
Sparse time series $\quad$ τ-step ahead forecasts

Modeling process:

- How to characterize time series dynamics in high-dimensional and sparse traffic data?

Tasks

We are working on spatiotemporal traffic data imputation and forecasting.



Tensor Factorization

Revisit tensor factorization

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

CP tensor factorization : Factorize $\mathcal{Y}$ into the combination of three rank-R factor matrices (i.e., low-dimensional latent factors).

$$\begin{cases} y_{i,j,t} \approx \sum_{r=1}^{R} u_{i,r} v_{j,r} x_{t,r} \\ \mathcal{Y} \approx \sum_{r=1}^{R} u_r \circ v_r \circ x_r \end{cases}$$

Matrix/Tensor Completion

Matrix nuclear norm ($\|X\|_*$) minimization
Candes & Recht'09

Singular value thresholding
Cai et al.'10

Truncated nuclear norm ($\|X\|_{r,*}, r \in \mathbb{N}^+$) minimization
Zhang et al.'12
Hu et al.'12

Tensor nuclear norm ($\|\mathcal{X}\|_*$) minimization
Liu et al.'13

Generalized nonconvex nonsmooth low-rank minimization
Lu et al.'14

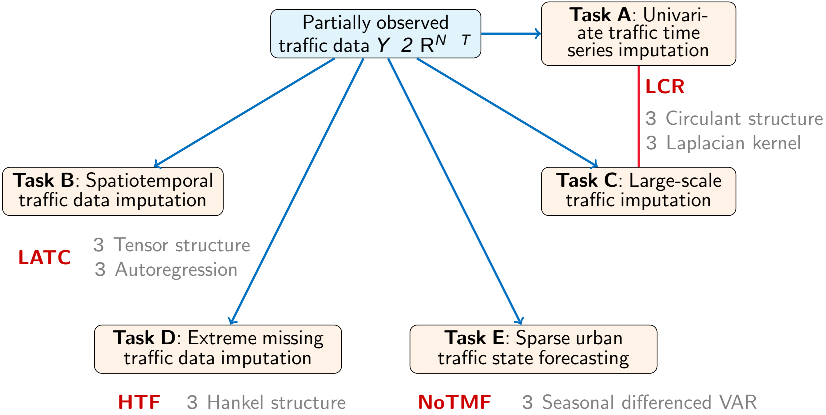
Tensor nuclear norm minimization with linear transform
Lu et al.'19

Low-rank Hankel matrix/tensor ($\mathcal{H}(\cdot)$) completion
Yokota et al.'18
Sedighin et al.'20
Cai et al.'21
Yamamoto et al.'22

Circulant/Convolution nuclear norm ($\|\mathcal{C}(X)\|_*$ or $\|\mathcal{C}_\ell(X)\|_*$) minimization
Liu'22
Liu & Zhang'23

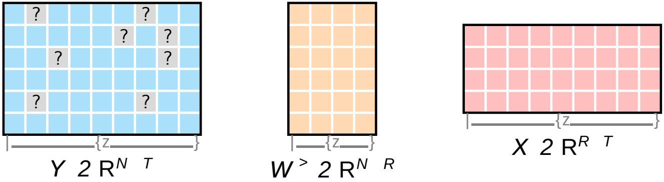
Overview

Matrix/Tensor methods + temporal modeling (e.g., smoothing & autoregression)



Matrix Factorization

A simple approach to reconstruct missing values.



MF (Koren et al.'09)

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

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

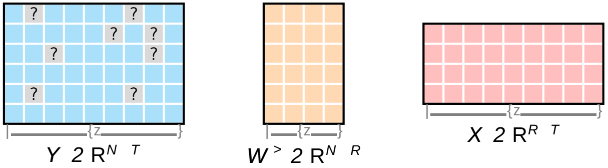
on data $\boldsymbol{Y}$ w/ observed index set Ω.

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

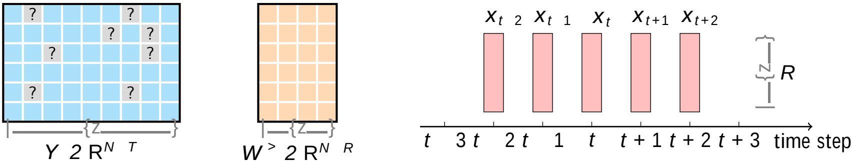
How to build temporal correlations on MF?

Temporal Matrix Factorization

Vector autoregression (VAR) on the temporal factor matrix.



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



Why? Temporal factor matrix $\boldsymbol{X}\in\mathbb{R}^{R\times T}$ 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.

[redacted]

Estimating low-dimensional W, X:

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

on data Y w/ observed index set Ω.

+

[redacted]

$$x_t = \sum_{k=1}^{d} A_k x_{t-k} + \underbrace{\epsilon_t}_{\sim \mathcal{N}(0, I)}$$

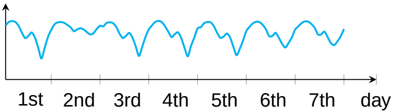
w/ coefficients $\{A_k\}$.

+ Yu et al.'16
Chen & Sun'21

$$\min_{W, X, \{A_k\}_{k=1}^{d}} \frac{1}{2}\underbrace{\|\mathcal{P}_{\Omega}(Y - W^\top X)\|_F^2}_{\text{MF on data } Y} + \frac{\gamma}{2}\underbrace{\sum_{t=d+1}^{T}\Big\|x_t - \sum_{k=1}^{d} A_k x_{t-k}\Big\|_2^2}_{\text{VAR on temporal factors } X}$$

Nonstationary Temporal Matrix Factorization

Traffic data are nonstationary due to daily patterns of traffic flow.



Season-m differencing ($m \in \mathbb{N}^{+}$, e.g., daily/weekly):

$$\boldsymbol{x}_t \approx \sum_{k=1}^{d} \boldsymbol{A}_k \boldsymbol{x}_{t-k} \Rightarrow \boldsymbol{x}_t - \boldsymbol{x}_{t-m} \approx \sum_{k=1}^{d} \boldsymbol{A}_k (\boldsymbol{x}_{t-k} - \boldsymbol{x}_{t-m-k})$$

(Ours) Optimization problem:

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

Nonstationary Temporal Matrix Factorization

Rewrite NoTMF

Optimization problem:[2]

$$\min_{\boldsymbol{W},\boldsymbol{X},\boldsymbol{A}}\ \frac{1}{2}\underbrace{\|\mathcal{P}_{\Omega}(\boldsymbol{Y}-\boldsymbol{W}^\top\boldsymbol{X})\|_F^2}_{\text{MF on data }\boldsymbol{Y}}+\frac{\rho}{2}\underbrace{(\|\boldsymbol{W}\|_F^2+\|\boldsymbol{X}\|_F^2)}_{\text{Regularization}}$$

$$+\frac{\lambda}{2}\underbrace{\|\boldsymbol{X}\boldsymbol{\Psi}_0^\top-\boldsymbol{A}(\boldsymbol{I}_d\otimes\boldsymbol{X})\boldsymbol{\Psi}^\top\|_F^2}_{\text{VAR on }\boldsymbol{X}}$$

where $\boldsymbol{\Psi}_0\in\mathbb{R}^{(T-dm)\times T},\boldsymbol{\Psi}\in\mathbb{R}^{(T-dm)\times(dT)}$ are temporal operators.

Alternating minimization (let f be the obj.):

$$\begin{cases}\text{Spatial factors} & \boldsymbol{W}:=\{\boldsymbol{W}\mid\frac{\partial f}{\partial\boldsymbol{W}}=\boldsymbol{0}\} & \text{(least squares)}\\ \text{Temporal factors} & \boldsymbol{X}:=\{\boldsymbol{X}\mid\frac{\partial f}{\partial\boldsymbol{X}}=\boldsymbol{0}\} & \text{(conjugate gradient)}\\ \text{VAR coefficients} & \boldsymbol{A}:=\{\boldsymbol{A}\mid\frac{\partial f}{\partial\boldsymbol{A}}=\boldsymbol{0}\} & \text{(least squares)}\end{cases}$$

[2] $\boldsymbol{A}\triangleq[\boldsymbol{A}_1,\cdots,\boldsymbol{A}_d]\in\mathbb{R}^{R\times(dR)}$ (coefficient matrix).

Nonstationary Temporal Matrix Factorization

NoTMF forecasting?

Implementation

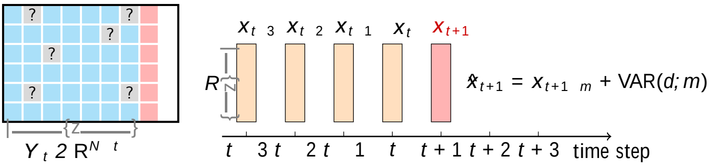
Estimate $\boldsymbol{W}; \boldsymbol{X}; \boldsymbol{A}$

Forecast $\hat{\boldsymbol{x}}_{t+1}$ with VAR

Return $\hat{\boldsymbol{y}}_{t+1} = \boldsymbol{W}^{>} \hat{\boldsymbol{x}}_{t+1}$

3 Sparse input $\boldsymbol{Y}_t$

3 Forecast in latent spaces



Nonstationary Temporal Matrix Factorization

NoTMF forecasting on streaming data?

Online forecasting (Gultekin & Paisley'18):

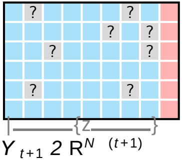
- Fix the spatial factor matrix $\boldsymbol{W}$
- Use input data $\boldsymbol{Y}_{t+1}$ to update the temporal factor matrix $\boldsymbol{X}$ and the coe cient matrix $\boldsymbol{A}$

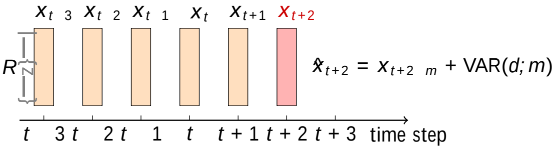
Implementation

- Estimate $\boldsymbol{X}; \boldsymbol{A}$
- Forecast $\hat{\boldsymbol{x}}_{t+2}$ with VAR
- Return $\hat{\boldsymbol{y}}_{t+2} = \boldsymbol{W}^{>} \hat{\boldsymbol{x}}_{t+2}$

3 Sparse input $\boldsymbol{Y}_{t+1}$

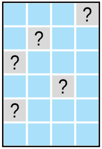
3 Forecast in latent spaces



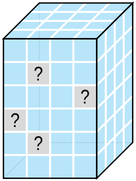


Matrix/Tensor Completion

Problem? Impute missing values in matrices/tensors.



$\mathcal{P}_{\Omega}(\boldsymbol{Y}) \in \mathbb{R}^{N \times T}$



$\mathcal{P}_{\Omega}(\mathcal{Y}) \in \mathbb{R}^{N \times I \times J}$

Cornerstone: Nuclear norm minimization

LRMC (Candes & Recht'09)

Estimating the matrix $\boldsymbol{X}$:

$$\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_{*} \quad \text{s.t. } \mathcal{P}_{\Omega}(\boldsymbol{X}) = \mathcal{P}_{\Omega}(\boldsymbol{Y})$$

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

vs.

LRTC (Liu et al.'13)

Estimating the tensor $\mathcal{X}$:

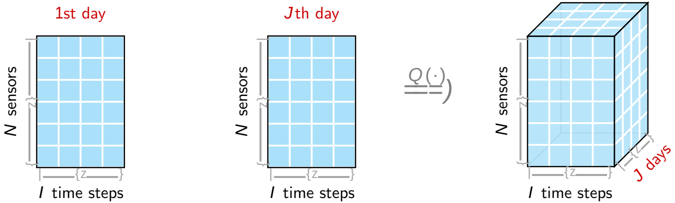
$$\min_{\mathcal{X}} \|\mathcal{X}\|_{*} \quad \text{s.t. } \mathcal{P}_{\Omega}(\mathcal{X}) = \mathcal{P}_{\Omega}(\mathcal{Y})$$

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

Limitation: Nuclear norm minimization only covers global consistency.

Low-Rank Autoregressive Tensor Completion

Introduce traffic tensors with day dimension[3] (Tan et al.'13, Chen et al.'19, ...).



Build temporal correlations with univariate autoregression.

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

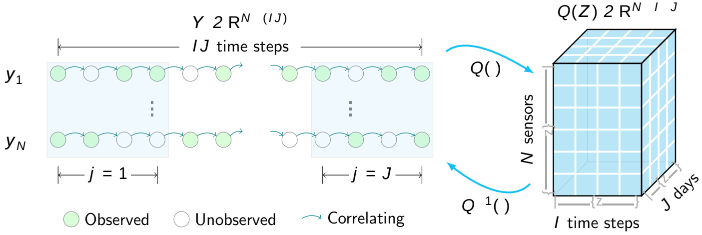
$$\|\boldsymbol{Y}\|_{\boldsymbol{A},\mathcal{H}} \triangleq \sqrt{\sum_{n,t}\left(y_{n,t} - \sum_{k} a_{n,k} y_{n,t-h_k}\right)^2}$$

w/ the time lag set $\mathcal{H} = \{h_1, \ldots, h_d\}$ and the coefficient matrix $\boldsymbol{A} \in \mathbb{R}^{N \times d}$.

[3] There are $T = IJ$ time steps in total.

Low-Rank Autoregressive Tensor Completion

Local consistency w/ autoregression — **Global consistency** w/ tensor structure



LATC

Optimization problem:

$$\min_{\boldsymbol{Z};\boldsymbol{A}} \underbrace{\|Q(\boldsymbol{Z})\|_{r;}}_{\text{global}} + \frac{}{2}\underbrace{\|\boldsymbol{Z}\|_{\boldsymbol{A};H}}_{\text{local}}$$

s.t. $P\ (\boldsymbol{Z}) = P\ (\boldsymbol{Y})$

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

$\Longrightarrow$

Two subproblems

$$\begin{cases} \boldsymbol{Z} := \underset{P\ (\boldsymbol{Z}) = P\ (\boldsymbol{Y})}{\arg\min}\ \|Q(\boldsymbol{Z})\|_{r;} + \frac{}{2}\|\boldsymbol{Z}\|_{\boldsymbol{A};H} \\ \boldsymbol{A} := \frac{1}{2}\|\boldsymbol{Z}\|_{\boldsymbol{A};H} \qquad \text{(Least squares)} \end{cases}$$

Low-Rank Autoregressive Tensor Completion

$\boldsymbol{Z}$-subproblem:

$$\boldsymbol{Z} := \underset{P\ (\boldsymbol{Z})=P\ (\boldsymbol{Y})}{\arg\min}\ \|Q(\boldsymbol{Z})\|_{r,} + \frac{}{2}\|\boldsymbol{Z}\|_{\boldsymbol{A},H}$$

Augmented Lagrangian function:[4]

$$L(X, \boldsymbol{Z}, W) = \|X\|_{r,} + \frac{}{2}\|\boldsymbol{Z}\|_{\boldsymbol{A},H} + \frac{}{2}\|X \quad Q(\boldsymbol{Z})\|_F^2 + \langle W, X \quad Q(\boldsymbol{Z})\rangle + \ (\boldsymbol{Z})$$

Implementation

Repeat

- Compute $\boldsymbol{Z}$
- Compute $\boldsymbol{A}$

$\Rightarrow$

Implementation

Repeat

- Repeat
  - # Alternating Direction Method of Multipliers (ADMM)
  - Compute X
  - Compute $\boldsymbol{Z}$
  - Compute W
- Compute $\boldsymbol{A}$

---

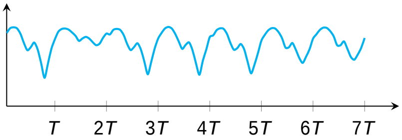
[4] $W \in \mathbb{R}^{N \quad I \quad J}$ (Lagrange multiplier); $\langle\ ,\ \rangle$ (inner product). The indicator function:

$$(\boldsymbol{Z}) = \begin{cases} 0, & \text{if } P\ (\boldsymbol{Z}) = P\ (\boldsymbol{Y}), \\ +\infty, & \text{otherwise.} \end{cases}$$

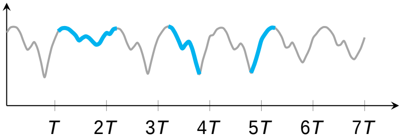
Laplacian Convolutional Representation

Motivation: Time series imputation

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



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

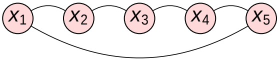


How to characterize both global and local trends in sparse time series?

Laplacian Convolutional Representation

Local trend modeling

Intuition of (circulant) Laplacian matrix



Undirected and circulant graph

$\xrightarrow{\text{Modeling}}$

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

(Circulant) Laplacian matrix

Define Laplacian kernel:

$$\boldsymbol{\ell} \triangleq (2, -1, 0, 0, -1)^\top$$

$$\Downarrow$$

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

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

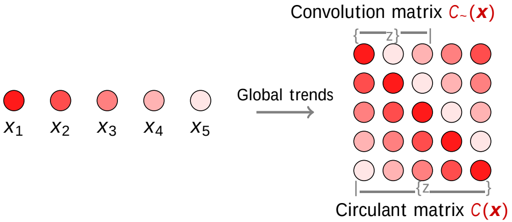
(Laplacian) Temporal regularization:

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

Reformulate temporal regularization with circular convolution.

Laplacian Convolutional Representation

Global trend modeling: Circulant matrix $C(\boldsymbol{x})$ vs. convolution matrix $C_{\sim}(\boldsymbol{x})$



Circulant/Convolution nuclear norm minimization

A balance between global and local trends modeling?

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

Estimating $\boldsymbol{x}$:

$$\min_{\boldsymbol{x}}\ \|C(\boldsymbol{x})\|$$
$$\text{s.t. } \|P\ (\boldsymbol{x}\quad \boldsymbol{y})\|_2$$

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

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

Estimating $\boldsymbol{x}$:

$$\min_{\boldsymbol{x}}\ \|C_{\sim}(\boldsymbol{x})\|$$
$$\text{s.t. } \|P\ (\boldsymbol{x}\quad \boldsymbol{y})\|_2$$

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

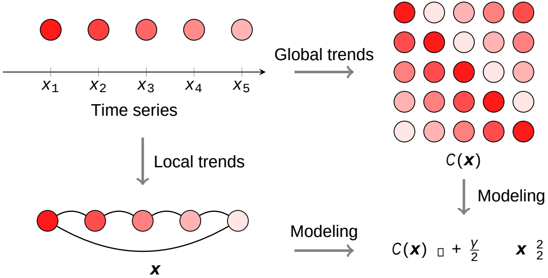
Laplacian Convolutional Representation

Laplacian Convolutional Representation (LCR)

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

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

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



Laplacian Convolutional Representation

Augmented Lagrangian function:[5]

$$L(x;z;w) = kC(x)k \quad + \quad \frac{}{2}k\grave{} \; ? \; x k_2^2 + \frac{}{2}kx \quad z k_2^2 + hw; x \quad z i + \frac{}{2}kP \quad (z \quad y) k_2^2$$

The ADMM scheme:

$$\begin{array}{lll} x := \arg\min_{x} \; L(x;z;w) & & \text{(Nuclear norm minimization)} \\ z := \arg\min_{z} \; L(x;z;w) & & \text{(Closed-form solution)} \\ w := w + \quad (x \quad z) & & \text{(Standard update)} \end{array}$$

Optimize x?

$$kC(x)k \quad = \; kF(x)k_1 \qquad \& \qquad \frac{1}{2}k\grave{} \; ? \; x k_2^2 = \frac{1}{2T}kF(\grave{}) \quad F(x)k_2^2$$

[redacted]

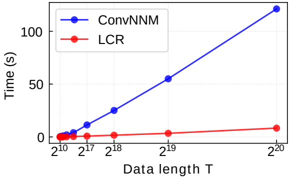
[5] $w \; 2 \; R^T$ (Lagrange multiplier); $h; i$ (inner product).

Laplacian Convolutional Representation

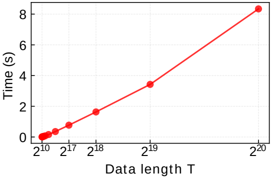
Empirical time complexity

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

- Ours: **LCR**
  - An FFT implementation in $O(T \log T)$
  - The logarithmic factor $\log T$ makes the FFT highly efficient
- Baseline: **ConvNNM**[6] (Liu'22, Liu & Zhang'23)
  - Convolution matrix $\mathcal{C}_{\tau}(\boldsymbol{y}) \in \mathbb{R}^{T \times \tau}$ with kernel size $\tau = 2^4$
  - Singular value thresholding in $O(\tau^2 T)$



ConvNNM vs. LCR



LCR

[6] Convolution nuclear norm minimization.

Hankel Tensor Factorization

Motivation: Spatiotemporal data reconstruction

Sparse speed eld reconstruction problem in vehicular tra c ow.

200-by-500 matrix
(NGSIM)

+

Reconstruct speed eld from
5% sparse trajectories?

[redacted]

Hankel Tensor Factorization

Hankel matrix

- Given $\boldsymbol{x} = (1, 2, 3, 4, 5)^\top$ and window length $\tau = 2$, we have

$$\textcolor{red}{\mathcal{H}_{\tau}(\boldsymbol{x})} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \\ 4 & 5 \end{bmatrix} \in \mathbb{R}^{4 \times 2}$$

- Automatic temporal modeling

Traffic time series

Hankel matrix

Hankel Tensor Factorization

Hankel tensor: Given any matrix $X\ 2\ R^{N\quad T}$, we have

$$X\ ,\ H_{1;\ 2}(X)$$

Window lengths: $_1;\ _2\ 2\ N^+$;
Tensor size: $(N\quad _1 + 1)\quad _1\quad (T\quad _2 + 1)\quad _2$;

(Figure) 4th-order Hankel tensor: A sequence of third-order tensors.

Slice: $X_{:;k_1;:;k_2}$; $8k_1; k_2$;
Slice size: $(N\quad _1 + 1)\quad (T\quad _2 + 1)$.

Hankel Tensor Factorization

Hankel indexing

Sampling function for the Hankel tensor:

$$_{k_1;k_2}(X) \, , \, [H_{\ 1;\ 2}(X)]_{:;k_1;:;k_2};$$

referring to as the tensor slice with k_1 2 f 1; : : : ; $_1$g; k_2 2 f 1; : : : ; $_2$g.

[Importance] Developing memory-e cient algorithms

Tensor slices $_{k_1;k_2}(X)$ vs. data matrix X

Hankel Tensor Factorization

Ours:

Convolutional tensor decomposition (circular convolution $?_{\text{row}}$):

$$_{k_1;k_2}(Y) \quad (Q \, ?_{\text{row}} \, s_{k_1}^{>})(U \, ?_{\text{row}} \, v_{k_2}^{>})^{>}$$

Baselines:

Tensor-train decomposition:

$$_{k_1;k_2}(Y) \quad (QS_{k_1})(UV_{k_2})^{>}$$

- f $S_{k_1}; V_{k_2}$ g are circulant matrices) convolutional decomposition
- f $S_{k_1}; V_{k_2}$ g are diagonal matrices) CP decomposition

CP tensor decomposition (Khatri-Rao product):

$$_{k_1;k_2}(Y) \quad (Q \quad s_{k_1}^{>})(U \quad v_{k_2}^{>})^{>}$$

Hankel Tensor Factorization

HTF (convolutional decomposition)

Optimization problem:

$$\min_{Q,S,U,V} \frac{1}{2} \underbrace{\sum_{k_1,k_2} \left\| \mathcal{P}_{k_1,k_2}(\mathcal{Y}) - (Q \star_{\text{row}} s_{k_1}^\top)(U \star_{\text{row}} v_{k_2}^\top)^\top \right\|_F^2}_{\text{Tensor decomposition on Hankel tensor slices}}$$
$$+ \frac{\lambda}{2}\left(\|Q\|_F^2 + \|S\|_F^2 + \|U\|_F^2 + \|V\|_F^2 \right)$$

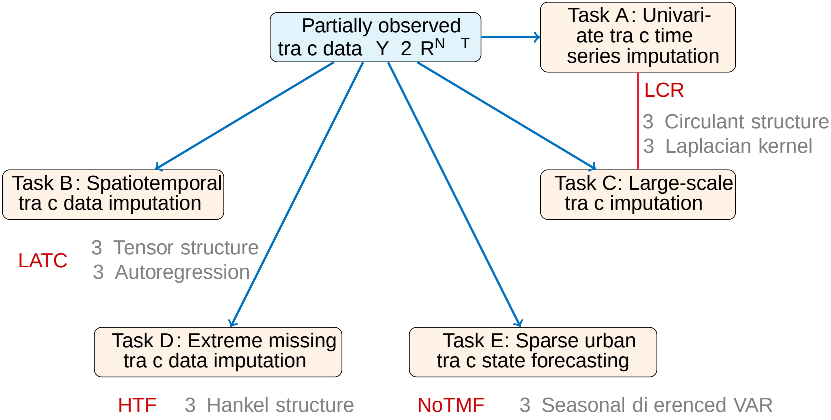
Alternating minimization (let f be the obj.):

$$\begin{cases}
Q := \{Q \mid \frac{\partial f}{\partial Q} = 0\} & \text{(conjugate gradient)} \\
s_{k_1} := \{s_{k_1} \mid \frac{\partial f}{\partial s_{k_1}} = 0\}, \forall k_1 & \text{(conjugate gradient)} \\
U := \{U \mid \frac{\partial f}{\partial U} = 0\} & \text{(conjugate gradient)} \\
v_{k_2} := \{v_{k_2} \mid \frac{\partial f}{\partial v_{k_2}} = 0\}, \forall k_2 & \text{(conjugate gradient)}
\end{cases}$$

Memory-efficient but still computationally costly!

Overview

We are working on spatiotemporal traffic data imputation and forecasting.



Task A: Univariate Traffic Time Series Imputation

CircNNM :

$$\min_{x} \underbrace{\|C(x)\|}_{\text{global}}$$

$$\text{s.t.} \quad \|P(x - y)\|_2$$

Plus temporal regularization

LCR:

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

$$\text{s.t.} \quad \|P(x - y)\|_2$$

Task A: Univariate Tra c Time Series Imputation

Mask 90% observations

Mask 95% observations

Reconstruct time series

Reconstruct time series

[redacted]

Task B: Spatiotemporal Tra c Data Imputation

LATC vs. baseline (in MAPE/RMSE)

On the Seattle freeway tra c speed dataset ($\mathcal{Y} \in \mathbb{R}^{323 \times 8064}$)

Missing rate	LATC	LAMC	LRTC-TNN	BTMF	SPC
30%, Random Missing	4.90/ 3.16	5.98/3.73	4.99/3.20	5.91/3.72	5.92/3.62
70%, Random Missing	5.96/ 3.71	8.02/4.70	6.10/3.77	6.47/3.98	7.38/4.30
90%, Random Missing	7.46/ 4.50	10.56/5.91	8.08/4.80	8.17/4.81	9.75/5.31
30%, Nonrandom Missing	7.10/4.33	6.99/4.25	6.85/ 4.21	9.26/5.36	8.87/4.99
70%, Nonrandom Missing	9.40/5.40	9.75/5.60	9.23/ 5.35	10.47/6.15	11.32/5.92
30%, Block-out Missing	9.43/ 5.36	27.05/13.66	9.52/5.41	14.33/13.60	11.30/5.84

On the Portland highway tra c volume dataset ($\mathcal{Y} \in \mathbb{R}^{1156 \times 2976}$)

Missing rate	LATC	LAMC	LRTC-TNN	BTMF	SPC
30%, Random Missing	16.95/ 15.99	17.93/16.03	17.27/16.08	18.22/19.14	21.29/56.73
70%, Random Missing	19.59/ 18.70	21.26/19.37	19.99/18.73	19.96/22.21	24.35/43.32
90%, Random Missing	23.15/22.83	25.64/23.75	22.90/ 22.68	23.90/25.71	28.45/39.65
30%, Nonrandom Missing	19.48/19.14	19.93/19.69	19.59/ 18.91	19.55/20.38	26.96/60.33
70%, Nonrandom Missing	27.67/45.03	25.75/28.25	30.26/60.85	23.86/ 26.74	33.42/47.34
30%, Block-out Missing	24.01/ 23.50	29.21/27.60	31.74/74.42	27.85/25.68	31.01/60.33

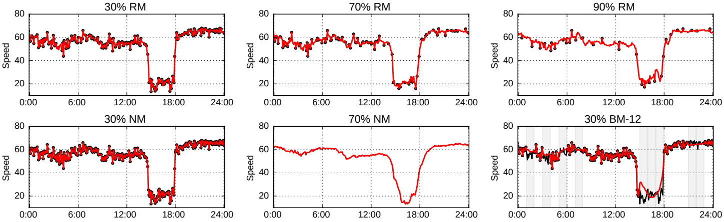
LATC vs. LAMC: The signi cance of tensor representation

LATC vs. LRTC-TNN: The signi cance of temporal autoregression

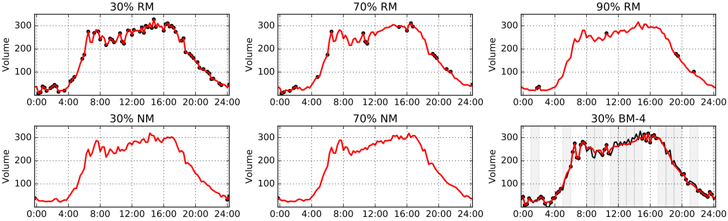
Task B: Spatiotemporal Traffic Data Imputation

LATC imputation

Seattle freeway tra c speed data



Portland highway tra c volume data



Task C: Large-Scale Traffic Data Imputation

LCR vs. baseline (in MAPE/RMSE)

PeMS-4W: California freeway traffic speed dataset ($\boldsymbol{Y} \in \mathbb{R}^{11160 \times 8064}$)

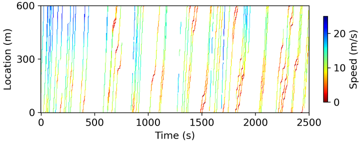
Model	Missing rate			
	30%	50%	70%	90%
LCR-2D	1.50/**1.49**	1.76/**1.69**	**2.07**/**2.06**	**3.19**/**3.05**
LCR_N	**1.48**/1.50	**1.73**/1.73	**2.07**/2.12	3.24/3.22
LCR	1.50/**1.49**	1.76/**1.69**	2.08/2.07	3.21/3.06
CTNNM	2.26/1.84	2.67/2.14	3.40/2.66	5.22/3.90
CircNNM	2.26/1.84	2.69/2.15	3.43/2.67	5.34/3.96
LRMC	2.04/1.80	2.43/2.12	3.08/2.66	6.05/4.43
HaLRTC	1.98/1.73	2.22/1.98	2.84/2.49	4.39/3.66
LRTC-TNN	1.68/1.55	1.93/1.77	2.33/2.14	3.40/3.10
NoTMF	2.95/2.65	3.05/2.73	3.33/2.97	5.22/4.71

Results

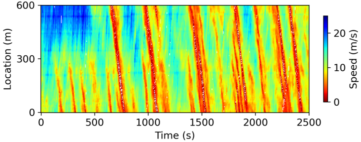
- LCR-2D > CTNNM: The importance of temporal regularization.
- CTNNM CircNNM: Ciculant tensor is superior to circulant matrix.
- LCR > LRMC/LRTC: The importance of global/local modeling.

$$O(NT\log(NT)) \quad \text{(FFT)} \quad \text{vs.} \quad O(\min\{N^2T, NT^2\}) \quad \text{(SVD)}$$

Task D: Extreme Missing Traffic Data Imputation

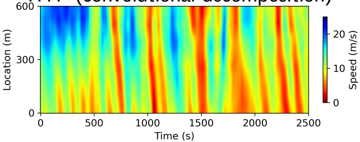


Sparse speed field



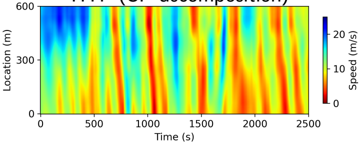
Ground truth speed field

HTF (convolutional decomposition)



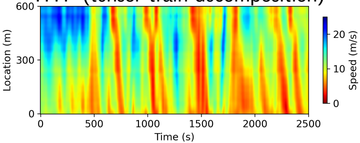
MAPE = **51.92%**

HTF (CP decomposition)



MAPE = 53.93%

HTF (tensor-train decomposition)



MAPE = 56.48%

Task D: Extreme Missing Traffic Data Imputation

HTF vs. baseline (in MAPE/RMSE)

On the Seattle freeway traffic speed dataset ($\boldsymbol{Y} \in \mathbb{R}^{323 \times 8064}$)

Model	Random missing rate			
	80%	85%	90%	95%
HTF (convolution)	**6.21/3.88**	**6.51/4.06**	**6.98/4.30**	**8.02/4.84**
HTF (tensor-train)	8.75/5.16	9.86/5.76	9.24/5.36	9.89/5.70
HTF (CP)	7.25/4.33	7.93/4.66	8.61/4.96	9.25/5.20
LATC	6.50/4.00	6.90/4.21	7.47/4.51	8.75/5.05
LRTC-TNN	6.97/4.24	7.43/4.43	8.19/4.81	9.60/5.55
LCR	6.75/4.15	7.31/4.38	7.96/4.71	9.78/5.39
BTMF	6.85/4.17	7.36/4.42	8.13/4.79	9.63/5.48

Results

- Convolutional tensor decomposition outperforms both tensor-train and CP tensor decomposition.
- Our HTF model performs better than state-of-the-art baseline models.

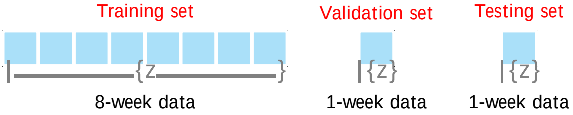
Task E: Sparse Urban Traffic State Forecasting

NoTMF forecasting

NYC Uber movement speed dataset:

10-week data of size 98210 1680; **66.56%** missing values

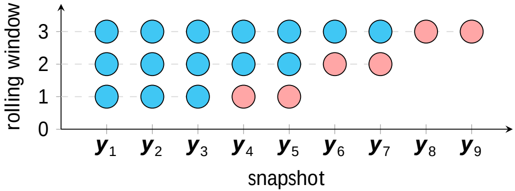
Rolling forecasting setup (Time horizon = 1; 2; 3; 6):



Weight parameter $2\ f10^{\ 2}; 10^{\ 1}; 10^0; 10^1; 10^2 g$

Weight parameter $= f10^{\ 1}\ ; 5\quad 10^{\ 1}\ ;\ \ ; 5\ \ ; 10\ \ g$

Rolling forecasting illustration (= 2):



Task E: Sparse Urban Traffic State Forecasting

NoTMF vs. baseline (in MAPE/RMSE)

On the NYC Uber movement speed dataset

	d	NoTMF (m = 24)	NoTMF (m = 168)	NoTMF-1st (m = 168)	TRMF	BTMF	BTRMF
1	1	13.63/2.88	13.53/2.86	**13.45**/**2.85**	14.50/3.12	14.94/3.13	15.93/3.33
	2	13.47/**2.84**	**13.41**/**2.84**	13.42/**2.84**	14.14/3.05	15.70/3.41	15.90/3.35
	3	13.46/2.84	**13.39**/**2.83**	13.43/2.84	13.87/2.96	15.80/3.34	16.08/3.43
	6	13.41/**2.83**	**13.39**/**2.83**	13.41/**2.83**	14.00/2.98	15.45/3.27	16.26/3.48
2	1	13.91/2.96	13.76/2.94	**13.70**/**2.92**	15.85/3.43	15.33/3.21	16.85/3.56
	2	13.77/2.92	**13.63**/**2.89**	13.72/2.92	15.04/3.31	15.87/3.32	17.27/3.71
	3	13.72/2.91	**13.61**/**2.89**	13.73/2.92	15.25/3.36	15.69/3.33	17.24/3.74
	6	13.59/**2.87**	**13.57**/2.88	13.68/2.91	14.92/3.24	15.91/3.39	18.18/3.97
3	1	14.30/3.05	14.06/3.02	**14.02**/**3.00**	17.52/3.83	15.86/3.32	18.61/3.91
	2	14.01/2.98	**13.84**/**2.94**	13.96/2.98	17.32/4.00	16.30/3.40	18.90/4.10
	3	13.95/2.97	**13.79**/**2.93**	13.98/2.98	16.91/3.71	16.56/3.49	18.68/4.05
	6	13.78/**2.92**	**13.73**/**2.92**	13.91/2.96	16.72/3.65	15.49/3.27	20.45/4.66
6	1	**14.61**/**3.11**	14.67/3.20	14.98/3.32	21.20/4.70	15.99/3.32	22.40/4.69
	2	**14.30**/**3.03**	14.33/3.09	14.90/3.28	20.87/5.01	16.04/3.33	23.56/5.63
	3	**14.26**/**3.03**	14.28/3.09	14.86/3.26	20.08/4.65	15.67/3.28	24.27/5.72
	6	**14.06**/**2.97**	14.16/3.06	14.80/3.23	20.40/4.35	16.38/3.50	26.34/6.60

NoTMF performs better than other TMF models (e.g., TRMF, BTMF & BTRMF).

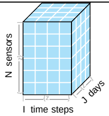
Task E: Sparse Urban Tra c State Forecasting

NoTMF forecasting (= 6)

- On the NYC Uber movement speed dataset

Task E: Sparse Urban Tra c State Forecasting

Conclusion

Sparse	
High-dimensional	
Multidimensional	
Noises & outliers	
Nonstationary	

)

Low-rank framework:

- NoTMF: matrix factorization
- LATC: low-rank tensor completion
- LCR: circulant matrix nuclear norm minimization
- HTF: tensor factorization

Temporal modeling:

- NoTMF: seasonal di erenced vector autoregression
- LATC: univariate autoregression
- LCR: temporal smoothing
- HTF: automatic temporal modeling with Hankel tensor

Highlights & Contributions

(Ours) LATC :
✓ Temporal autoregression

Matrix nuclear norm ($\|\boldsymbol{X}\|_*$) minimization
Candès & Recht'09

Singular value thresholding
Cai et al.'10

Truncated nuclear norm ($\|\boldsymbol{X}\|_{r,*}$; $r \in \mathbb{N}^+$) minimization
Zhang et al.'12
Hu et al.'12

Tensor nuclear norm ($\|\boldsymbol{\mathcal{X}}\|_*$) minimization
Liu et al.'13

Generalized nonconvex nonsmooth low-rank minimization
Lu et al.'14

Tensor nuclear norm minimization with linear transform
Lu et al.'19

Low-rank Hankel matrix/tensor ($\mathcal{H}(\cdot)$) completion
Yokota et al.'18
Sedighin et al.'20
Cai et al.'21
Yamamoto et al.'22

Circulant/Convolution nuclear norm ($\|\mathcal{C}(\boldsymbol{X})\|_*$ or $\|\mathcal{C}_\tau(\boldsymbol{X})\|_*$) minimization
Liu'22
Liu & Zhang'23

(Ours) LCR:
✓ Local trend modeling
✓ An FFT implementation

(Ours) HTF :
✓ Memory-efficient
✓ Conv. para.

References

A short list:

(Candes & Recht'09) \Exact matrix completion via convex optimization." Foundations of Computational Mathematics. 2009, 9(6): 717-772.

(Cai et al.'10) \A singular value thresholding algorithm for matrix completion." SIAM Journal on optimization. 2010, 20(4): 1956-1982.

(Zhang et al.'12) \Matrix completion by truncated nuclear norm regularization." IEEE Conference on computer vision and pattern recognition. 2012.

(Hu et al.'12) \Fast and accurate matrix completion via truncated nuclear norm regularization." IEEE transactions on pattern analysis and machine intelligence. 2012, 35(9): 2117-2130.

(Lu et al.'14) \Generalized nonconvex nonsmooth low-rank minimization." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

(Gultekin & Paisley'18) \Online forecasting matrix factorization." IEEE Transactions on Signal Processing. 2018, 67(5): 1223-1236.

(Yokota et al.'18) \Missing slice recovery for tensors using a low-rank model in embedded space." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

(Lu et al.'19) \Tensor robust principal component analysis with a new tensor nuclear norm." IEEE transactions on pattern analysis and machine intelligence. 2019, 42(4): 925-938.

(Cai et al.'21) \Accelerated structured alternating projections for robust spectrally sparse signal recovery." IEEE Transactions on Signal Processing. 2021, 69: 809-821.

(Chen & Sun'22) \Bayesian temporal factorization for multidimensional time series prediction." IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022, 44(9): 4659-4673.

(Liu'22) \Time series forecasting via learning convolutionally low-rank models." IEEE Transactions on Information Theory. 2022, 68(5): 3362-3380.

(Liu & Zhang'23) \Recovery of future data via convolution nuclear norm minimization." IEEE Transactions on Information Theory. 2023, 69(1): 650-665.

Thanks for your attention!

Any Questions?

Slides: https://xinychen.github.io/slides/sustech23.pdf

About me:

Homepage: https://xinychen.github.io

How to reach me: chenxy346@gmail.com