Robust low rank tensor completion with regularized redescending M-estimator

Yuning Yang, Yunlong Feng and Johan A.K. Suykens

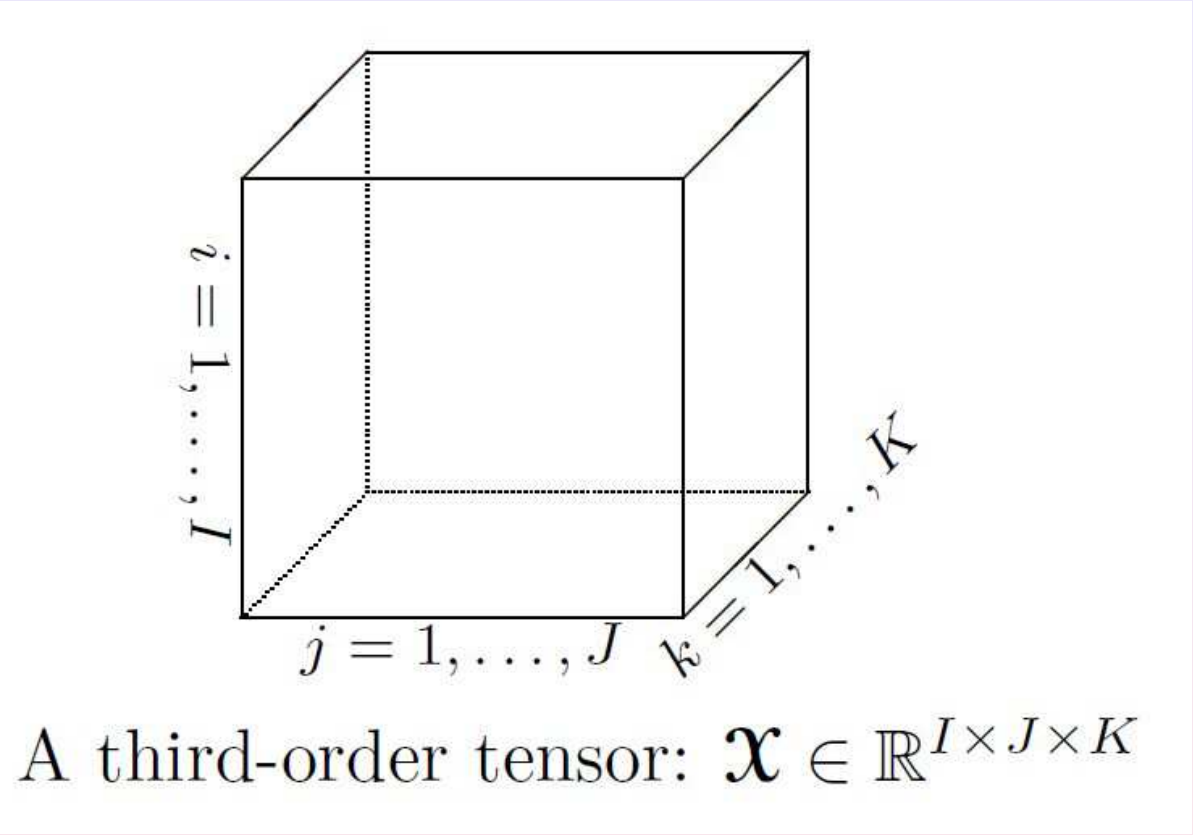
Department of Electrical Engineering
ESAT-STADIUS, KU Leuven, Belgium

July 14, 2014

Outline

- Tensor completion problems
- Robust tensor completion with regularized redescending M-estimator
- Computational algorithms and convergence results
- Applications
- Conclusions

Tensors



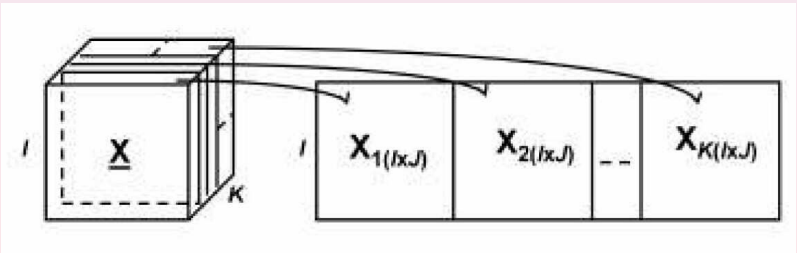
An N-th order tensor is denoted as $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$. We use $\mathbb{T} := \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$ to denote an N-th order tensor space

Tensor Completion Problems

- Tensor completion aims at recovering a low rank tensor from partial observations:

$$\min_{\mathcal{X}\in\mathbb{T}} \sum_{i=1}^{N} \text{rank}(X_{(i)}) \text{ s.t. } \mathcal{X}_{i_1\cdots i_N} = \mathcal{B}_{i_1\cdots i_N}, (i_1\cdots i_N) \in \Omega, \tag{1}$$

- $X_{(i)}$ is the mode-i unfolding matrix (matricization) of $\mathcal{X}$, and is of size $n_i \times \prod_{j\neq i}^{N} n_j$.



Tensor Completion Problems

- The nuclear norm based convex relaxation of (1)

$$\min_{\mathcal{X}\in\mathbb{T}} \sum_{i=1}^{N} \|X_{(i)}\|_* \text{ s.t. } \mathcal{X}_{i_1\cdots i_N} = \mathcal{B}_{i_1\cdots i_N}, (i_1\cdots i_N) \in \Omega. \quad (2)$$

Tensor Completion: Noise Case

- The observations might be contaminated by noise

$$\mathcal{B}_\Omega = \mathcal{X}_\Omega + \mathcal{E}, \tag{3}$$

where $\mathcal{E}$ refers to the noise.

- The model using the least squares loss:

$$\min_{\mathcal{X}\in\mathbb{T}} \ \frac{1}{2}\|\mathcal{X}_\Omega - \mathcal{B}_\Omega\|_F^2 + \lambda \sum_{i=1}^{N} \|X_{(i)}\|_* \tag{4}$$

- The model using the L_1 loss:

$$\min_{\mathcal{X}\in\mathbb{T}} \ \|\mathcal{X}_\Omega - \mathcal{B}_\Omega\|_1 + \lambda \sum_{i=1}^{N} \|X_{(i)}\|_* \tag{5}$$

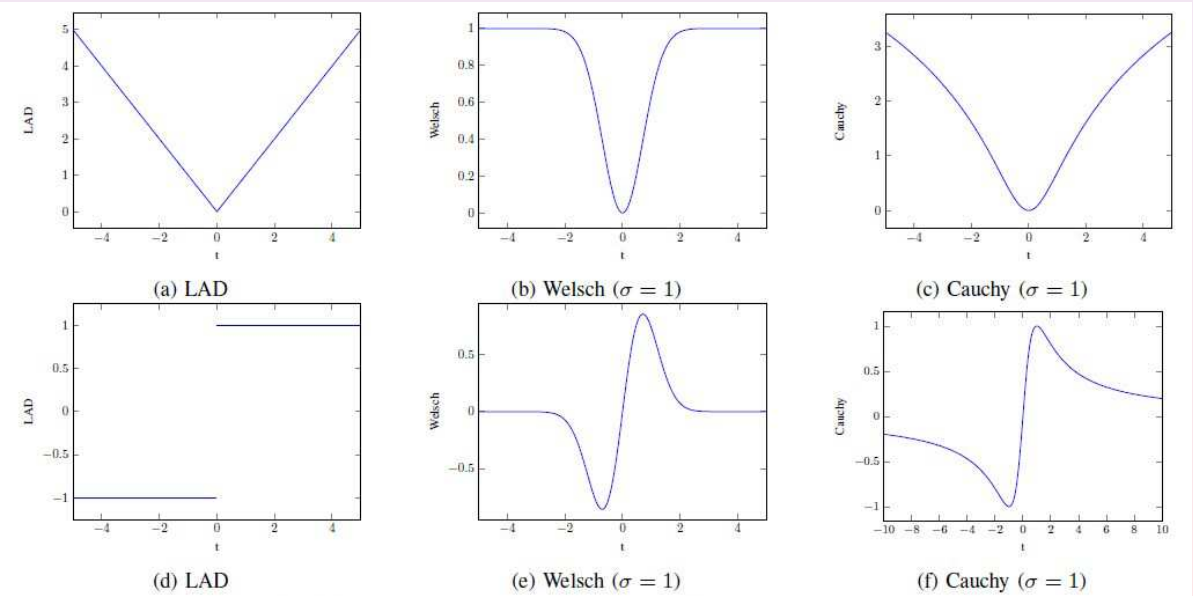
Our Approach

We adopt the following robust loss functions into the tensor completion problems

- Cauchy loss: $\rho_\sigma(t) = \frac{\sigma^2}{2} \log\left(1 + t^2/\sigma^2\right)$;
- Welsch loss: $\rho_\sigma(t) = \frac{\sigma^2}{2}\left(1 - \exp(-t^2/\sigma^2)\right)$.

Their influence functions:

- $\psi_\sigma^{\text{Cauchy}}(t) = t/(1 + t^2/\sigma^2)$,
- $\psi_\sigma^{\text{Welsh}}(t) = \exp(-t^2/\sigma^2)t$.



(a) LAD (b) Welsch ($\sigma = 1$) (c) Cauchy ($\sigma = 1$)

(d) LAD (e) Welsch ($\sigma = 1$) (f) Cauchy ($\sigma = 1$)

Fig. 1: Plots of different losses (top) and their influence functions (bottom).

Our Approach

The tensor model: Suppose the tensor to be recovered is of order N. we assume the tensor to be recovered can be approximated by the sum of N factor tensors, each being low rank in only one mode:

$$S(\boldsymbol{\mathcal{X}}) = \sum_{j=1}^{N} \mathcal{X}_j,$$

$$\boldsymbol{\mathcal{X}} := \{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_N\}.$$

with $X_{j,(j)}$ being a low rank matrix, $1 \leq j \leq N$, where $X_{j,(j)}$ represents the mode-j unfolding matrix of the j-th factor tensor.

Our Approach

Let the empirical risk term J_σ be

$$J_\sigma(S(\boldsymbol{\mathcal{X}})) = \sum_{(i_1 \cdots i_N) \in \Omega} \rho_\sigma \left(S(\boldsymbol{\mathcal{X}})_{i_1 \cdots i_N} - \mathcal{B}_{i_1 \cdots i_N} \right),$$

where ρ_σ is either the Cauchy loss or the Welsch loss,

$$\boldsymbol{\mathcal{X}} := \{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_N\}.$$

- Nuclear norm regularization model

$$\min_{\boldsymbol{\mathcal{X}} \in \mathbb{T}^N} \; J_\sigma \left(S(\boldsymbol{\mathcal{X}}) \right) + \lambda \sum_{j=1}^{N} \| X_{j,(j)} \|_*, \tag{6}$$

  where $\lambda > 0$ is a regularization parameter.

- Rank constraint model

$$\min_{\boldsymbol{\mathcal{X}} \in \mathbb{T}^N} \; J_\sigma \left(S(\boldsymbol{\mathcal{X}}) \right) \;\; \text{s.t.} \;\; \operatorname{rank}(X_{j,(j)}) \leq R_j, \; 1 \leq j \leq N. \tag{7}$$

Algorithm: Proximal and Linearized Block Coordinate Descent Methods

Solve the nuclear norm regularization model

$$\min_{\mathcal{X}_1,\ldots,\mathcal{X}_N} J_\sigma\left(\sum_{j=1}^{N}\mathcal{X}_j\right)+\lambda\sum_{j=1}^{N}\|X_{j,(j)}\|_*. \tag{8}$$

Block coordinate descent (BCD) method?

$$\mathcal{X}_j^{(k+1)}\in\arg\min_{\mathcal{Z}\in\mathbb{T}}\Phi(\mathcal{X}_1^{(k+1)},\ldots,\mathcal{X}_{j-1}^{(k+1)},\mathcal{Z},\mathcal{X}_{j+1}^{(k)},\mathcal{X}_N^{(k)}).$$

The subproblems are nonconvex nor have closed-form solutions.

To obtain $\mathcal{X}_j^{(k+1)}$, we solve the following subproblem

$$\mathcal{X}_j^{(k+1)}=\arg\min_{\mathcal{X}\in\mathbb{T}}\ \|X_{(j)}\|_*+\frac{\alpha}{2}\|\mathcal{X}-\mathcal{X}_j^{(k)}\|_F^2 \tag{9}$$
$$+\left\langle\nabla_{\mathcal{X}_j}J_\sigma\left(\sum_{i=1}^{j-1}\mathcal{X}_i^{(k+1)}+\sum_{i=j}^{N}\mathcal{X}_i^{(k)}\right),\mathcal{X}-\mathcal{X}_j^{(k)}\right\rangle.$$

Algorithm: Proximal and Linearized Block Coordinate Descent Methods

$$\mathcal{X}_j^{(k+1)} = \arg\min_{\mathcal{X}\in\mathbb{T}} \frac{1}{\alpha}\|X_{(j)}\|_* + \frac{1}{2}\|\mathcal{X} - \mathcal{Y}_j^{(k+1)}\|_F^2. \qquad (10)$$

Here

$$\mathcal{Y}_j^{(k+1)} = \mathcal{X}_j^{(k)} - \frac{1}{\alpha}\nabla_{\mathcal{X}_j} J_\sigma \left(\sum_{i=1}^{j-1} \mathcal{X}_i^{(k+1)} + \sum_{i=j}^{N} \mathcal{X}_i^{(k)} \right), \qquad (11)$$

- (10) has a closed-form solution, which is given by the matrix soft thresholding operator. Let $Y_{j,(j)}^{k+1} = U\mathrm{diag}\left(\{\sigma_i\}_{1\le i\le r}\right)V^T$ be the SVD of $Y_{j,(j)}^{k+1}$. Then

$$X_{j,(j)}^{(k+1)} = S_{1/\alpha}(Y_{j,(j)}^{(k+1)}) = U\mathrm{diag}\left(\max\{\sigma_i - 1/\alpha, 0\}\right)V^T.$$

Algorithm: Proximal and Linearized Block Coordinate Descent Methods

Algorithm 1 Proximal and Linearized BCD Method

Input: initial guess $\boldsymbol{\mathcal{X}}^{(0)} \in \mathbb{T}^N$, parameter $\sigma > 0$, $\lambda > 0$.

Output: the recovered matrix

$$\boldsymbol{\mathcal{X}}^{(k+1)} = \{\mathcal{X}_1^{(k+1)}, \ldots, \mathcal{X}_N^{(k+1)}\}.$$

for $k = 1$ **to** $\ldots$ **do**

for $j = 1$ **to** N **do**

- Compute

$$\mathcal{Y}_j^{(k+1)} = \mathcal{X}_j^{(k)} - \frac{1}{\alpha}\nabla_{\mathcal{X}_j} J_\sigma \left(\sum_{i=1}^{j-1} \mathcal{X}_i^{(k+1)} + \sum_{i=j}^{N} \mathcal{X}_i^{(k)} \right).$$

- Perform the matrix soft thresholding operator on $Y_j^{(k+1)}$ such that

$$X_{j,(j)}^{(k+1)} = S_{1/\alpha}\left(Y_{j,(j)}^{(k+1)}\right).$$

- Fold the matrix $X_{j,(j)}^{(k+1)}$ back to the tensor $\mathcal{X}_j^{(k+1)}$.

end for

end for

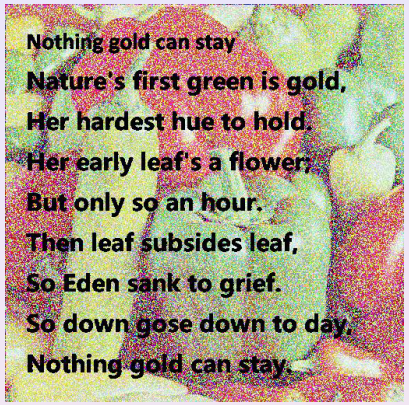
Convergence Results

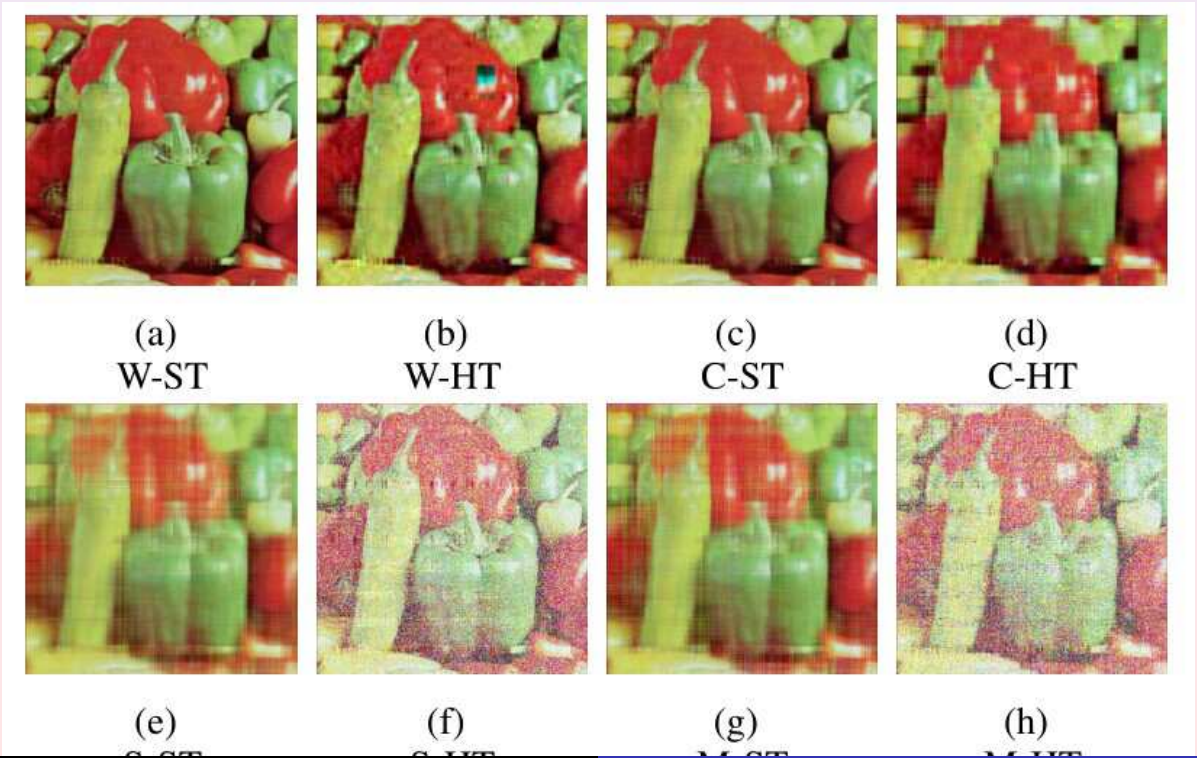
Theorem

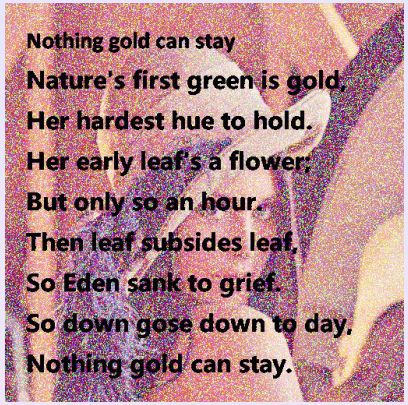
Let $\{\boldsymbol{\mathcal{X}}^{(k)}\}$ be a sequence generated by Algorithm 1, and let the fixed step-size satisfy $\alpha > 1$. Then $\{\Phi(\boldsymbol{\mathcal{X}}^{(k)})\}$ is monotonically decreasing, and $\{\boldsymbol{\mathcal{X}}^{(k)}\}$ converges to a critical point of Φ, where Φ denotes the objective function of the nuclear norm regularization model.

Numerical Experiments

- Color image inpainting






Nothing gold can stay
Nature's first green is gold,
Her hardest hue to hold.
Her early leaf's a flower;
But only so an hour.
Then leaf subsides leaf,
So Eden sank to grief.
So down gose down to day,
Nothing gold can stay.


W-ST
W-HT
C-ST
C-HT
S-ST
S-HT
M-ST
M-HT

Traffic data flow recovery

- In traffic systems, the traffic data flow, collected by detectors, are used to analyze the traffic situations.
- Traffic flow data: 8928 records. Record = flow volume per 5 min

$$8928 = 12 \text{ (records per hour)} \times 24 \text{ (hours per day)} \times 31\text{(days in a month)}$$

Can be represented by a third order tensor of size $12 \times 24 \times 31$.

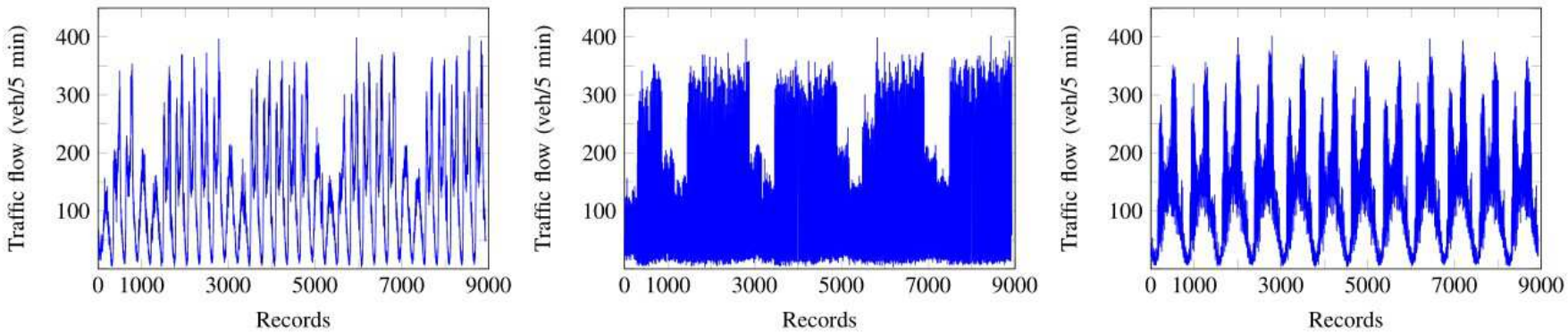
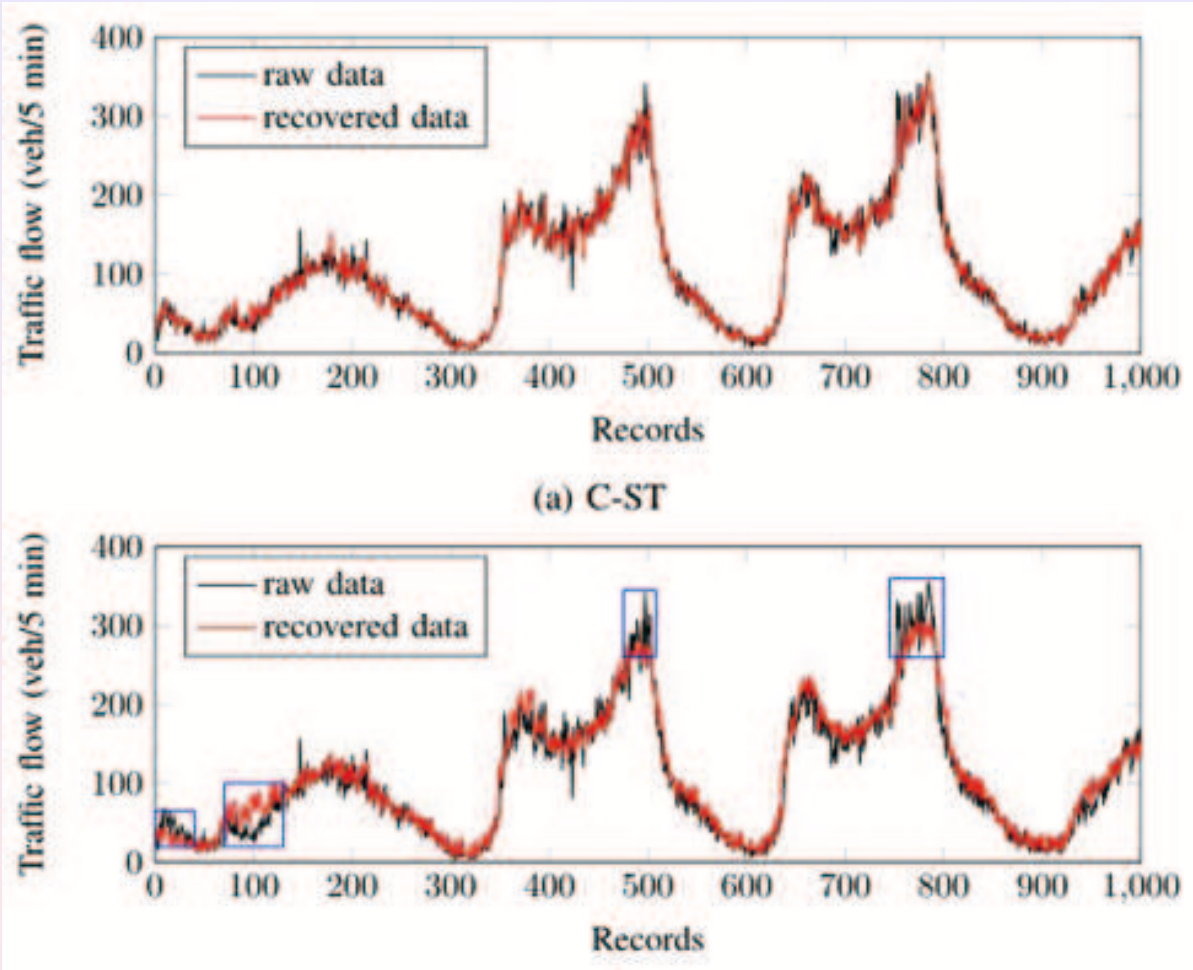


Fig. 6: Plots of vectorizations of the traffic data tensor along mode-1, mode-2 and mode-3, respectively.

20% entries miss randomly; 30% outliers are added.



(a) C-ST

Conclusions

- A robust tensor completion approach with regularized redescending M-estimators
- Linearized and proximal block coordinated descent method to solve the models
- Applications

The End

Thank you for your attention!