Robust Tensor Completion and its Applications

Michael K. Ng
Department of Mathematics
Hong Kong Baptist University
mng@math.hkbu.edu.hk

MLA'18 - The 16th China Symposium on Machine Learning and Applications

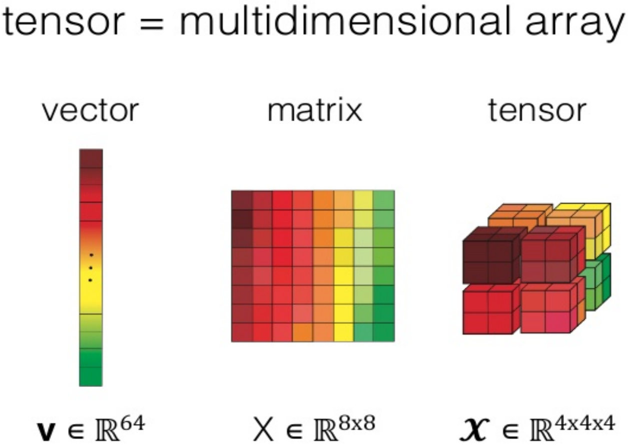
3 November 2018

Outline

- Introduction
- Low Rank Tensor Recovery and t-SVD
- t-SVD Completion
- Correction Model
- Generalization
- Multi-view Clustering
- Summary

Introduction

Example


tensor = multidimensional array
vector
matrix
tensor
$\mathbf{v} \in \mathbb{R}^{64}$
$X \in \mathbb{R}^{8x8}$
$\mathcal{X} \in \mathbb{R}^{4x4x4}$

Application: Moment and Cumulant Tensors

Let $\mathbf{x}$ be a random vector of dimension n with components x_i. Its moment and cumulant tensors of order m as

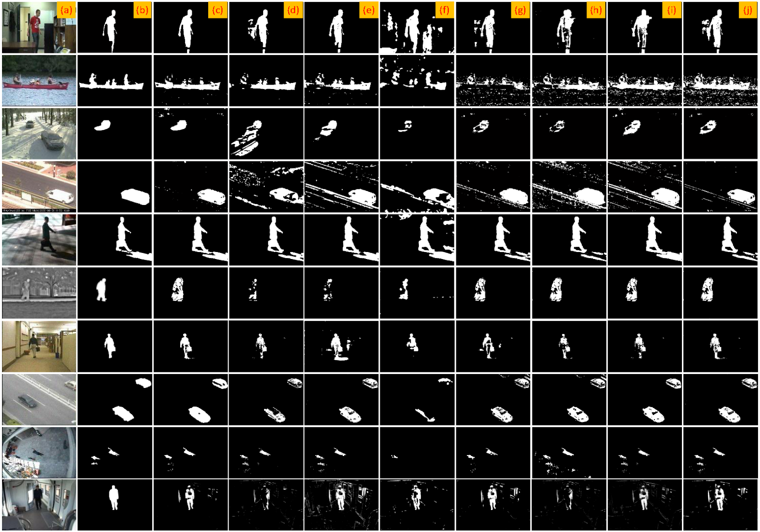
$$\mathcal{M}(\mathbf{x}) = [\mu_{i_1,i_2,\cdots,i_m}] \quad \text{with} \quad \mu_{i_1,i_2,\cdots,i_m} = \mathtt{E}\{x_{i_1}x_{i_2}\cdots x_{i_m}\}$$

and

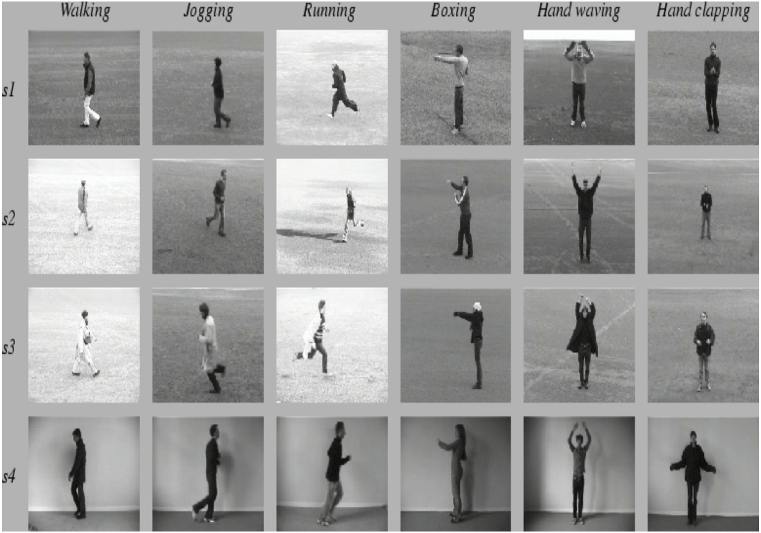
$$\mathcal{C}(\mathbf{x}) = [c_{i_1,i_2,\cdots,i_m}] \quad \text{with} \quad c_{i_1,i_2,\cdots,i_m} = \mathtt{Cum}\{x_{i_1}x_{i_2}\cdots x_{i_m}\}$$

(3rd order: skewness and 4th order: kurtosis)

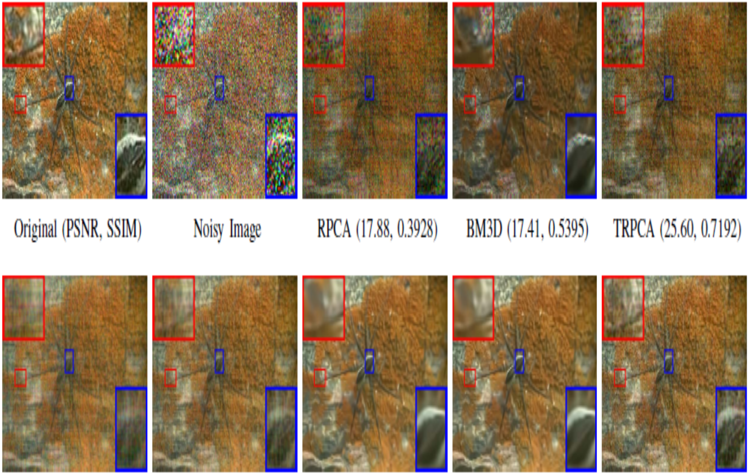
Application: Background and Foreground Separation in Video (spatial dimensions + time)



Application: Recognition (spatial dimensions + time)



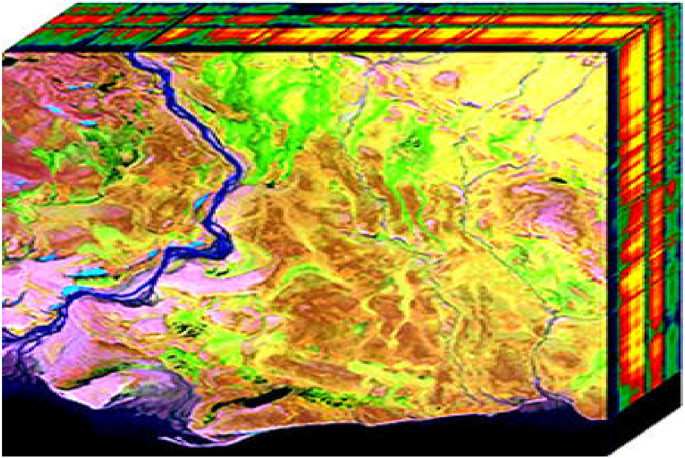
Application: Color Images Completion and Denoising (spatial dimensions + RGB)



Original (PSNR, SSIM) | Noisy Image | RPCA (17.88, 0.3928) | BM3D (17.41, 0.5395) | TRPCA (25.60, 0.7192)

RMC (24.03, 0.6380) | SNN (25.47, 0.7399) | BM3D+ (25.60, 0.6279) | BM3D++ (26.18, 0.6668) | RTC (28.39, 0.8593)

Application: Hyperspectral Images Completion and Denoising (spatial dimensions + frequencies)



Application: Adjacency and Laplacian Tensors

As a generalization of a graph, a uniform hypergraph $G = (V, E)$ with $V = \{1, 2, \cdots, n\}$ the vertex set and $E = \{e_1, e_2, \cdots, e_m\}$ the edge set, is defined to satisfy that $|e_p| = k$ for any $e_p \subset V$, $p = 2, \cdots, m$ and $k \geq 2$. Such a uniform hypergraph is also called a k-graph. If $k = 2$, G is exactly an ordinary graph.

Given a k-graph G, its adjacency tensor $\mathcal{A}$ ($\mathcal{A}(G)$) of G, is a k-th order n-dimensional symmetric tensor, defined as $\mathcal{A} = [a_{i_1,i_2,\cdots,i_m}]$ where $a_{i_1,i_2,\cdots,i_m} = \frac{1}{(k-1)!}$ if $(i_1, i_2, \cdots, i_m) \in E$, and 0 otherwise.

For $i \in V$, its degree $d(i)$ is defined as $d(i) = |\{e_i : i \in e_p \in E\}|$. The degree tensor $\mathcal{D}$ of G is a k-th order n-dimensional tensor: $d_{i,i,\cdots,i} = d(i)$. The Laplacian tensor is defined $\mathcal{D} - \mathcal{A}$.

Multiple Relations Tensor

- Tensor can be used to describe the multiple relationships between objects. A tensor is a multidimensional array. Here a three-way array (third-order tensor) is used:

	O_1	O_2	$\cdots$	O_n
O_1	$a_{1,1,1}$	$a_{1,2,1}$	$\cdots$	$a_{1,n,1}$
O_2	$a_{2,1,1}$	$a_{2,2,1}$	$\cdots$	$a_{2,n,1}$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
O_n	$a_{n,1,1}$	$a_{n,2,1}$	$\cdots$	$a_{n,n,1}$

	O_1	O_2	$\cdots$	O_n
O_1	$a_{1,1,2}$	$a_{1,2,2}$	$\cdots$	$a_{1,n,2}$
O_2	$a_{2,1,2}$	$a_{2,2,2}$	$\cdots$	$a_{2,n,2}$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
O_n	$a_{n,1,2}$	$a_{n,2,2}$	$\cdots$	$a_{n,n,2}$

$\ldots$

	O_1	O_2	$\cdots$	O_n
O_1	$a_{1,1,p}$	$a_{1,2,p}$	$\cdots$	$a_{1,n,p}$
O_2	$a_{2,1,p}$	$a_{2,2,p}$	$\cdots$	$a_{2,n,p}$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
O_n	$a_{n,1,p}$	$a_{n,2,p}$	$\cdots$	$a_{n,n,p}$

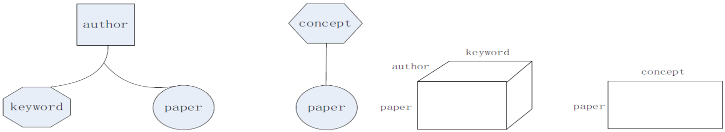
- p relationships among n objects

Application: Information Retrieval

- Web information retrieval is significantly more challenging than that based on web hyperlink structure
- One main difference is the multiple links based on the other features (text, images, etc)
- Example: 100,000 webpages from .GOV Web collection in 2002 TREC and 50 topic distillation topics in TREC 2003 Web track as queries
- Multiple links among webpages via different anchor texts
- 39,255 anchor terms (multiple relations), and 479,122 links with these anchor terms among the 100,000 webpages

Application: Networks

- In a social network where objects are connected via multiple relations, via sharing, comments, stories, photos, tags, keywords, topics, etc
- In a publication network where the interactions among items in three entities: author, keyword and paper



- A tensor: the interactions among items in three dimensions/entities: author, keyword and paper; A matrix: the interactions between items in two dimensions/entities: concept and paper

Tensor Decomposition

CANDECOMP/PARAFAC Decomposition:

$$\mathcal{X} = \sum_{i=1}^{r} \lambda_i \mathbf{a}^{i,1} \otimes \cdots \otimes \mathbf{a}^{i,m}$$

The minimal value of r is called the rank of $\mathcal{A}$.

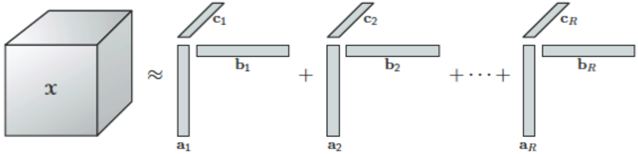


Fig. 3.1 *CP decomposition of a three-way array.*

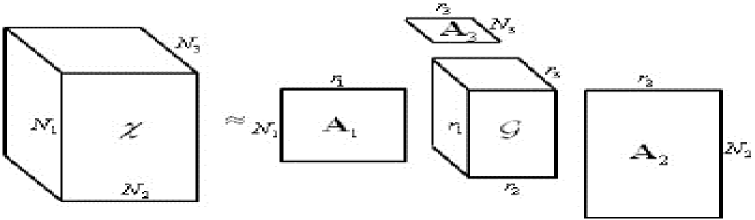
Tensor Decomposition

Tucker Decomposition:

$$\mathcal{X} = \mathcal{G} \times \mathbf{A}_1 \times \mathbf{A}_2 \cdots \times \mathbf{A}_m$$

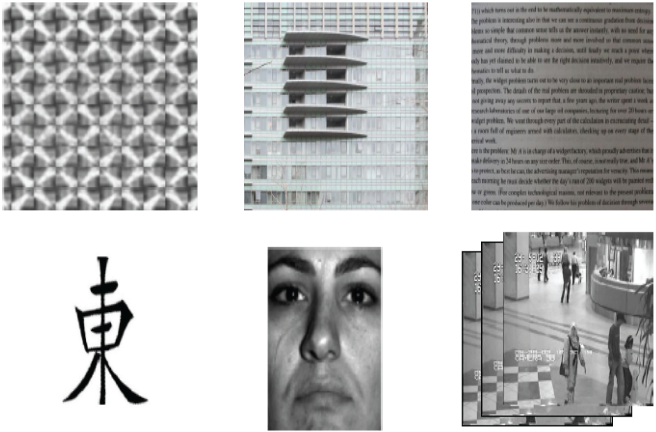
$$\mathcal{X} = \sum_{i_1=1}^{r_1} \cdots \sum_{i_m=1}^{r_m} g_{i_1,i_2,\cdots,i_m} \mathbf{a}^{i_1,1} \otimes \cdots \otimes \mathbf{a}^{i_m,m}$$

It can be obtained by using singular value decomposition to each unfolded matrix $\mathbf{X}_{i_j}$ from $\mathcal{X}$. The Tucker rank is $(rank(\mathbf{X}_1), rank(\mathbf{X}_2), \cdots, rank(\mathbf{X}_m)) = (r_1, r_2, \cdots, r_m)$.



Low Rank Tensor Recovery

Low-dimensional Structure



Data in many real applications exhibit low-dimensional structures due to local regularities, global symmetries, repetitive patterns, redundant sampling, ... (low-dimensional structure $\rightarrow$ low-rank data matrices)

Example

Customer/Item	I	II	III	IV	$\cdots$
A	5	1	?	?	$\cdots$
B	?	2	3	?	$\cdots$
C	?	?	4	2	$\cdots$
D	1	?	?	?	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$

For example (Netflix Challenge 2009), it is about 0.5 million users and about 18,000 movies

Matrix Completion

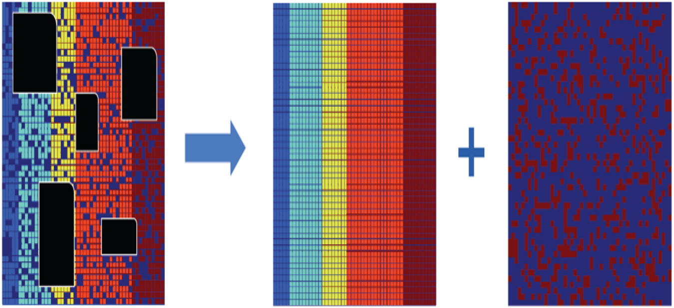
$$\min_{\mathbf{X}} \ \mathit{rank}(\mathbf{X}) \quad \text{subject to} \quad P_\Omega(\mathbf{X}) = P_\Omega(\mathbf{M})$$

Example

Matrix RPCA

$$\min_{\mathbf{X}} \ \mathit{rank}(\mathbf{X}) + \lambda \|\mathbf{E}\|_0 \quad \text{subject to} \quad \mathbf{X} + \mathbf{E} = \mathbf{M}$$

Example



Robust Matrix Completion

$$\min_{\mathbf{X}} \; rank(\mathbf{X}) + \lambda \|\mathbf{E}\|_0 \quad \text{subject to} \quad P_\Omega(\mathbf{X} + \mathbf{E}) = P_\Omega(\mathbf{M})$$

Low Rank Matrix Recovery

- Matrix Completion

$$\min_{\mathbf{X}} \ rank(\mathbf{X}) \quad \text{subject to} \quad P_\Omega(\mathbf{X}) = P_\Omega(\mathbf{M})$$

- Matrix RPCA

$$\min_{\mathbf{X}} \ rank(\mathbf{X}) + \lambda \|\mathbf{E}\|_0 \quad \text{subject to} \quad \mathbf{M} = \mathbf{X} + \mathbf{E}$$

- Robust Matrix Completion

$$\min_{\mathbf{X}} \ rank(\mathbf{X}) + \lambda \|\mathbf{E}\|_0 \quad \text{subject to} \quad P_\Omega(\mathbf{M}) = P_\Omega(\mathbf{X} + \mathbf{E})$$

Low Rank Matrix Recovery

- Matrix Completion

$$\min_{\mathbf{X}} \ \|\mathbf{X}\|_* \quad \text{subject to} \quad P_\Omega(\mathbf{X}) = P_\Omega(\mathbf{M})$$

- Matrix RPCA

$$\min_{\mathbf{X}} \ \|\mathbf{X}\|_* + \lambda\|\mathbf{E}\|_1 \quad \text{subject to} \quad \mathbf{M} = \mathbf{X} + \mathbf{E}$$

- Robust Matrix Completion

$$\min_{\mathbf{X}} \ \|\mathbf{X}\|_* + \lambda\|\mathbf{E}\|_1 \quad \text{subject to} \quad P_\Omega(\mathbf{M}) = P_\Omega(\mathbf{X} + \mathbf{E})$$

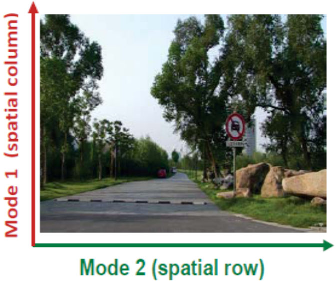
Nuclear norm $\|\cdot\|_*$: sum of singular values (convex envelop of rank)

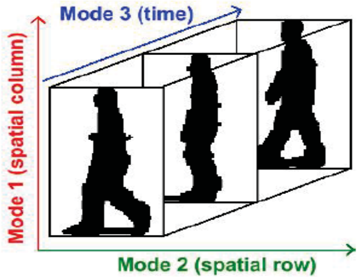
Low Rank Matrix Recovery Results

- (RPCA) Candes, E. J., Li, X., Ma, Y., and Wright, J. Journal of the ACM, 58(3):173, 2011.
- (Matrix Completion) Recht, B. Journal of Machine Learning Research, 12(4):34133430, 2011.
- (Matrix Completion) Chen, Y. IEEE Transactions on Information Theory, 61(5):29092923, 2013.
- many papers ...

Low Rank Tensor Recovery

Data are usually in multi-dimensional array.





“Vectorization” probably break the inherent structures and correlations in the original data.

Low Rank Tensor Recovery

- Tensor Completion

$$\min_{\mathcal{X}} \ \mathit{rank}(\mathcal{X}) \quad \text{subject to} \quad P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{M})$$

- Tensor Robust PCA

$$\min_{\mathcal{X}} \ \mathit{rank}(\mathcal{X}) + \lambda \|\mathcal{E}\|_0 \quad \text{subject to} \quad \mathcal{M} = \mathcal{X} + \mathcal{E}$$

- Robust Tensor Completion

$$\min_{\mathcal{X}} \ \mathit{rank}(\mathcal{X}) + \lambda \|\mathcal{E}\|_0 \quad \text{subject to} \quad P_\Omega(\mathcal{M}) = P_\Omega(\mathcal{X} + \mathcal{E})$$

Low Rank Tensor Recovery

- CP decomposition/rank cannot be computed efficiently
- Matrix rank can be replaced by matrix nuclear norm (the sum of singular values), it is a convex envelope
- Replace Tucker rank by the sum of nuclear norms of unfolding tensors, interdependent matrix trace norm is involved
- The use of the sum of nuclear norms of unfolding matrices of a tensor may be challenged since it is suboptimal[1]
- The tensor trace norm (the average of trace norms of unfolding matrices) is not a tight convex relaxation of the tensor rank (the average rank of unfolding matrices) [2]

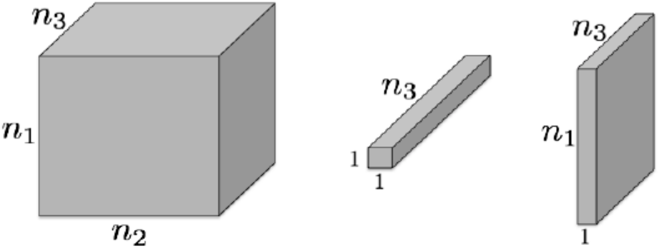
[1] C. Mu, B. Huang, J. Wright, and D. Goldfarb. Square deal: Lower bounds and improved relaxations for tensor recovery. In ICML, pages 7381, 2014.

[2] B. Romera-Paredes and M. Pontil. A new convex relaxation for tensor completion. In Adv. Neural Inf. Process. Syst., pages 29672975, 2013.

t-SVD

t-SVD Decomposition

A third-order tensor of size $n_1 \times n_2 \times n_3$ can be viewed as an $n_1 \times n_2$ matrix of tubes which lie in the third-dimension. [Kilmer, M. E. and Martin, C. D. Linear Algebra & Its Applications, 435(3):641658, 2011]



t-SVD Decomposition

Definition: The t-product $\mathcal{A} * \mathcal{B}$ of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ is a tensor $\mathcal{C} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$ whose (i, j)th tube is given by

$$\mathcal{C}(i, j, :) = \sum_{k=1}^{n_2} \mathcal{A}(i, k, :) * \mathcal{B}(k, j, :),$$

where $*$ denotes the circular convolution between two tubes of same size.

The tube at (i, k) position in $\mathcal{A}$ convolutes with the tube at (k, j) position in $\mathcal{B}$. Both have sizes n_3. Put all the correlations at (i, j) position in $\mathcal{C}$.

The multiplication of between the scalars is replaced by circular convolution between the tubes.

t-SVD Decomposition

Definition: The identity tensor $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$ is defined to be a tensor whose first frontal slice $\mathcal{I}^{(1)}$ is the $n \times n$ identity matrix and whose other frontal slices $\mathcal{I}^{(i)}, i = 2, \ldots, n_3$ are zero matrices.

Definition: The conjugate transpose of a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the tensor $\mathcal{A}^H \in \mathbb{R}^{n_2 \times n_1 \times n_3}$ obtained by conjugate transposing each of the frontal slice and then reversing the order of transposed frontal slices 2 through n_3, i.e.,

$$\begin{aligned} \left(\mathcal{A}^H\right)^{(1)} &= \left(\mathcal{A}^{(1)}\right)^H, \\ \left(\mathcal{A}^H\right)^{(i)} &= \left(\mathcal{A}^{(n_3+2-i)}\right)^H, \quad i = 2, \ldots, n_3. \end{aligned}$$

t-SVD Decomposition

Definition: A tensor $\mathcal{Q} \in \mathbb{R}^{n \times n \times n_3}$ is orthogonal if it satisfies

$$\mathcal{Q}^H * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^H = \mathcal{I},$$

where $\mathcal{I}$ is the identity tensor of size $n \times n \times n_3$.

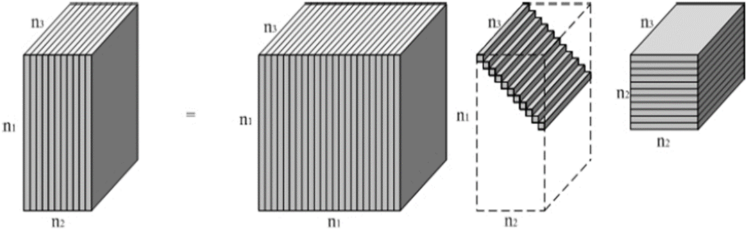
Definition: A tensor $\mathcal{A}$ is called f-diagonal if each frontal slice $\mathcal{A}^{(i)}$ is a diagonal matrix.

t-SVD Decomposition

For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the t-SVD of $\mathcal{A}$ is given by

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^H,$$

where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors, and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a f-diagonal tensor, respectively. The entries in $\mathcal{S}$ are called the singular tubes of $\mathcal{A}$.



t-SVD Decomposition

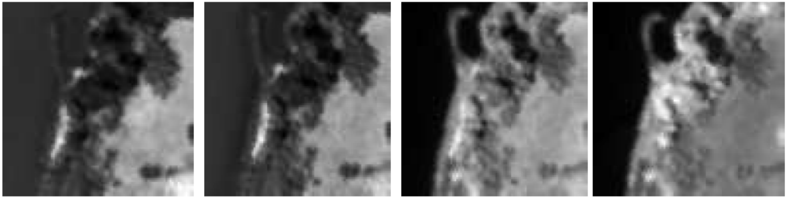
The tensor tubal-rank, denoted as $rank_t(\mathcal{A})$, is defined as the number of nonzero singular tubes of $\mathcal{S}$, where $\mathcal{S}$ comes from the t-SVD of $\mathcal{A}$, i.e.,

$$rank_t(\mathcal{A}) = \#\{i : \mathcal{S}(i, i, :) \neq \vec{0}\}.$$

It can be shown that it is equal to $\max_i rank(\hat{\mathcal{A}}^{(i)})$ where $\hat{\mathcal{A}}^{(i)}$ is the i-th slice of $\hat{\mathcal{A}}$ and $\hat{\mathcal{A}}$ represents a third-order tensor obtained by taking the Discrete Fourier Transform (DFT) of all the tubes along the third dimension of $\mathcal{A}$.

Example of t-SVD Decomposition

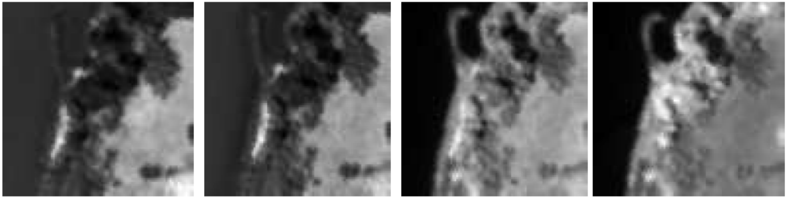
Original Images



(a) 20st (b) 60st (c) 100st (d) 140st

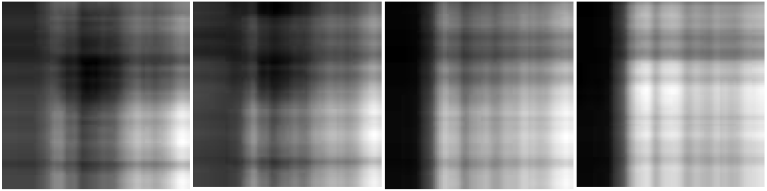
Original images with different bands for the Samson hyperspectral data.

Tubal Rank 1



(a) 20st (b) 60st (c) 100st (d) 140st

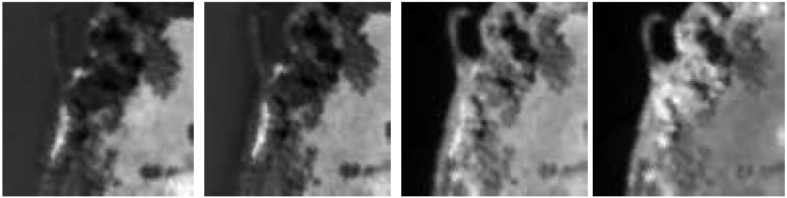
Original images with different bands for the Samson hyperspectral data.



(a) 20st (b) 60st (c) 100st (d) 140st

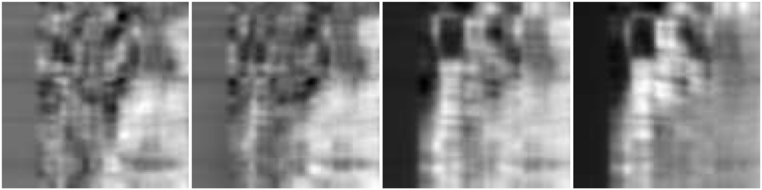
The $\mathcal{U} * \mathcal{S} * \mathcal{V}^H$ by using 1st tube in $\mathcal{S}$ and the corresponding 1st $\mathcal{U}$ and $\mathcal{V}^H$ components for the Samson hyperspectral data.

Tubal Rank 5



(a) 20st (b) 60st (c) 100st (d) 140st

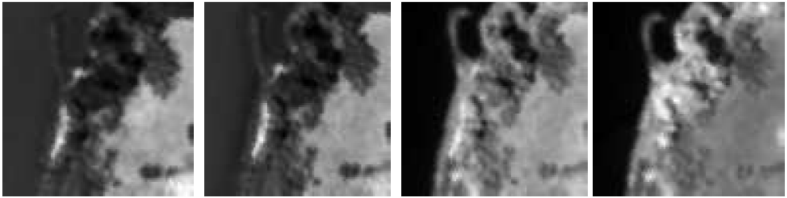
Original images with different bands for the Samson hyperspectral data.



(a) 20st (b) 60st (c) 100st (d) 140st

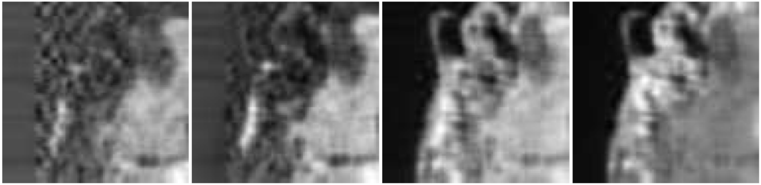
Figure 0.3: The $\mathcal{U} * \mathcal{S} * \mathcal{V}^H$ by using 5st tube in $\mathcal{S}$ and the corresponding 5st $\mathcal{U}$ and $\mathcal{V}^H$ components for the Samson hyperspectral data.

Tubal Rank 10



(a) 20st (b) 60st (c) 100st (d) 140st

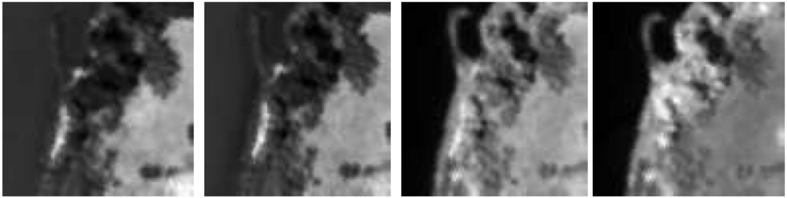
Original images with different bands for the Samson hyperspectral data.



(a) 20st (b) 60st (c) 100st (d) 140st

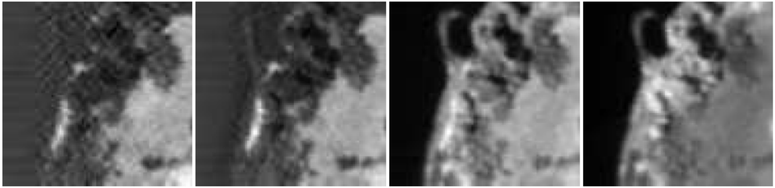
The $\mathcal{U} * \mathcal{S} * \mathcal{V}^H$ by using 10st tube in $\mathcal{S}$ and the corresponding 10st $\mathcal{U}$ and $\mathcal{V}^H$ components for the Samson hyperspectral data.

Tubal Rank 20



(a) 20st (b) 60st (c) 100st (d) 140st

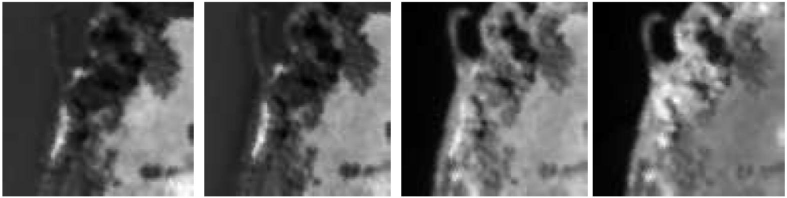
Original images with different bands for the Samson hyperspectral data.



(a) 20st (b) 60st (c) 100st (d) 140st

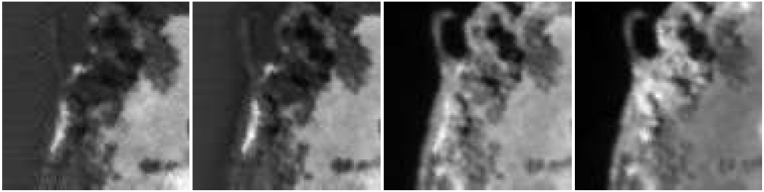
The $\mathcal{U} * \mathcal{S} * \mathcal{V}^H$ by using 20st tube in $\mathcal{S}$ and the corresponding 20st $\mathcal{U}$ and $\mathcal{V}^H$ components for the Samson hyperspectral data.

Tubal Rank 40



(a) 20st (b) 60st (c) 100st (d) 140st

Original images with different bands for the Samson hyperspectral data.



(a) 20st (b) 60st (c) 100st (d) 140st

Figure 0.6: The $\mathcal{U} * \mathcal{S} * \mathcal{V}^H$ by using 40st tube in $\mathcal{S}$ and the corresponding 40st $\mathcal{U}$ and $\mathcal{V}^H$ components for the Samson hyperspectral data.

Low Tubal Rank Tensor Recovery

- Tensor Completion

$$\min_{\mathcal{X}} \; \mathit{rank}(\mathcal{X}) \quad \text{subject to} \quad P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{M})$$

- Tensor Robust PCA

$$\min_{\mathcal{X}} \; \mathit{rank}(\mathcal{X}) + \lambda \|\mathcal{E}\|_0 \quad \text{subject to} \quad \mathcal{M} = \mathcal{X} + \mathcal{E}$$

- Robust Tensor Completion

$$\min_{\mathcal{X}} \; \mathit{rank}(\mathcal{X}) + \lambda \|\mathcal{E}\|_0 \quad \text{subject to} \quad P_\Omega(\mathcal{M}) = P_\Omega(\mathcal{X} + \mathcal{E})$$

TNN

Definition: The tubal nuclear norm of a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, denoted as $\|\mathcal{A}\|_{\mathsf{TNN}}$, is the nuclear norm of all the frontal slices of $\hat{\mathcal{A}}$.

Theorem

For any tensor $\mathcal{X} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, $\|\mathcal{X}\|_{\mathsf{TNN}}$ is the convex envelope of the function $\sum_{i=1}^{n_3} \operatorname{rank}(\widehat{\mathcal{A}}^{(i)})$ on the set $\{\mathcal{X} \mid \|\mathcal{X}\| \leq 1\}$.

Low Tubal Rank Tensor Recovery (Relaxation)

- Tensor Completion

$$\min_{\mathcal{X}} \ \|\mathcal{X}\|_{\mathsf{TNN}} \quad \text{subject to} \quad P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{M})$$

- Tensor Robust PCA

$$\min_{\mathcal{X}} \ \|\mathcal{X}\|_{\mathsf{TNN}} + \lambda\|\mathcal{E}\|_1 \quad \text{subject to} \quad \mathcal{M} = \mathcal{X} + \mathcal{E}$$

- Robust Tensor Completion

$$\min_{\mathcal{X}} \ \|\mathcal{X}\|_{\mathsf{TNN}} + \lambda\|\mathcal{E}\|_1 \quad \text{subject to} \quad P_\Omega(\mathcal{M}) = P_\Omega(\mathcal{X} + \mathcal{E})$$

Can we recover low-tubal-rank tensor from partial and grossly corrupted observations exactly ?

Tensor Incoherence Conditions

Assume that $rank_t(\mathcal{L}_0) = r$ and its t-SVD $\mathcal{L}_0 = \mathcal{U} * \mathcal{S} * \mathcal{V}^H$. $\mathcal{L}_0$ is said to satisfy the tensor incoherence conditions with parameter $\mu > 0$ if

$$\max_{i=1,\cdots,n_1} \|\mathcal{U}^H * \vec{e}_i\|_F \leq \sqrt{\frac{\mu r}{n_1}},$$

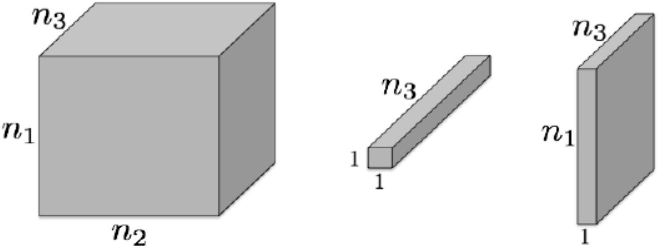
$$\max_{j=1,\cdots,n_2} \|\mathcal{V}^H * \vec{e}_j\|_F \leq \sqrt{\frac{\mu r}{n_2}},$$

and (joint incoherence condition)

$$\|\mathcal{U} * \mathcal{V}^H\|_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2 n_3}}.$$

Tensor Incoherence Conditions

The **column basis**, denoted as $\vec{e}_i$, is a tensor of size $n_1 \times 1 \times n_3$ with its $(i, 1, 1)$th entry equaling to 1 and the rest equaling to 0. The **tube basis**, denoted as $\mathring{\mathbf{e}}_k$, is a tensor of size $1 \times 1 \times n_3$ with its $(1, 1, k)$th entry equaling to 1 and the rest equaling to 0.



Low Rank Tensor Recovery

Theorem

Suppose $\mathcal{L}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *obeys tensor incoherence conditions, and the observation set* Ω *is uniformly distributed among all sets of cardinality* $m = \rho n_1 n_2 n_3$*. Also suppose that each observed entry is independently corrupted with probability* γ*. Then, there exist universal constants* $c_1, c_2 > 0$ *such that with probability at least* $1 - c_1(n_{(1)} n_3)^{-c_2}$*, the recovery of* $\mathcal{L}_0$ *with* $\lambda = 1/\sqrt{\rho n_{(1)} n_3}$ *is exact, provided that*

$$r \leq \frac{c_r n_{(2)}}{\mu(\log(n_{(1)} n_3))^2} \quad \text{and} \quad \gamma \leq c_\gamma$$

where c_r *and* c_γ *are two positive constants.*

$n_{(1)} = \max\{n_1, n_2\}$ and $n_{(2)} = \min\{n_1, n_2\}$

Low Rank Tensor Recovery

Theorem

(Tensor Completion): Suppose $\mathcal{L}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ obeys tensor incoherence conditions, and m entries of $\mathcal{L}_0$ are observed with locations sampled uniformly at random, then there exist universal constants $c_0, c_1, c_2 > 0$ such that if

$$m \geq c_0 \mu r n_{(1)} n_3 (\log(n_{(1)} n_3))^2,$$

$\mathcal{L}_0$ is the unique minimizer to the convex optimization problem with probability at east $1 - c_1 (n_{(1)} n_3)^{-c_2}$.

Low Rank Tensor Recovery

Theorem

(Tensor Robust PCA): Suppose $\mathcal{L}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ obeys tensor incoherence conditions and joint incoherence condition and $\mathcal{E}_0$ has support uniformly distributed with probability γ. Then, there exist universal constants $c_1, c_2 > 0$ such that with probability at least $1 - c_1(n_{(1)} n_3)^{-c_2}$, $(\mathcal{L}_0, \mathcal{E}_0)$ is the unique minimizer to the convex optimization problem with $\lambda = 1/\sqrt{n_{(1)} n_3}$, provided that

$$r \leq \frac{c_r n_{(2)}}{\mu(\log(n_{(1)} n_3))^2} \quad \text{and} \quad \gamma \leq c_\gamma$$

where c_r and c_γ are two positive constants.

Convex Optimization Problem

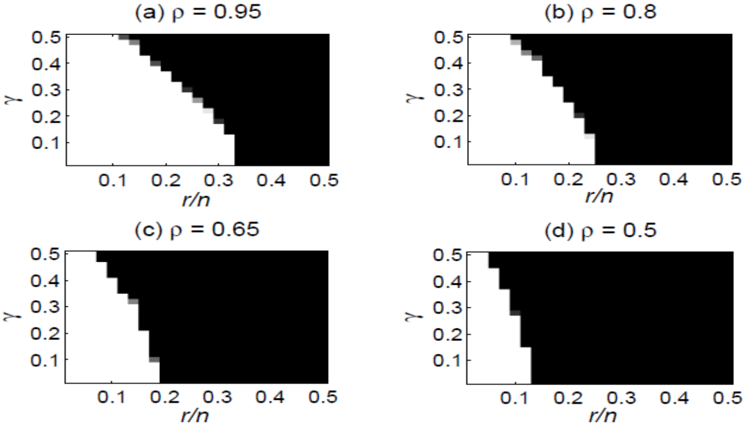
Input: $\mathcal{X}$, Ω and λ.

Initialize: $\mathcal{L}^0 = \mathcal{E}^0 = \mathcal{Y}^0 = 0$, $\rho = 1.1$, $\mu^0 = $ 1e-4, $\mu_{\max} = $ 1e8.

- WHILE not converged
  1. Update $\mathcal{L}^{k+1}$ by
  $$\min_{\mathcal{L}} \|\mathcal{L}\|_{\text{TNN}} + \frac{\mu^k}{2}\left\|\mathcal{L} + \mathcal{E}^k - \mathcal{X} + \frac{\mathcal{Y}^k}{\mu^k}\right\|_F^2;$$
  2. Update $P_\Omega(\mathcal{E}^{k+1})$ by
  $$\min_{\mathcal{E}} \lambda\|P_\Omega(\mathcal{E})\|_1 + \frac{\mu^k}{2}\left\|P_\Omega\left(\mathcal{E} + \mathcal{L}^{k+1} - \mathcal{X} + \frac{\mathcal{Y}^k}{\mu^k}\right)\right\|_F^2;$$
  3. Update $P_{\Omega^c}(\mathcal{E}^{k+1})$ by $P_{\Omega^c}(\mathcal{E}^{k+1}) = P_{\Omega^c}(\mathcal{X} - \mathcal{L}^{k+1} - \mathcal{Y}^k/\mu^k)$;
  4. Update the multipliers $\mathcal{Y}^{k+1}$ by $\mathcal{Y}^{k+1} = \mathcal{Y}^k + \mu^k(\mathcal{L}^{k+1} + \mathcal{E}^{k+1} - \mathcal{X})$;
  5. Update μ^{k+1} by $\mu^{k+1} = \min(\rho\mu^k, \mu_{\max})$;
  6. Check the convergence condition
- ENDWHILE

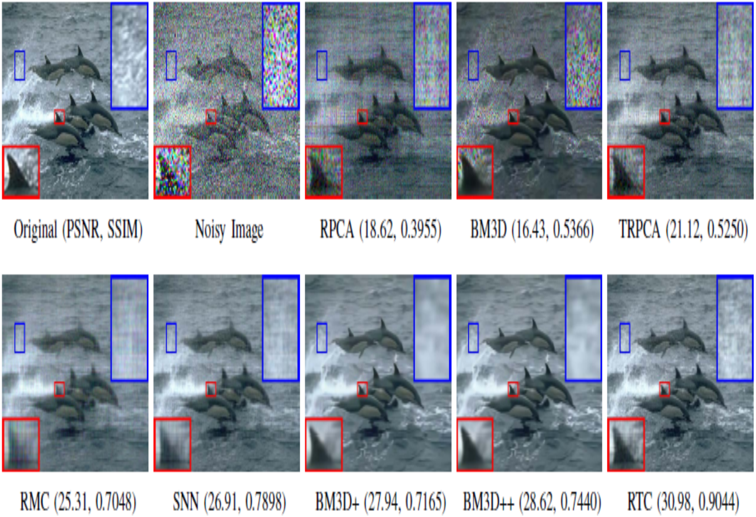
Output: $\mathcal{L}$

Phase Transition



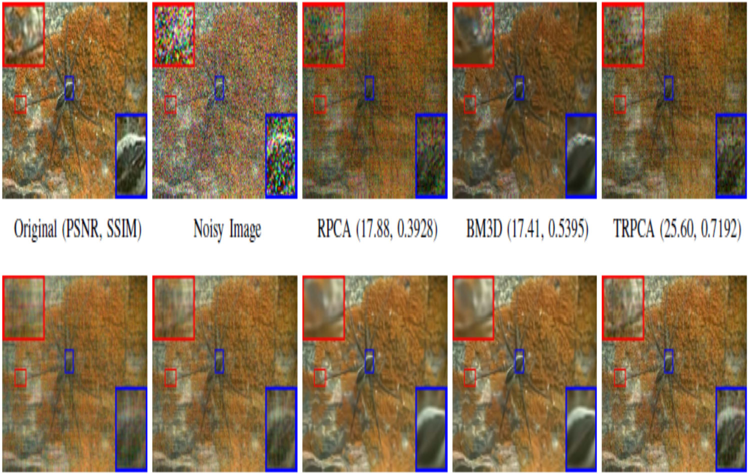
ρ (data observation) and γ (data corruption)

Application: Completion and Denoising



$\rho = 70\%$ (data observation) and $\gamma = 30\%$ (data corruption)

Application: Completion and Denoising



Original (PSNR, SSIM) | Noisy Image | RPCA (17.88, 0.3928) | BM3D (17.41, 0.5395) | TRPCA (25.60, 0.7192)

RMC (24.03, 0.6380) | SNN (25.47, 0.7399) | BM3D+ (25.60, 0.6279) | BM3D++ (26.18, 0.6668) | RTC (28.39, 0.8593)

$\rho = 70\%$ (data observation) and $\gamma = 30\%$ (data corruption)

Application: Completion and Denoising

Method	$\rho = 90\%$						$\rho = 70\%$					
	$\gamma = 10\%$		$\gamma = 20\%$		$\gamma = 30\%$		$\gamma = 10\%$		$\gamma = 20\%$		$\gamma = 30\%$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
RPCA	27.67	0.8535	27.30	0.8367	26.88	0.8122	21.69	0.5609	20.62	0.4744	19.64	0.4081
BM3D	25.42	0.7766	24.99	0.7649	24.59	0.7559	17.80	0.5793	17.67	0.5740	17.55	0.5690
TRPCA	31.40	0.9370	30.51	0.9153	29.63	0.8851	23.82	0.6763	22.52	0.5810	21.31	0.4979
RMC	28.11	0.8552	27.82	0.8423	27.53	0.8276	26.33	0.7865	26.09	0.7736	25.84	0.7599
SNN	30.13	0.9128	29.60	0.8972	29.11	0.8797	27.74	0.8426	27.35	0.8248	26.97	0.8063
BM3D+	30.72	0.8289	30.51	0.8245	30.28	0.8203	29.75	0.8060	29.45	0.7993	29.18	0.7935
BM3D++	30.94	0.8338	30.74	0.8297	30.52	0.8257	30.42	0.8221	30.10	0.8152	29.82	0.8093
RTC	**33.03**	**0.9566**	**32.10**	**0.9400**	**31.27**	**0.9185**	**31.30**	**0.9296**	**30.58**	**0.9091**	**29.91**	**0.8831**

For RPCA and RMC, we apply them on each channel with $\lambda = 1/\sqrt{n_1}$; For SNN, unfolding with three parameters suggested in the literature; For TRPCA, $\lambda = 1/\sqrt{n_1 n_3}$; For BM3D, standard denoising method using nonlocal information; For BM3D+, two-step method with BM3D and image completion using HaLRTC (tensor unfolding to matrix); For BM3D++, two-step method with BM3D and image completion using TNMM.

Video Background Modeling

Background: Low-Tubal-Rank Component and Moving Objects: Sparse Component

Bootstrap (Resolution: 120×140)

ρ	Method				
	RPCA	RMC	SNN	TRPCA	RTC
100%	0.7190	0.7190	0.6849	**0.7548**	**0.7548**
80%	0.6783	0.7084	0.6417	0.7441	**0.7529**
50%	NA	0.6869	0.4081	NA	**0.7476**
20%	NA	0.5492	NA	NA	**0.7208**

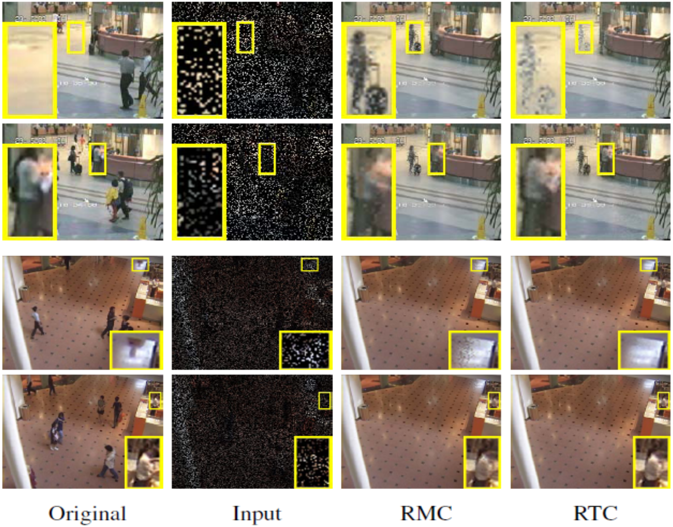
Hall (Resolution: 144×176)

ρ	Method				
	RPCA	RMC	SNN	TRPCA	RTC
100%	0.6296	0.6296	0.5539	**0.6345**	**0.6345**
80%	0.5697	0.6247	0.5516	0.4495	**0.6340**
50%	NA	0.6110	0.4720	NA	**0.6218**
20%	NA	0.5650	NA	NA	**0.6158**

Shopping Mall (Resolution: 256×320)

ρ	Method				
	RPCA	RMC	SNN	TRPCA	RTC
100%	0.7466	0.7466	0.7361	0.7708	**0.7708**
80%	0.7424	0.7422	0.7036	0.7559	**0.7675**
50%	NA	0.7353	0.4770	NA	**0.7614**
20%	NA	0.7176	NA	NA	**0.7496**

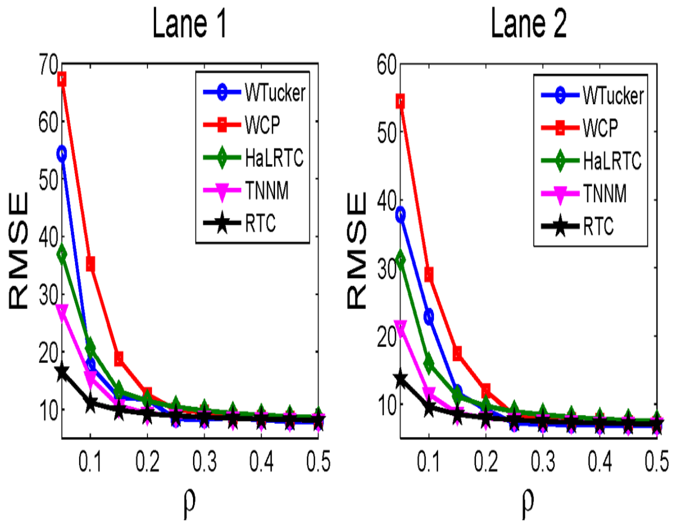
Video Background Modeling



Traffic Data Estimation

- Traffic flow data such as traffic volumes, occupancy rats and flow speeds are usually contaminated by missing values and outliers due to the hardware or software malfunctions.
- Performance Measurement System (PeMS) pems.dot.ca.gov
- Third-order tensor (day) x (time) x (week) of traffic volume

Traffic Data Estimation



The Correction Model

The Corrected Model

Issue: The nuclear norm minimization of a matrix may be challenged under general sampling distribution. Salakhutdinov et al.[3] showed that when certain rows and/or columns were sampled with high probability, the matrix nuclear norm minimization may fail in the sense that the number of observations required for recovery was much more than the setting of most matrix completion problems.

Miao et al. proposed a rank-corrected model for low-rank matrix recovery with fixed basis coefficients[4].

[3] R. Salakhutdinov and N. Srebro. Collaborative filtering in a non-uniform world: Learning with the weighted trace norm. In Adv. Neural Inform. Process. Syst., pages 20562064, 2010.

[4] W. Miao, S. Pan, and D. Sun. A rank-corrected procedure for matrix completion with fixed basis coefficients. Math. Program., 159(1):289338, 2016.

The Corrected Method

For any given index set
$\Omega \subset \{1, 2, \ldots, n_1\} \times \{1, 2, \ldots, n_2\} \times \{1, 2, \ldots, n_3\}$, we define the sampling operator $\mathfrak{D}_\Omega : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^{|\Omega|}$ by

$$\mathfrak{D}_\Omega(\mathcal{X}) = (\langle \mathcal{E}_{ijk}, \mathcal{X} \rangle)^T_{(i,j,k)\in\Omega},$$

where $|\Omega|$ denotes the number of entries in Ω.

Let $\mathcal{X}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be an unknown true tensor. The observed model can be described in the following form:

$$\mathbf{y} = \mathfrak{D}_\Omega(\mathcal{X}_0) + \sigma\varepsilon,$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_m)^T \in \mathbb{R}^m$ and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m)^T \in \mathbb{R}^m$ are the observation vector and the noise vector, respectively, ε_i are the independent and identically distributed (i.i.d.) noises with $\mathbb{E}(\varepsilon_i) = 0$ and $\mathbb{E}(\varepsilon_i^2) = 1$, and $\sigma > 0$ controls the magnitude of noise.

The Corrected Method

Assumption: Each entry is sampled with positive probability, i.e., there exists a positive constant $\kappa_1 \geq 1$ such that

$$p_{ijk} \geq \frac{1}{\kappa_1 n_1 n_2 n_3}.$$

It implies

$$\mathbb{E}(\langle \mathcal{E}, \mathcal{X} \rangle^2) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} p_{ijk} x_{ijk}^2 \geq \frac{1}{\kappa_1 n_1 n_2 n_3} \|\mathcal{X}\|_F^2.$$

The Corrected Method

In the matrix case, the nuclear norm penalization may fail when some columns or rows are sampled with very high probability. In the third-order tensor, we also need to avoid this case that each fiber is sampled with very high probability. Let

$$R_{jk} = \sum_{i=1}^{n_1} p_{ijk}, \quad C_{ik} = \sum_{j=1}^{n_2} p_{ijk}, \quad T_{ij} = \sum_{k=1}^{n_3} p_{ijk},$$

Assumption: There exists a positive constant $\kappa_2 \geq 1$ such that

$$\max_{i,j,k}\{R_{jk}, C_{ik}, T_{ij}\} \leq \frac{\kappa_2}{\min\{n_1, n_2, n_3\}}.$$

The Corrected Method

$$\min \frac{1}{2m}\|\mathbf{y} - \mathfrak{D}_\Omega(\mathcal{X})\|^2 + \mu\Big(\|\mathcal{X}\|_{TNN} - \langle F(\mathcal{X}_m), \mathcal{X}\rangle\Big)$$
$$\text{s.t. } \|\mathcal{X}\|_\infty \le c,$$

where the spectral function $F : \mathbb{R}^{n_1\times n_2\times n_3} \to \mathbb{R}^{n_1\times n_2\times n_3}$ is given as follows: $F(\mathcal{X}_m) := \mathcal{U} * \Sigma * \mathcal{V}^H$, associated with

$$\Sigma = \text{ifft}(\widehat{\mathcal{M}}, [\,], 3) \quad \text{with} \quad \widehat{M}^{(i)} = f(\widehat{S}^{(i)}) := \text{Diag}\Big(f\Big(\text{diag}(\widehat{S}^{(i)})\Big)\Big),$$

f is defined by

$$f_i(\mathbf{x}) := \begin{cases} \phi\left(\frac{x_i}{\|\mathbf{x}\|_\infty}\right), & \text{if } \mathbf{x} \ne 0, \\ 0, & \text{otherwise,} \end{cases}$$

and the scalar function $\phi : \mathbb{R} \to \mathbb{R}$, is defined by

$$\phi(z) = (1+\varepsilon^\tau)\frac{|z|^\tau}{|z|^\tau + \varepsilon^\tau}.$$

The Corrected Method

- The correction function F is used to get a lower tubal rank solution.
- For the small singular values of the frontal slices in the Fourier domain, we would like to penalize more in the correction procedure. Then these small singular values will approximate to zero in the next correction procedure. In this case, the model can generate a lower tubal rank solution by the correction method.

The Corrected Method

Theorem

Suppose the two assumptions hold. Let $\tau > 1$ be given. Then, for $m \geq \widetilde{n} n_3 \log^3(n_1 n_3 + n_2 n_3)/\kappa_2$, there exists constants $\widetilde{C}, C_1 > 0$ such that

$$\frac{\|\mathcal{X}_c - \mathcal{X}_0\|_F^2}{n_1 n_2 n_3} \leq \frac{n_1 n_2 \kappa_1^2 \kappa_2 \log((n_1 + n_2) n_3)}{m \widetilde{n}} \left(32 C_1^2 \left(\frac{\sqrt{2r}}{\tau} + \alpha_m \right)^2 \tau^2 \sigma^2 + 4096 \widetilde{C} c^2 \left(\frac{\tau(\sqrt{2r} + \alpha_m)}{\tau - 1} \right)^2 \right)$$

with probability at least $1 - \frac{2}{n_1 + n_2 + n_3}$, where $\alpha_m = \|\widetilde{\mathcal{U}}_1 * \widetilde{\mathcal{V}}_1^T - F(\mathcal{X}_m)\|_F$.

Here $\widetilde{\mathcal{U}}_1$, $\widetilde{\mathcal{V}}_1^T$ are the associated orthogonal tensors in t-SVD of $\mathcal{X}_0$.

The Symmetric Gauss-Seidel Multi-Block ADMM

Let $\mathfrak{U}(\mathcal{X}) := \{\mathcal{X} | \|\mathcal{X}\|_\infty \leq c\}$. By introducing $\mathbf{z} = \mathbf{y} - \mathfrak{D}_\Omega(\mathcal{X})$ and $\mathcal{X} = \mathcal{S}$, the model is given by

$$\begin{array}{ll} \min & \frac{1}{2m}\|\mathbf{z}\|^2 + \mu\Big(\|\mathcal{X}\|_{TNN} - \langle F(\mathcal{X}_m), \mathcal{X}\rangle\Big) + \delta_{\mathfrak{U}}(\mathcal{S}) \\ \text{s.t.} & \mathbf{z} = \mathbf{y} - \mathfrak{D}_\Omega(\mathcal{X}), \mathcal{X} = \mathcal{S}. \end{array}$$

Since the TNN is the dual norm of the tensor spectral norm, its Lagrangian dual is given as follows:

$$\begin{array}{ll} \max\limits_{\mathbf{u}, \mathcal{W}} & -\frac{m}{2}\|\mathbf{u}\|^2 + \langle \mathbf{u}, \mathbf{y}\rangle - \delta^*_{\mathfrak{U}}(-\mathcal{W}) \\ \text{s.t.} & \|\mu F(\mathcal{X}_m) + \mathfrak{D}^*_\Omega(\mathbf{u}) + \mathcal{W}\| \leq \mu. \end{array}$$

The Symmetric Gauss-Seidel Multi-Block ADMM

Let $\mathcal{Z} := \mu F(\mathcal{X}_m) - \mathfrak{D}^*_\Omega(\mathbf{u}) + \mathcal{W}$ and $\mathfrak{X}(\mathcal{X}) := \{\mathcal{X} | \|\mathcal{X}\| \leq \mu\}$.

$$\begin{array}{ll} \min\limits_{\mathbf{u},\mathcal{W},\mathcal{Z}} & \frac{m}{2}\|\mathbf{u}\|^2 - \langle \mathbf{u}, \mathbf{y} \rangle + \delta^*_{\mathfrak{U}}(-\mathcal{W}) + \delta_{\mathfrak{X}}(\mathcal{Z}) \\ \text{s.t.} & \mathcal{Z} = \mu F(\mathcal{X}_m) + \mathfrak{D}^*_\Omega(\mathbf{u}) + \mathcal{W}. \end{array}$$

The augmented Lagrangian function is defined by

$$\begin{aligned} L(\mathbf{u}, \mathcal{W}, \mathcal{Z}, \mathcal{X}) \quad := \quad & \frac{m}{2}\|\mathbf{u}\|^2 - \langle \mathbf{u}, \mathbf{y} \rangle + \delta^*_{\mathfrak{U}}(-\mathcal{W}) + \delta_{\mathfrak{X}}(\mathcal{Z}) \\ & - \langle \mathcal{X}, \mathcal{Z} - \mu F(\mathcal{X}_m) - \mathfrak{D}^*_\Omega(\mathbf{u}) - \mathcal{W} \rangle \\ & + \frac{\beta}{2}\|\mathcal{Z} - \mu F(\mathcal{X}_m) - \mathfrak{D}^*_\Omega(\mathbf{u}) - \mathcal{W}\|_F^2, \end{aligned}$$

where $\beta > 0$ is the penalty parameter and $\mathcal{X}$ is the Lagrangian multiplier.

The Symmetric Gauss-Seidel Multi-Block ADMM

The iteration system of sGS-ADMM is described as follows:

$$
\begin{aligned}
\mathbf{u}^{k+\frac{1}{2}} &= \arg\min_{\mathbf{u}} \left\{ L(\mathbf{u}, \mathcal{W}^k, \mathcal{Z}^k, \mathcal{X}^k) \right\}, \\
\mathcal{W}^{k+1} &= \arg\min_{\mathcal{W}} \left\{ L(\mathbf{u}^{k+\frac{1}{2}}, \mathcal{W}, \mathcal{Z}^k, \mathcal{X}^k) \right\}, \\
\mathbf{u}^{k+1} &= \arg\min_{\mathbf{u}} \left\{ L(\mathbf{u}, \mathcal{W}^{k+1}, \mathcal{Z}^k, \mathcal{X}^k) \right\}, \\
\mathcal{Z}^{k+1} &= \arg\min_{\mathcal{Z}} \left\{ L(\mathbf{u}^{k+1}, \mathcal{W}^{k+1}, \mathcal{Z}, \mathcal{X}^k) \right\}, \\
\mathcal{X}^{k+1} &= \mathcal{X}^k - \gamma\beta\Big(\mathcal{Z}^{k+1} - \mu F(\mathcal{X}_m) - \mathfrak{D}_\Omega^*(\mathbf{u}^{k+1}) - \mathcal{W}^{k+1} \Big),
\end{aligned}
$$

where $\gamma \in (0, (1+\sqrt{5})/2)$ is the step-length.

The Symmetric Gauss-Seidel Multi-Block ADMM

The optimal solution with respect to $\mathbf{u}$ is given explicitly by

$$\mathbf{u} = \frac{1}{m+\beta}\Big(\mathbf{y} - \mathfrak{D}_\Omega\Big(\mathcal{X} + \beta(\mu F(\mathcal{X}_m) + \mathcal{W} - \mathcal{Z})\Big)\Big).$$

The optimal solution with respect to $\mathcal{W}$ is given explicitly by

$$\begin{aligned}\mathcal{W}^{k+1} = & \quad -\mathrm{Prox}_{\frac{1}{\beta}\delta^*_{\mathfrak{U}}}\Big(\tfrac{1}{\beta}\mathcal{X}^k + \mu F(\mathcal{X}_m) + \mathfrak{D}^*_\Omega(\mathbf{u}^{k+\frac{1}{2}}) - \mathcal{Z}^k\Big) \\ = & \quad -\Big(\tfrac{1}{\beta}\mathcal{X}^k + \mu F(\mathcal{X}_m) + \mathfrak{D}^*_\Omega(\mathbf{u}^{k+\frac{1}{2}}) - \mathcal{Z}^k\Big) \\ & +\tfrac{1}{\beta}\mathrm{Prox}_{\beta\delta_{\mathfrak{U}}}\Big(\beta\Big(\tfrac{1}{\beta}\mathcal{X}^k + \mu F(\mathcal{X}_m) + \mathfrak{D}^*_\Omega(\mathbf{u}^{k+\frac{1}{2}}) - \mathcal{Z}^k\Big)\Big).\end{aligned}$$

The Symmetric Gauss-Seidel Multi-Block ADMM

For the subproblem with respect to $\mathcal{Z}$, it is a projection onto $\mathfrak{X}$, which has a closed-form solution.

Theorem

For any $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\rho > 0$, let $\mathcal{Y} = \mathcal{U} * \mathcal{S} * \mathcal{V}^H$ be the t-SVD. Then the optimal solution $\mathcal{X}^*$ of the following problem

$$\min_{\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}} \{\|\mathcal{X} - \mathcal{Y}\|_F^2, \|\mathcal{X}\| \leq \rho\}$$

is given by

$$\mathcal{X}^* = \mathcal{U} * \mathcal{S}_\rho * \mathcal{V}^H,$$

where $\mathcal{S}_\rho = \texttt{ifft}(\min\{\overline{\mathcal{S}}, \rho\}, [], 3)$.

The Symmetric Gauss-Seidel Multi-Block ADMM

The optimal solution with respect to $\mathcal{Z}$ in (1) is given by

$$\begin{aligned} \mathcal{Z}^{k+1} &= \mathrm{Prox}_{\frac{1}{\beta}\delta_{\mathfrak{X}}}\Big(\mu F(\mathcal{X}_m) + \mathfrak{D}_{\Omega}^{*}(\mathbf{u}^{k+1}) + \mathcal{W}^{k+1} + \tfrac{1}{\beta}\mathcal{X}^{k+1}\Big) \\ &= \mathcal{U}^{k+1} * \mathcal{S}_{\mu}^{k+1} * (\mathcal{V}^{k+1})^{T}, \end{aligned}$$

where $\mathcal{S}_{\mu}^{k+1} = \mathrm{ifft}(\min\{\overline{\mathcal{S}^{k+1}}, \mu\}, [\,], 3)$ and

$$\mu F(\mathcal{X}_m) - \mathfrak{D}_{\Omega}^{*}(\mathbf{u}^{k+1}) + \mathcal{W}^{k+1} + \frac{1}{\beta}\mathcal{X}^{k+1} = \mathcal{U}^{k+1} * \mathcal{S}^{k+1} * (\mathcal{V}^{k+1})^{T}.$$

The Symmetric Gauss-Seidel Multi-Block ADMM

Theorem

The optimal solution set is nonempty and compact.

Only two blocks with respect to $\mathcal{W}, \mathcal{Z}$ are nonsmooth and other blocks are quadratic.

Theorem

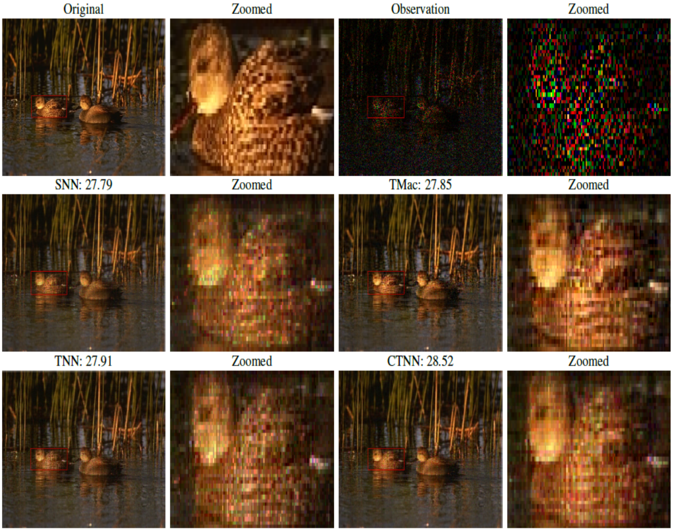
Suppose that $\beta > 0$ *and* $\gamma \in (0, (1+\sqrt{5})/2)$. *Let the sequence* $\{(\mathcal{W}^k, \mathbf{u}^k, \mathcal{Z}^k, \mathcal{X}^k)\}$ *be generated by the algorithm. Then* $\{(\mathcal{W}^k, \mathbf{u}^k, \mathcal{Z}^k)\}$ *converges to an optimal solution and* $\{\mathcal{X}^k\}$ *converges to an optimal solution of the dual problem.*

Numerical Examples

Table: Relative errors of the TNN and CTNN with different tensors, tubal ranks, and sampling ratios for low-rank tensor recovery.

Tensor	r	σ	SR	TNN	CTNN-1	CTNN-2	CTNN-3
$30 \times 40 \times 50$	2	0.1	0.15	5.12e-1	3.57e-1	1.91e-1	4.09e-2
			0.20	2.30e-1	1.63e-2	1.33e-2	1.33e-2
			0.30	1.69e-2	1.01e-2	1.01e-2	1.01e-2
$30 \times 40 \times 50$	3	0.01	0.20	5.46e-1	4.58e-1	3.82e-1	3.07e-1
			0.25	3.19e-1	1.51e-2	1.29e-3	1.26e-3
			0.30	8.61e-2	1.08e-3	1.04e-3	1.04e-3
$50 \times 50 \times 50$	4	0.01	0.15	5.17e-1	3.70e-1	2.12e-1	2.83e-2
			0.20	2.29e-1	1.31e-3	1.08e-3	1.08e-3
			0.25	2.31e-3	9.03e-4	9.03e-4	9.03e-4
$100 \times 100 \times 50$	3	0.05	0.10	3.73e-1	1.44e-2	5.96e-3	5.96e-3
			0.15	1.08e-2	4.45e-3	4.45e-3	4.45e-3
			0.20	6.04e-3	3.93e-3	3.93e-3	3.93e-3
$100 \times 100 \times 50$	6	0.01	0.15	5.37e-1	3.88e-1	2.38e-1	5.79e-2
			0.20	2.41e-1	1.36e-3	1.13e-3	1.13e-3
			0.25	2.36e-3	9.68e-4	9.68e-4	9.68e-4
$100 \times 100 \times 100$	4	0.1	0.10	5.98e-1	4.75e-1	3.63e-1	2.35e-1
			0.15	1.73e-1	6.41e-3	5.83e-3	5.83e-3
			0.20	1.05e-2	4.92e-3	4.92e-3	4.92e-3

Color Image



Original
Zoomed
Observation
Zoomed
SNN: 27.79
Zoomed
TMac: 27.85
Zoomed
TNN: 27.91
Zoomed
CTNN: 28.52
Zoomed

Color Image


Original
Zoomed
Observation
Zoomed
SNN: 26.28
Zoomed
TMac: 26.55
Zoomed
TNN: 26.10
Zoomed
CTNN: 27.02
Zoomed

Other t-SVDs

Revisit t-SVD

We use $\tilde{\mathcal{X}} \in \mathbb{C}^{m_1 \times m_2 \times m_3}$ to represent the discrete Fourier transform of $\mathcal{X} \in \mathbb{C}^{m_1 \times m_2 \times m_3}$ along each tube, i.e., $\tilde{\mathcal{X}} = \mathrm{fft}(\mathcal{X}, [\,], 3)$. The block circulant matrix is defined as

$$\mathrm{bcirc}(\mathcal{X}) := \begin{bmatrix} \mathbf{X}^{(1)} & \mathbf{X}^{(m_3)} & \cdots & \mathbf{X}^{(2)} \\ \mathbf{X}^{(2)} & \mathbf{X}^{(1)} & \cdots & \mathbf{X}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}^{(m_3)} & \mathbf{X}^{(m_3-1)} & \cdots & \mathbf{X}^{(1)} \end{bmatrix}.$$

The block diagonal matrix and the corresponding inverse operator are defined as

$$\mathrm{bdiag}(\mathcal{X}) := \begin{bmatrix} \mathbf{X}^{(1)} & & & \\ & \mathbf{X}^{(2)} & & \\ & & \ddots & \\ & & & \mathbf{X}^{(m_3)} \end{bmatrix},$$

$$\mathrm{unbdiag}(\mathrm{bdiag}(\mathcal{X})) = \mathcal{X}.$$

Revisit t-SVD

Theorem

$$bdiag(\tilde{\mathcal{X}}) = (\mathbf{F}_{m_3} \otimes \mathbf{I}_{m_1})bcirc(\mathcal{X})(\mathbf{F}_{m_3}^H \otimes \mathbf{I}_{m_2}),$$

where $\otimes$ *denotes the Kronecker product,* $\mathbf{F}_{m_3}$ *is an* $m_3 \times m_3$ *DFT matrix and* $\mathbf{I}_m$ *is an* $m \times m$ *identity matrix.*

Revisit t-SVD

The unfold and fold operators in t-SVD are defined as

$$\text{unfold}(\mathcal{X}) := \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \\ \vdots \\ \mathbf{X}^{(m_3)} \end{bmatrix}, \quad \text{fold}(\text{unfold}(\mathcal{X})) = \mathcal{X}.$$

Given $\mathcal{X} \in \mathbb{C}^{m_1 \times m_2 \times m_3}$ and $\mathcal{Y} \in \mathbb{C}^{m_2 \times m_4 \times m_3}$, the t-product $\mathcal{X} * \mathcal{Y}$ is a third-order tensor of size $m_1 \times m_4 \times m_3$

$$\mathcal{Z} = \mathcal{X} * \mathcal{Y} := \text{fold}(\text{bcirc}(\mathcal{X})\text{unfold}(\mathcal{Y})).$$

Since the corresponding block circulant matrices can be diagonalized by DFT, the DFT based t-SVD can be efficiently implemented via fast Fourier transform (fft).

Cosine-Transform based t-SVD

The first work is given by E. Kernfeld, M. Kilmer and S. Aeron, Tensor tensor products with invertible linear transforms, LAA, Vol 485, pp. 545-570 (2015).

We define the shift of tensor $\mathcal{A} = \text{fold}\begin{bmatrix} \mathbf{A}^{(1)} \\ \mathbf{A}^{(2)} \\ \vdots \\ \mathbf{A}^{(m_3)} \end{bmatrix}$ as

$$\sigma(\mathcal{A}) = \text{fold}\begin{bmatrix} \mathbf{A}^{(2)} \\ \mathbf{A}^{(3)} \\ \vdots \\ \mathbf{A}^{(m_3)} \\ \mathbf{O} \end{bmatrix}.$$

Any tensor $\mathcal{X}$ can be uniquely divided into $\mathcal{A} + \sigma(\mathcal{A})$.

Cosine-Transform based t-SVD

We use $\bar{\mathcal{X}} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$ to represent the DCT along each tube of $\mathcal{X}$, i.e., $\bar{\mathcal{X}} = \mathrm{dct}(\mathcal{X}, [\,], 3) = \mathrm{dct}(\mathcal{A} + \sigma(\mathcal{A}), [\,], 3)$. We define the block Toeplitz matrix of $\mathcal{A}$ as

$$\mathrm{bt}(\mathcal{A}) := \begin{bmatrix} \mathbf{A}^{(1)} & \mathbf{A}^{(2)} & \cdots & \mathbf{A}^{(m_3-1)} & \mathbf{A}^{(m_3)} \\ \mathbf{A}^{(2)} & \mathbf{A}^{(1)} & \cdots & \mathbf{A}^{(m_3-2)} & \mathbf{A}^{(m_3-1)} \\ \vdots & \vdots & \ddots & \vdots & \\ \mathbf{A}^{(m_3-1)} & \mathbf{A}^{(m_3-2)} & \cdots & \mathbf{A}^{(1)} & \mathbf{A}^{(2)} \\ \mathbf{A}^{(m_3)} & \mathbf{A}^{(m_3-1)} & \cdots & \mathbf{A}^{(2)} & \mathbf{A}^{(1)} \end{bmatrix}.$$

The block Hankel matrix is defined as

$$\mathrm{bh}(\mathcal{A}) := \begin{bmatrix} \mathbf{A}^{(2)} & \mathbf{A}^{(3)} & \cdots & \mathbf{A}^{(m_3)} & \mathbf{O} \\ \mathbf{A}^{(3)} & \mathbf{A}^{(4)} & \cdots & \mathbf{O} & \mathbf{A}^{(m_3)} \\ \vdots & \vdots & \ddots & \vdots & \\ \mathbf{A}^{(m_3)} & \mathbf{O} & \cdots & \mathbf{A}^{(4)} & \mathbf{A}^{(3)} \\ \mathbf{O} & \mathbf{A}^{(m_3)} & \cdots & \mathbf{A}^{(3)} & \mathbf{A}^{(2)} \end{bmatrix}.$$

Cosine-Transform based t-SVD

The block Toeplitz-plus-Hankel matrix of $\mathcal{A}$ is defined as

$$\mathrm{btph}(\mathcal{A}) := \mathrm{bt}(\mathcal{A}) + \mathrm{bh}(\mathcal{A}).$$

The block Toeplitz-plus-Hankel matrix can be diagonalized.

Theorem

$$bdiag(\bar{\mathcal{X}}) = (\mathbf{C}_{m_3} \otimes \mathbf{I}_{m_1}) btph(\mathcal{A}) (\mathbf{C}_{m_3}^T \otimes \mathbf{I}_{m_2}),$$

where $\otimes$ denotes the Kronecker product, $\mathbf{C}_{m_3}$ is an $m_3 \times m_3$ DCT matrix.

Cosine-Transform based t-SVD

Definition: Given $\mathcal{X} \in \mathbb{C}^{m_1 \times m_2 \times m_3}$ and $\mathcal{Y} \in \mathbb{C}^{m_2 \times m_4 \times m_3}$, the t-product $\mathcal{X} * \mathcal{Y}$ is a third-order tensor of size $m_1 \times m_4 \times m_3$

$$\mathcal{Z} = \mathcal{X} * \mathcal{Y} := \mathrm{fold}(\mathrm{btph}(\mathcal{A})\mathrm{unfold}(\mathcal{Y})),$$

where $\mathcal{X} = \mathcal{A} + \sigma(\mathcal{A})$.

Cosine-Transform based t-SVD

Theorem

Given a tensor $\mathcal{X} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, the DCT-based t-SVD of $\mathcal{X}$ is given by

$$\mathcal{X} = \mathcal{U} *_{dct} \mathcal{S} *_{dct} \mathcal{V}^H,$$

where $\mathcal{U} \in \mathbb{R}^{m_1 \times m_1 \times m_3}$,$\mathcal{V} \in \mathbb{R}^{m_2 \times m_2 \times m_3}$ are orthogonal tensors, $\mathcal{S} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$ is a f-diagonal tensor, and $\mathcal{V}^H$ is the tensor transpose of $\mathcal{V}$.

Cosine-Transform based t-SVD

Table: The time cost of t-SVD and DCT-based t-SVD on the random tensors of different size.

size	*100*100*100*	*100*100*400*	*200*200*100*	*400*400*100*
FFT	0.0041	0.0175	0.0176	0.0653
SVD after FFT	0.0818	0.3250	0.3641	1.9015
original t-SVD	0.0859	0.3425	0.3817	1.9668
DCT	0.0042	0.0150	0.0162	0.0601
SVD after DCT	0.0439	0.1649	0.1978	0.8922
new t-SVD	0.0481	0.1799	0.2140	0.9523

Video Examples

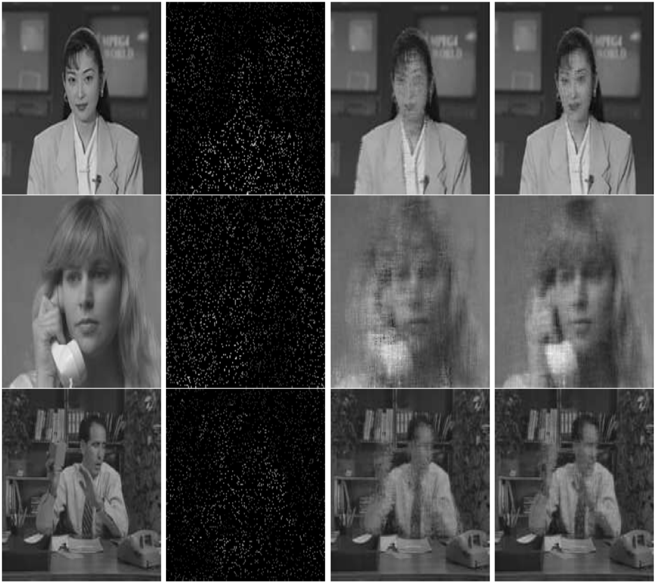
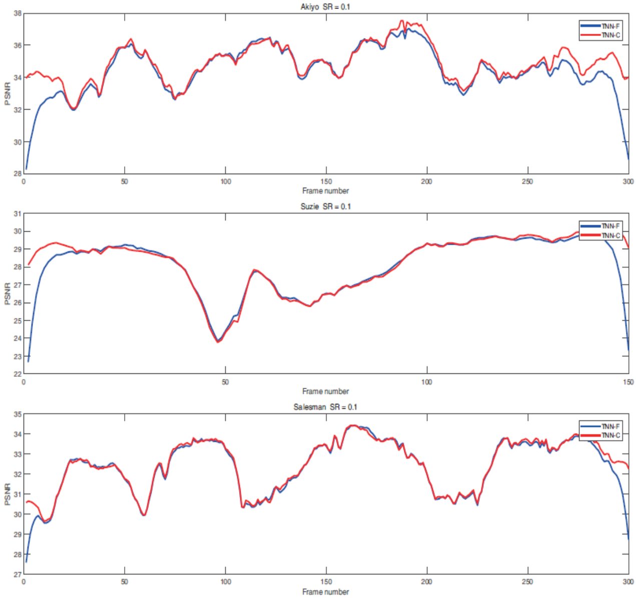


Table: PSNR, SSIM, and time of two methods in video completion. In brackets, they are the time required for transformation and time required for performing SVD. The best results are highlighted in bold.

video		*akiyo*		*suzie*		*salesman*	
SR	metric	TNN-F	TNN-C	TNN-F	TNN-C	TNN-F	TNN-C
0.05	PSNR	32.00	**32.57**	25.50	**26.02**	30.12	**30.22**
	SSIM	0.934	**0.941**	0.681	**0.700**	0.895	**0.897**
	time	156.2	**91.9**	69.6	**40.1**	148.5	**85.6**
0.1	PSNR	34.20	**34.75**	27.73	**27.93**	32.13	**32.29**
	SSIM	0.958	**0.963**	0.759	**0.766**	0.928	**0.931**
	time	141.8	**86.3**	64.5	**39.3**	139.5	**84.9**
0.2	PSNR	37.44	**38.11**	30.29	**30.51**	35.01	**35.20**
	SSIM	0.979	**0.983**	0.838	**0.844**	0.960	**0.961**
	time	145.2	**79.8**	62.5	**37.2**	135.1	**81.3**

Video Examples



Transform-based t-SVD

Fourier-Transform based t-SVD

$$\mathcal{Z} = \mathcal{X} *_{fft} \mathcal{Y} = \text{fold}(\text{bcirc}(\mathcal{X})\text{unfold}(\mathcal{Y}))$$

The DFT based t-SVD can be efficiently implemented via fast Fourier transform (fft).

Cosine-Transform based t-SVD

$$\mathcal{Z} = \mathcal{X} *_{dct} \mathcal{Y} = \text{fold}(\text{btph}(\mathcal{A})\text{unfold}(\mathcal{Y}))$$

The DCT based t-SVD can be efficiently implemented via fast cosine transform (dct).

Transform-based t-SVD

Fourier-Transform based t-SVD

$$\mathcal{Z} = \mathcal{X} *_{fft} \mathcal{Y} = \text{fft}\left[\text{fold}(\text{blockdiag}(\hat{X}_{fft}) \times \text{blockdiag}(\hat{Y}_{fft}))\right]$$

The DFT based t-SVD can be efficiently implemented via fast Fourier transform (fft).

Cosine-Transform based t-SVD

$$\mathcal{Z} = \mathcal{X} *_{dct} \mathcal{Y} = \text{dct}\left[\text{fold}(\text{blockdiag}(\hat{X}_{dct}) \times \text{blockdiag}(\hat{Y})_{dct})\right]$$

The DCT based t-SVD can be efficiently implemented via fast cosine transform (dct).

Transform-based t-SVD

The first work is given by E. Kernfeld, M. Kilmer and S. Aeron, Tensor tensor products with invertible linear transforms, LAA, Vol 485, pp. 545-570 (2015).

We generalize tensor singular value decomposition by using other unitary transform matrices instead of discrete Fourier/cosine transform matrix.

The motivation is that a lower transformed tubal tensor rank may be obtained by using other unitary transform matrices than that by using discrete Fourier/cosine transform matrix, and therefore this would be more effective for robust tensor completion.

Transform-based t-SVD

- Let $\mathbf{\Phi}$ be the unitary transform matrix with $\mathbf{\Phi}\mathbf{\Phi}^H = \mathbf{\Phi}^H\mathbf{\Phi} = \mathbf{I}$.
- $\hat{\mathcal{A}}_{\mathbf{\Phi}}$ represents a third-order tensor obtained via multiplying by $\mathbf{\Phi}$ on all tubes along the third dimension of $\mathcal{A}$.
- The $\mathbf{\Phi}$-product of $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{C}^{n_2 \times n_4 \times n_3}$ is a tensor $\mathcal{C} \in \mathbb{C}^{n_1 \times n_4 \times n_3}$, which is given by

$$\mathcal{C} = \mathcal{A} \diamond_{\mathbf{\Phi}} \mathcal{B} = \mathbf{\Phi}^H \left[\text{fold} \left(\text{blockdiag}(\hat{\mathcal{A}}_{\mathbf{\Phi}}) \times \text{blockdiag}(\hat{\mathcal{B}}_{\mathbf{\Phi}}) \right) \right],$$

  where " $\times$ " denotes the usual matrix product.

Transform-based t-SVD

Theorem

Suppose that $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$. Then $\mathcal{A}$ can be factorized as follows:

$$\mathcal{A} = \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{S} \diamond_{\mathbf{\Phi}} \mathcal{V}^H,$$

where $\mathcal{U} \in \mathbb{C}^{n_1 \times n_1 \times n_3}$, $\mathcal{V} \in \mathbb{C}^{n_2 \times n_2 \times n_3}$ are unitary tensors with respect to $\mathbf{\Phi}$-product, and $\mathcal{S} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is a diagonal tensor.

Transform-based t-SVD

Definition: The transformed tubal multi-rank of a tensor $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is a vector $\boldsymbol{r} \in \mathbb{R}^{n_3}$ with its i-th entry as the rank of the i-th frontal slice of $\hat{\mathcal{A}}_{\boldsymbol{\Phi}}$, i.e., $r_i = \text{rank}(\hat{\mathcal{A}}_{\boldsymbol{\Phi}}^{(i)})$. The transformed tubal tensor rank, denoted as $\text{rank}_{tt}(\mathcal{A})$, is defined as the number of nonzero singular tubes of $\mathcal{S}$, where $\mathcal{S}$ comes from the tt-SVD of $\mathcal{A} = \mathcal{U} \diamond_{\boldsymbol{\Phi}} \mathcal{S} \diamond_{\boldsymbol{\Phi}} \mathcal{V}^H$.

Definition: The transformed tubal nuclear norm of a tensor $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, denoted as $\|\mathcal{A}\|_{\text{TTNN}}$, is the sum of the nuclear norms of all the frontal slices of $\widehat{\mathcal{A}}_{\boldsymbol{\Phi}}$, i.e.,

$$\|\mathcal{A}\|_{\text{TTNN}} = \sum_{i=1}^{n_3} \|\widehat{\mathcal{A}}_{\boldsymbol{\Phi}}^{(i)}\|_*.$$

Transform-based t-SVD

Theorem

For any tensor $\mathcal{X} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, $\|\mathcal{X}\|_{\mathrm{TTNN}}$ *is the convex envelope of the function* $\sum_{i=1}^{n_3} \mathrm{rank}(\widehat{\mathcal{A}}_{\mathbf{\Phi}}^{(i)})$ *on the set* $\{\mathcal{X} \mid \|\mathcal{X}\| \leq 1\}$.

Transform-based t-SVD

$$\min_{\mathcal{L},\mathcal{E}} \ \|\mathcal{L}\|_{\mathsf{TTNN}} + \lambda\|\mathcal{E}\|_1, \ \ \text{s.t.}, \mathcal{P}_\Omega(\mathcal{L}+\mathcal{E}) = \mathcal{P}_\Omega(\mathcal{X}),$$

where λ is a penalty parameter and $\mathcal{P}_\Omega$ is a linear projection such that the entries in the set Ω are given while the remaining entries are missing.

Transform-based t-SVD

Assume that $\text{rank}_{tt}(\mathcal{L}_0) = r$ and its skinny tt-SVD is $\mathcal{L}_0 = \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{S} \diamond_{\mathbf{\Phi}} \mathcal{V}^H$. $\mathcal{L}_0$ is said to satisfy the transformed tensor incoherence conditions with parameter $\mu > 0$ if

$$\max_{i=1,\ldots,n_1} \|\mathcal{U}^H \diamond_{\mathbf{\Phi}} \vec{\boldsymbol{e}}_i\|_F \leq \sqrt{\frac{\mu r}{n_1}},$$

$$\max_{j=1,\ldots,n_2} \|\mathcal{V}^H \diamond_{\mathbf{\Phi}} \vec{e}_j\|_F \leq \sqrt{\frac{\mu r}{n_2}},$$

and

$$\|\mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{V}^H\|_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2 n_3}},$$

where $\vec{\boldsymbol{e}}_i$ and $\vec{\boldsymbol{e}}_j$ are the tensor basis with respect to $\mathbf{\Phi}$.

Transform-based t-SVD

Theorem

Suppose that $\mathcal{L}_0 \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ obeys transformed tensor incoherence conditions, and the observation set Ω is uniformly distributed among all sets of cardinality $m = \rho n_1 n_2 n_3$. Also suppose that each observed entry is independently corrupted with probability γ. Then, there exist universal constants $c_1, c_2 > 0$ such that with probability at least $1 - c_1(n_{(1)} n_3)^{-c_2}$, the recovery of $\mathcal{L}_0$ with $\lambda = 1/\sqrt{\rho n_{(1)} n_3}$ is exact, provided that

$$r \leq \frac{c_r n_{(2)}}{\mu(\log(n_{(1)} n_3))^2} \quad \text{and} \quad \gamma \leq c_\gamma,$$

where c_r and c_γ are two positive constants.

Numerical Illustration

Table: The transformed tubal ranks of randomly generated ten tensors.

Transform/Tensor	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
level-1 Haar	10	20	15	7	3	12	30	5	18	2
level-2 Haar	10	20	15	7	3	12	30	5	18	2
Fourier	28	67	45	23	11	24	84	21	50	6

Numerical Illustration

Table: The relative errors of tensor completion for Tensors #1 and #2 with sampling ratios.

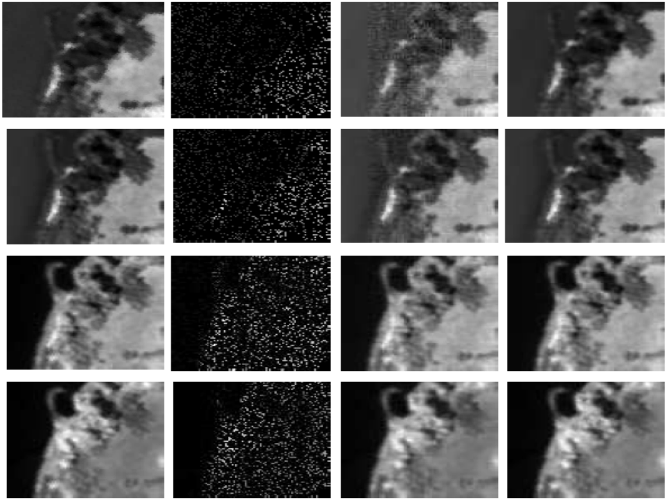
ρ	Tensor #1			Tensor #2		
	level-1 Haar	level-2 Haar	Fourier	level-1 Haar	level-2 Haar	Fourier
0.2	8.22e-2	**2.79e-4**	3.79e-1	4.72e-1	**4.58e-2**	5.68e-1
0.3	3.48e-3	**2.39e-4**	2.29e-1	1.50e-1	**2.46e-4**	4.02e-1
0.4	1.81e-4	**1.57e-4**	1.43e-2	2.58e-3	**1.74e-4**	2.81e-1

Numerical Illustration

Table: The relative errors of robust tensor completion for Tensors #1 and #2 with sampling ratios and noise levels.

ρ	γ	Tensor #1			Tensor #2		
		level-1 Haar	level-2 Haar	Fourier	level-1 Haar	level-2 Haar	Fourier
0.6	0.1	5.47e-3	**6.91e-4**	9.60e-1	9.18e-2	**2.05e-3**	9.78e-1
	0.2	2.70e-2	**1.26e-3**	1.32e0	2.26e-1	**1.41e-2**	1.33e0
	0.3	5.87e-2	**2.79e-3**	1.58e0	3.67e-1	**8.60e-2**	1.59e0
0.8	0.1	**7.87e-5**	5.51e-4	1.01e0	2.63e-2	**7.13e-4**	1.01e0
	0.2	**1.26e-4**	7.55e-4	1.39e0	3.35e-2	**1.67e-3**	1.39e0
	0.3	1.00e-2	**9.96e-4**	1.67e0	1.77e-1	**9.35e-3**	1.66e0

Hyperspectral Image



(a) Original images (b) Observed images (c) Fourier (d) Wavelet

Figure 3: Recovered images with different bands by Fourier and wavelet transforms for hyperspectral data with 15% sampling ratio. Form top to bottom: 20th band, 60th band, 100th band, and 140th band. (a) Original images. (b) Observed images. (c) Recovered images by Fourier transform with PSNR 35.55. (d) Recovered images by wavelet transform with PSNR 40.85.

Hyperspectral Image

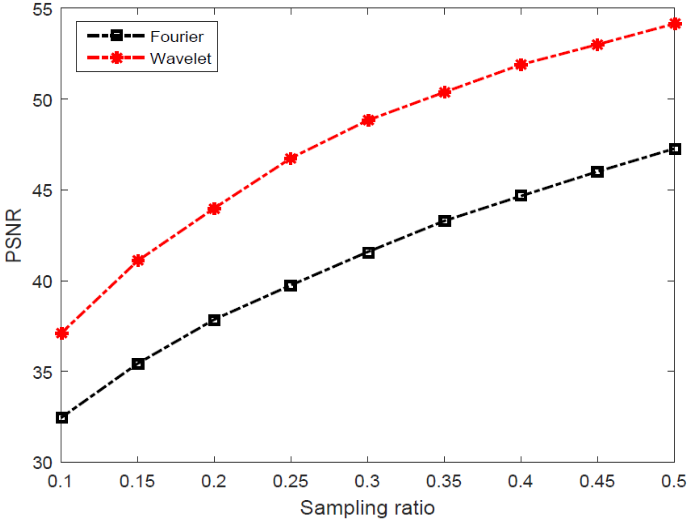


Figure 2: PSNR values versus sampling ratios for Hyperspectral data.

Hyperspectral Image

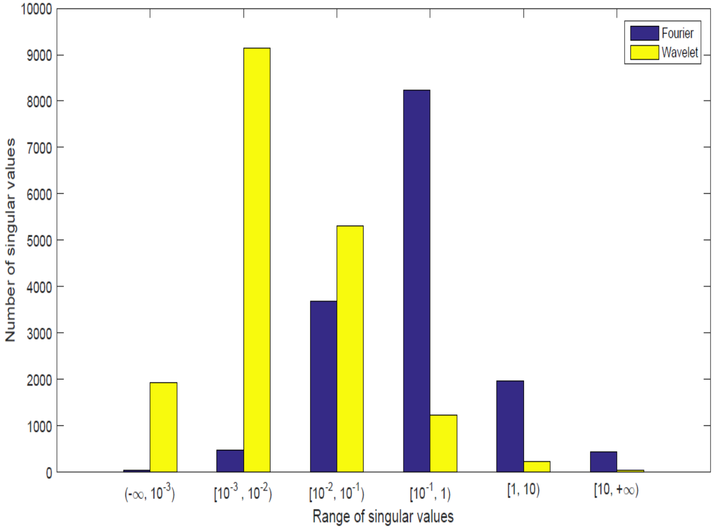


Figure 1: Distributions of singular values of all frontal slices for Hyperspectral data after Fourier and wavelet transforms, repsectively.

Data-Dependent Transform

Theorem

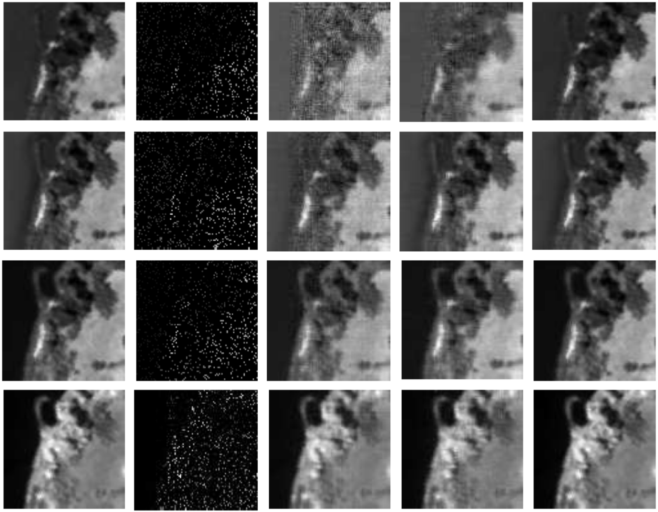
Let $\mathcal{A}$ be a given tensor with size $n_1 \times n_2 \times n_3$ and $\mathbf{A}$ be its unfolding matrix along the third-dimension. Suppose rank$(\mathbf{A}) = k$. Then a global optimal solution of the following problem

$$\min_{\text{rank}(\mathbf{B})=k, \mathbf{\Phi}} \|\mathbf{\Phi}\mathbf{A} - \mathbf{B}\|_F^2$$

$$s.t.\ \mathbf{\Phi}^H\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Phi}^H = \mathbf{I},$$

is given by $\mathbf{\Phi} = \mathbf{U}^H$ and $\mathbf{B} = \mathbf{\Sigma}\mathbf{V}^H$.

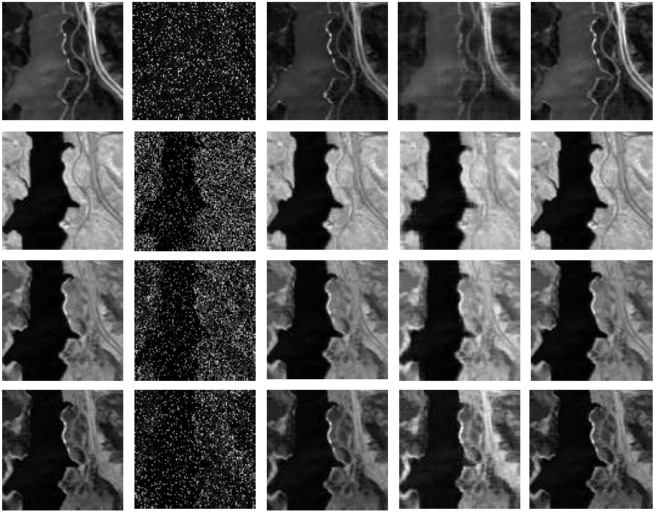
Hyperspectral Image



(a) Original images (b) Observed images (c) Fourier (d) Wavelet (e) U

Fig. 4. Recovered images with different bands by using Fourier, wavelet and unitary transforms in tensor completion for the Samson dataset with 10% sampling ratio. From top to bottom: 15th band, 55th band, 95th band, and 135th band. (a) Original images. (b) Observed images. (c) Recovered images by using Fourier transform with PSNR 32.44. (d) Recovered images by using wavelet transform with PSNR 33.10. (e) Recovered images by using unitary transform with PSNR 48.01.

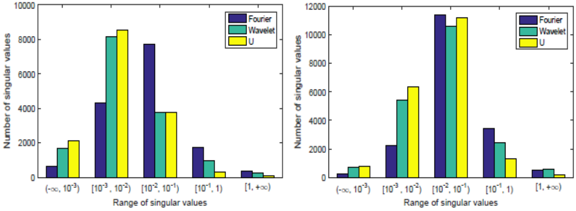
Hyperspectral Image



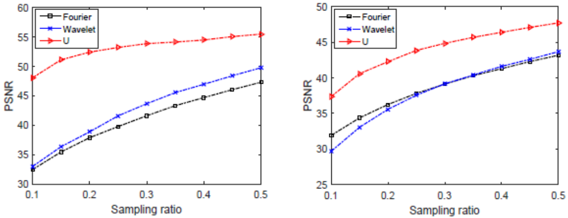
(a) Original images (b) Observed images (c) Fourier (d) Wavelet (e) U

Fig. 5. Recovered images with different bands by using Fourier, wavelet and unitary transforms in robust tensor component principal analysis for Japser Ridge dataset with 60% sampling ratio and 30% corrupted entries. From top to bottom: 10th band, 70th band, 130th band, and 190th band. (a) Original images. (b) Observed images. (c) Recovered images by using Fourier transform with PSNR 33.63. (d) Recovered images by using wavelet transform with PSNR 33.59. (e) Recovered images by using unitary transform with PSNR 37.38.

Hyperspectral Image



(a) Samson (b) Japser Ridge



(a) Samson (b) Japser Ridge

Multi-view Clustering

The original multi-view data with n objects from m views is taken as a tensor. More specifically, each object from different views is twisted into a third-order tensor $D \times 1 \times m$ (D is the total number of features in all the views), then the whole data set can be organized as a tensor $\mathbb{R}^{D\times 1\times m}$

Then the multi-view data can be represented by collecting all objects tensors along the second mode to obtain a tensor $\mathcal{X} \in \mathbb{R}^{D\times n\times m}$.

Multi-view Clustering

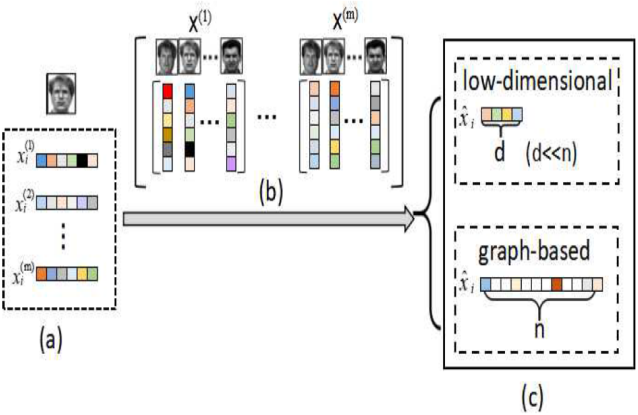


Fig. 1. Demonstration of the multi-view representation learning (MRL). Given an object i with multiple views (a), MRL takes advantage of the multi-view information (b) to learn the new representation of the object (c).

Multi-view Clustering

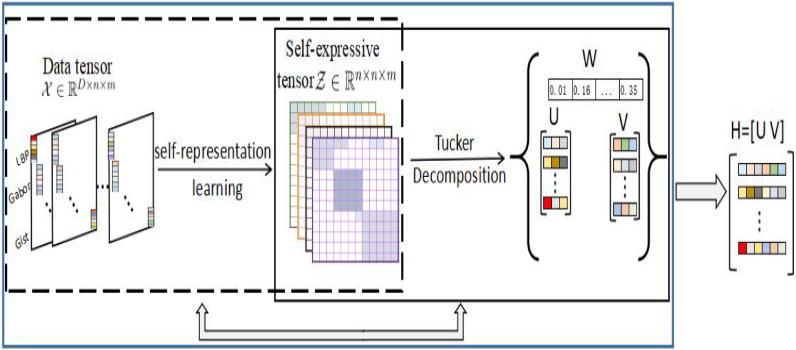


Fig. 2. The framework of tensor-based representation learning for multi-view clustering (**tRLMvC**). $\mathcal{X}$ is the multi-view data tensor, $\mathcal{Z}$ is the self-expressive tensor, $\{\mathrm{U}, \mathrm{V}, \mathrm{W}\}$ is obtained by Tucker decomposition on $\mathcal{Z}$. Among them, $\mathrm{W} \in \mathbb{R}^{1\times m}$ shows the contributions of different views, $\mathrm{U} \in \mathbb{R}^{n\times k_1}$ and $\mathrm{V} \in \mathbb{R}^{n\times k_2}$ indicate the representation of multi-view data in latent space from two aspects, and $\mathrm{H} = [\mathrm{UV}]$ is the final low-dimensional representation.

Latent space with size k_1 and k_2 respectively, which can essentially capture the object cluster structure. $W \in \mathbb{R}^{m\times 1}$ is a column vector recording the contributions of different views.

Multi-view Clustering

DESCRIPTION OF THE TEST DATASETS

Dataset	type	♯views	♯objects	♯clusters
BDGP	text-image	2	2500	5
NUS-WIDE-C5	text-image	2	4000	5
MSRC-v1	scene	6	210	7
Scene-15	scene	3	4485	15
Yale	Face	3	165	15
Extend YaleB	Face	3	640	10
Caltech101-7	Generic Object	6	441	7
COIL	Generic Object	3	1440	20

Multi-view Clustering

TABLE III

CLUSTERING RESULTS OBTAINED BY APPLYING TEN METHODS ON BDGP DATABASE.

Methods	ACC	F-score	Precision	Recall	NMI	AR
LT-MSC[4]	0.945±0.002	0.897±0.001	0.896 ±0.001	0.898 ±0.002	0.860±0.002	0.871 ±0.003
t-SVD-MSC [9]	0.984±0.004	0.947±0.003	0.968±0.004	0.969±0.001	0.947 ±0.003	0.961±0.003
SCMV-3DT [8]	0.990±0.004	0.975±0.003	0.978±0.002	0.975±0.003	0.960±0.000	0.969±0.004
MLAN [3]	0.474±0.004	0.427±0.002	0.328±0.002	0.711±0.004	0.265±0.005	0.165±0.002
MVSC[6]	0.827±0.002	0.695±0.003	0.686±0.004	0.705±0.002	0.600±0.002	0.618 ±0.001
ECMSC[33]	0.485±0.002	0.393±0.003	0.388±0.002	0.398±0.004	0.286 ±0.002	0.239±0.002
DCCAE [14]	0.503±0.002	0.405±0.002	0.403±0.001	0.407±0.003	0.310±0.003	0.2556±0.002
DSemi-NMF [15]	0.593±0.004	0.469±0.001	0.464±0.003	0.475±0.002	0.372±0.002	0.335±0.004
tRLMvC$_K$	0.924±0.001	0.862±0.001	0.862±0.003	0.862±0.002	0.821±0.002	0.823±0.003
tRLMvC$_S$	**0.999**±0.003	**0.999**±0.004	**0.999**±0.001	**0.999**±0.003	**0.998**±0.002	**0.999**±0.004

TABLE IV

CLUSTERING RESULTS OBTAINED BY APPLYING TEN METHODS ON NUS-WIDE-C5 DATABASE.

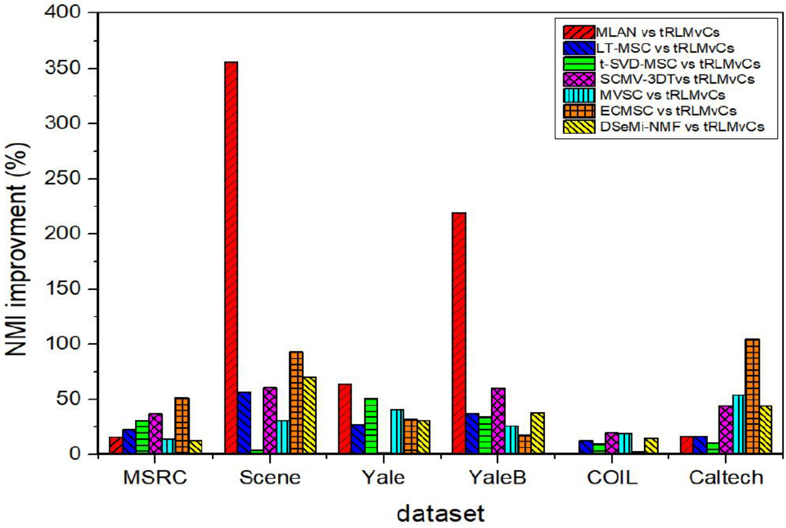
Methods	ACC	F-score	Precision	Recall	NMI	AR
LT-MSC[4]	0.821±0.003	0.686±0.003	0.680±0.004	0.692±0.002	0.597±0.002	0.606±0.003
t-SVD-MSC [9]	0.990±0.003	0.981±0.003	0.981±0.003	0.981±0.003	0.962±0.004	0.976±0.002
SCMV-3DT [8]	0.842±0.002	0.718±0.003	0.710±0.001	0.726±0.002	0.638±0.002	0.648±0.003
MLAN [3]	0.800±0.003	0.653±0.003	0.649±0.003	0.658±0.003	0.552±0.003	0.556±0.003
MVSC[6]	0.750±0.003	0.738±0.003	0.721±0.004	0.755±0.002	0.740 ±0.002	0.687±0.003
ECMSC[33]	0.787±0.002	0.634±0.002	0.633±0.002	0.638±0.002	0.530±0.002	0.545±0.002
DCCAE [14]	0.780±0.001	0.653±0.002	0.649±0.002	0.658±0.003	0.552±0.000	0.566±0.001
DSemi-NMF [15]	0.785±0.003	0.638±0.001	0.631±0.004	0.636±0.003	0.528±0.004	0.542±0.001
tRLMvC$_K$	0.933±0.003	0.900±0.006	0.912±0.002	0.917±0.005	0.900±0.002	0.903±0.006
tRLMvC$_S$	**0.995**±0.002	**0.990**±0.001	**0.990**±0.001	**0.990**±0.001	**0.978**±0.005	**0.987**±0.001

Multi-view Clustering

Image								
t-SVD-MSC	toy	sun	bird	toy	bird	tower	sun	toy
tRLMvC$_S$	bird		food		sun		tower	

Fig. 6. The image examples from bi-view NUS-WIDE-C5 dataset with right clustering results obtained by tRLMvC$_S$, while the best baseline gives the wrong results.

Multi-view Clustering



Concluding Remarks

- More and more applications involving tensor data
- Theory and Algorithms to be studied