Robust Tensor Completion and its Applications

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

29 August 2018

Matrix

- Matrix (second-order tensor) can be used to describe the relationship between objects, and objects with different attributes:

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

	A_1	A_2	$\cdots$	A_m
O_1	$a_{1,1}$	$a_{1,2}$	$\cdots$	$a_{1,m}$
O_2	$a_{2,1}$	$a_{2,2}$	$\cdots$	$a_{2,m}$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
O_n	$a_{n,1}$	$a_{n,2}$	$\cdots$	$a_{n,m}$

  Examples: (left) a similarity matrix, an image, a Google matrix; (right) a gene expression data, multivariate data, terms and documents.

- large data (n is large); high-dimensional data (m is large)

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	$a_{n,1,p}$	$a_{n,2,p}$	$\cdots$	$a_{n,n,p}$

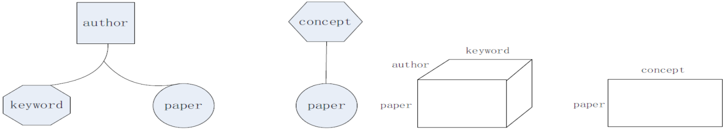
- p relationships among n objects

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 $\mathcal{A}$.

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-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

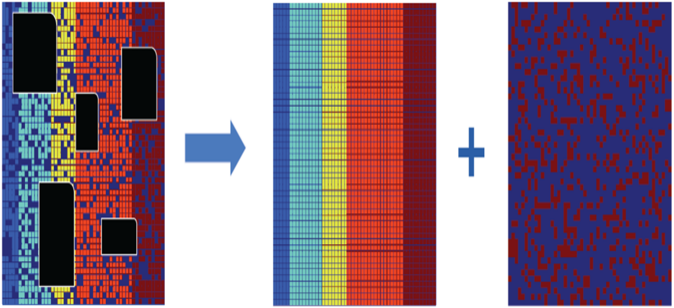
$$\min_{\mathbf{X}} \ rank(\mathbf{X}) \quad \text{subject to} \quad P_\Omega(\mathbf{X}) = P_\Omega(\mathbf{M})$$

Example

Matrix RPCA

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

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

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

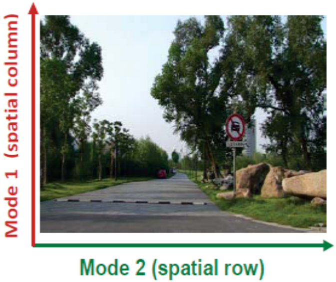
Nuclear norm $\|\cdot\|_*$: sum of singular values (convex envelop of rank)

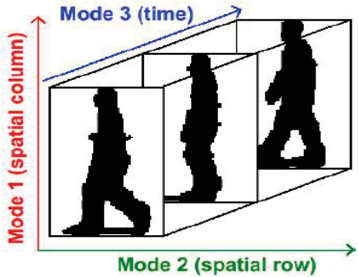
Low Rank Matrix Recovery Results

- (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.

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

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

- Tensor Robust PCA

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

- Robust Tensor Completion

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

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 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 $\mathcal{A}$.

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

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

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

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

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

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

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

The detailed theoretical and numerical results can be found in
https://arxiv.org/abs/1708.00601

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

The detailed theoretical and numerical results can be found in
http://www.math.hkbu.edu.hk/~mng/RTC.pdf

Summary

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