# Weighted tensor nuclear norm minimization for tensor completion using tensor-SVD

Yang Mu, Liangfu Lu

School of Mathematics, Tianjin Univerisity
liangfulv@tju.edu.cn

2018-04-15

# Contents

# Matrix completion

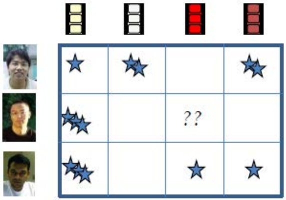



Figure: Netflix Challenge

Matrix completion problem can be expressed as:

$$\begin{aligned} & \textit{minimize } \textit{rank}(\mathbf{X}) \\ & \textit{subject to } P_\Omega(\mathbf{X}) = P_\Omega(\mathbf{M}), \end{aligned} \tag{1}$$

# Tensor completion

$$Evaluation(Feature) \to (Feature_1, Feature_2, ...Feature_n)$$

$$Matrix\ completion \to Tensor\ completion$$

Then tensor completion problem can be expressed as:

$$\begin{aligned} & minimize\ rank(\mathcal{X}) \\ & subject\ to\ P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{M}), \end{aligned} \tag{2}$$

# Tensor multiplication

$$bcirc(\mathcal{A}) = \begin{bmatrix} \mathbf{A}^{(1)} & \mathbf{A}^{(n_3)} & \dots & \mathbf{A}^{(2)} \\ \mathbf{A}^{(2)} & \mathbf{A}^{(1)} & \dots & \mathbf{A}^{(3)} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{A}^{(n_3)} & \mathbf{A}^{(n_3-1)} & \dots & \mathbf{A}^{(1)} \end{bmatrix},$$

$$unfold(\mathcal{A}) = \begin{bmatrix} \mathbf{A}^{(1)} \\ \mathbf{A}^{(2)} \\ \vdots \\ \mathbf{A}^{(n_3)} \end{bmatrix}, fold(unfold(\mathcal{A})) = \mathcal{A}.$$

# Tensor multiplication

$$(F_{n_3} \otimes I_{n_1}) \cdot bcirc(\mathcal{A}) \cdot (F_{n_3} \otimes I_{n_2}) = blkdiag(\hat{A}),$$

where

$$blkdiag(\hat{\mathcal{A}}) = \begin{bmatrix} \hat{A}^{(1)} & & & \\ & \hat{A}^{(2)} & & \\ & & \ddots & \\ & & & \hat{A}^{(n_3)} \end{bmatrix}$$

$$\mathcal{A} * \mathcal{B} = fold(bcirc(\mathcal{A}) \cdot (unfold(\mathcal{B}))).$$

# Tensor-SVD

## Theorem (Tensor-SVD)

*Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then it can be decomposed as*

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T, \tag{3}$$

*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 and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a $f$-diagonal tensor.*

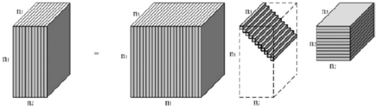


Figure: Tensor-SVD

## Innovation

- **Weighted tensor nuclear norm** is used to approximate the tensor completion problem using Tensor-SVD.
- The **convergence** of the algorithm is discussed under certain condition.
- The **KKT condition** is proofed under this situation.

# Weighted tensor nuclear norm

$$\|\mathcal{X}\|_\omega = \sum_{i=1}^{min(n_1,n_2)} \sum_{j=1}^{n_3} \omega_i \sigma_i$$

- when $\omega_i = 1$, $\|\mathcal{X}\|_w = \|\mathcal{X}\|_*$,
- when $\omega_i \leq \omega_j$ $(i \geq j)$, $\|\mathcal{X}\|_\omega$ is nonsmooth convex function,
- when $\omega_i \geq \omega_j$ $(i \geq j)$, $\|\mathcal{X}\|_\omega$ is nonsmoth nonconvex function(better performance for tensor completion).

# Convergence condition

$$\mathcal{L}(\hat{\mathcal{X}},\hat{\mathcal{Z}},\hat{\mathcal{Q}}) = \arg\min_{\hat{\mathcal{X}},\hat{\mathcal{Z}},\hat{\mathcal{Q}}} \|(\hat{\mathcal{Z}})\|_{\omega} + \delta_{\hat{\mathcal{Y}}=\mathcal{G}(\hat{X})} + \langle \hat{\mathcal{Q}}(:), \hat{\mathcal{X}}(:) - \hat{\mathcal{Z}}(:)\rangle + \frac{\beta}{2}\|\hat{\mathcal{X}} - \hat{\mathcal{Z}}\|_F^2, \quad (4)$$

The convergence condition:

$$\mathcal{L}(\hat{\mathcal{X}}_{k+1},\hat{\mathcal{Z}}_{k+1},\hat{\mathcal{Q}}_{k+1}) - \mathcal{L}(\hat{\mathcal{X}}_{k},\hat{\mathcal{Z}}_{k},\hat{\mathcal{Q}}_{k}) \le 0$$

- $\|(\hat{\mathcal{Z}})\|_{\omega} + \delta_{\hat{\mathcal{Y}}=\mathcal{G}(\hat{X})}$

**Assumption**

*Suppose that exist $\beta^{'}$, $\beta^{''}$ such that subproblem are strongly convex, and the strongly convex parameters are $\gamma_{'}$ and $\gamma_{''}$.*

- $\langle \hat{\mathcal{Q}}(:), \hat{\mathcal{X}}(:) - \hat{\mathcal{Z}}(:) \rangle + \frac{\beta}{2}\|\hat{\mathcal{X}} - \hat{\mathcal{Z}}\|_F^2$

**Assumption**

*Let $\gamma = min(\gamma_{'}, \gamma_{''})$, the following inequation exist*
$\|P_{\Omega}(\mathcal{M} - \mathcal{Z}_{k+1})\|_F \leq \sqrt{\frac{\gamma}{\beta_k}}(\|\mathcal{Z}_{k+1} - \mathcal{Z}_k\|_F + \|\mathcal{X}_{k+1} - \mathcal{X}_k\|_F)$.

# Assumption

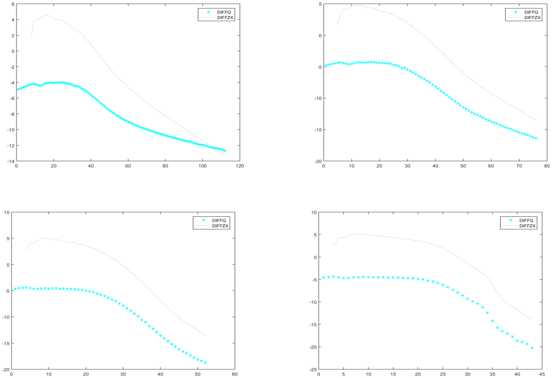



Figure: The difference between DIFFZX and DIFFQ with varying missing ratio and tubal rank.The tensor size is $100 \times 100 \times 50$ The tubal rank is 5. The missing ratio from up to down respectively are 45% 60%, 75% and 90%.

# Subgradient of weighted tensor nuclear norm

### Theorem

*The function of weighted sum of singular values* $h(\mathcal{X}) = \|\mathcal{X}\|_w = \sum_{i=1}^{min(n_1,n_2)} \sum_{j=1}^{n_3} w_i\sigma_i$ *(where* $w_i$ *denotes the weight of the singular value) for* $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *is Lipschitz continuous. Namely, there exists a constant scalar* $K$ *satisfying*

$$|h(\mathcal{X}_1) - h(\mathcal{X}_2)| \leq K\|\mathcal{X}_1 - \mathcal{X}_2\|_F$$

*for all* $\mathcal{X}_1, \mathcal{X}_2 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$.

**Lemma**

*Let $f : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}$ be a locally Lipschitz continuous function at $\mathcal{X}$ Then the differential $\partial_C f(\mathcal{X})$ is non-empty, convex and compact set.*

The subgradient of $h(\mathcal{X})$ exits.

**Lemma**

*Let $f : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}$ be a locally Lipschitz continuous function at $\mathcal{X}$, then a subgradient $\mathbf{W} \in \partial_C f(\mathcal{X})$ is bounded as*

$$\|\mathbf{W}\|_F \leq B.$$

Let $\mathcal{Q}'_{k+1} = \mathcal{Q}_k + \beta_k(\mathcal{X}_k - \mathcal{Z}_{k+1})$ , $\{\mathcal{Q}'_{k+1}\}_{k=1}^{\infty}$ is bounded.

### Theorem

*Suppose that* $S_k = \{\mathcal{X}_k, \mathcal{Z}_k, \mathcal{Q}_k, \mathcal{Q}_k^{'}, \}$. $\{\mathcal{Q}_{k+1}^{'}\}_{k=1}^{\infty}$ *is bounded.* $\lim\limits_{k\to\infty}(\mathcal{Q}_{k+1} - \mathcal{Q}_k) = 0$ *and* $\mu_k$ *is non-decreasing. Then the limit point* $(\mathcal{X}^*, \mathcal{Z}^*, \mathcal{Q}^*)$ *is a stationary point, which is* $0 \in \partial\mathcal{L}_\beta(\mathcal{X}^*, \mathcal{Z}^*, \mathcal{Q}^*)$*, or equivalently,*

$$\mathcal{Q}^* \in \partial\|\mathcal{Z}\|_\omega, \quad \mathcal{X}^* - \mathcal{Z}^* = 0, \quad \mathcal{Q}^* \in \partial 1_{\mathcal{X}_\Omega = \mathcal{M}_\Omega}$$

.

- SNN where $\alpha_i = \frac{1}{3}$
- TNN where $\omega_i = 1$;
- TTNN2 where $\omega_1 = 0$, $\omega_2 = 0$ and $\omega_i = 1; (i \neq 1, 2)$
- TTNN3 where $\omega_1 = 0$, $\omega_2 = 0$ ,$\omega_3$ and $\omega_i = 1; (i \neq 1, 2, 3)$
- TTNN5 where $\omega_j = 0; (i = 1, 2, 3, 4, 5)$ and $\omega_i = 1$ other wise;
- WWNN where

$$\omega(i, i, j) = \frac{1}{\sigma_1(i, i, j) + \epsilon} \quad \textit{where} \quad \epsilon > 0. \tag{5}$$

- RWTN where

$$\omega_k(i, i, j) = \frac{1}{\sigma_{k-1}(i, i, j) + \epsilon} \quad \textit{where} \quad \epsilon > 0. \tag{6}$$

## Image Recovery

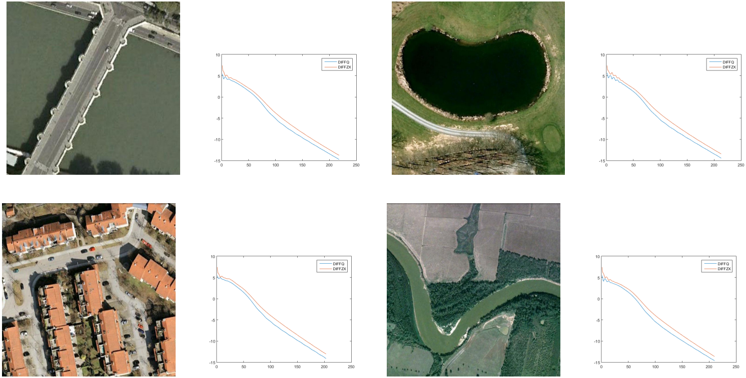



Figure: Oral image selected form WHU-RS19 dataset. The curve is the DIFFZX and DIFFQ, and their relation satisfied the Assumption 2

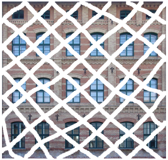
**Table:** The color image with 80% percentage missing ratio, the PSNR value and computation time of the each method list in the table.The sampling mode is using i.i.d.

| PSNR | Bridge | Pond | Residential | River |
|---|---|---|---|---|
| SNN | 27.42 | 26.35 | 21.59 | 26.70 |
| | 122 | 116 | 99 | 100 |
| TNN | 30.03 | 27.57 | 22.18 | 27.80 |
| | 128 | 123 | 122 | 121 |
| TTNN2 | 30.23 | 27.81 | 22.39 | 28.00 |
| | 132 | 126 | 126 | 124 |
| TTNN3 | 30.14 | 27.87 | 22.49 | 28.04 |
| | 136 | 129 | 130 | 128 |
| TTNN5 | 30.28 | 27.86 | 22.55 | 27.81 |
| | 143 | 132 | 135 | 132 |
| RWTN | **33.86** | **29.44** | **23.33** | **28.95** |
| | 159 | 161 | 155 | 152 |

Table: Fist row is the PSNR value of each method. Second row is the computation time

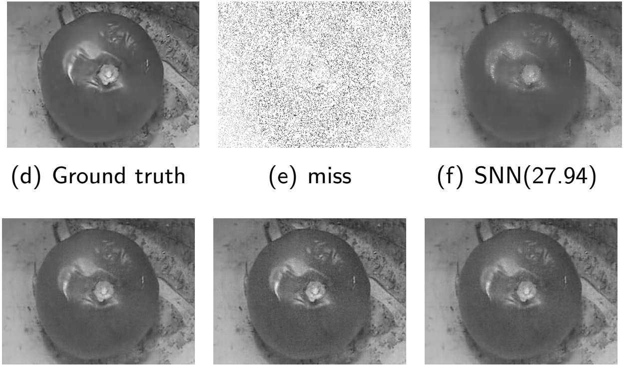
| Mask | SNN | TNN | TTNN2 | TTNN3 | TTNN5 | RWTN |
|---|---|---|---|---|---|---|
| Facede1 | 27.99 | 27.96 | 28.17 | 28.18 | **28.19** | 27.33 |
| | **70** | 104 | 100 | 101 | 104 | 121 |
| Facede2 | 31.82 | 31.81 | 32.01 | 32.03 | **32.21** | 31.59 |
| | **68** | 111 | 109 | 105 | 113 | 130 |

(a) Ground truth



(b) Facede1

(c) Facede2



(d) Ground truth (e) miss (f) SNN(27.94)

(g) TNN(31.75 ) (h) RWTN(31.17) (i) WWNN(31.11)

Figure: The recovering result on the tomato video with 80% missing ratio,The PSNR is corresponding for the whole data.

Table: The result of MRI video with different missing ratio is showed in this table. The fist column is the PSNR value, and the second column is the computation time.

| Missing ratio | 55% | 60% | 65% | 70% | 75% | 80% | 85% | 90% |
|---|---|---|---|---|---|---|---|---|
| SNN | 153 | 151 | 134 | 111 | 88 | 68 | 50 | 36 |
| | 516 | 493 | 507 | 525 | 590 | 585 | 644 | 585 |
| TNN | 122 | 114 | 106 | 91 | 81 | 73 | 69 | 59 |
| | 989 | 996 | 992 | 1016 | 993 | 996 | 1050 | 1053 |
| RWTN | 130 | 122 | 113 | 102 | 96 | 81 | 79 | 68 |
| | 906 | 1013 | 995 | 1029 | 1017 | 1024 | 1119 | 972 |
| WWNN | 124 | 117 | 107 | 93 | 82 | 74 | 69 | 58 |
| | 784 | 765 | 758 | 766 | 817 | 805 | 918 | 808 |

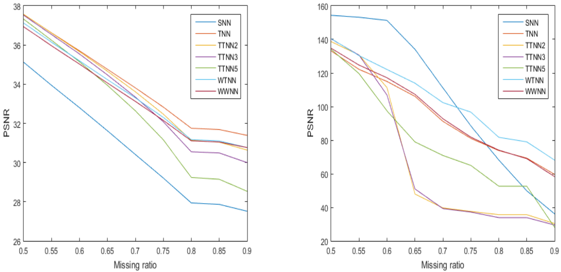



Figure: Tomato video and MRI data is design to recovery

*Thank you for your attention* ! ! !