A Dual Framework for Low-rank Tensor Completion

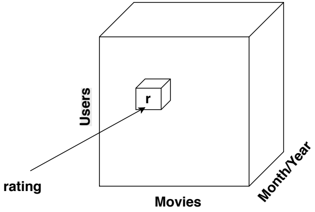
Madhav Nimishakavi
IISc, Bangalore

Outline

1. Preliminaries
2. A Dual Framework for Tensor Completion
   - Introduction
   - Proposed Variant of Trace Norm
   - Dual Formulation
   - Optimization
   - Experimental Results
   - Conclusion

Preliminaries

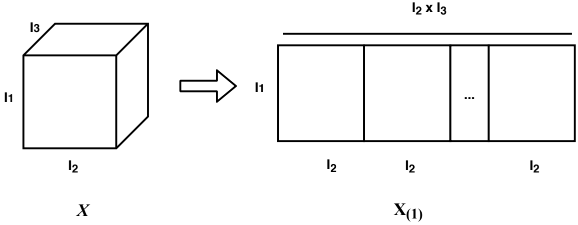
- A **Tensor** is a multi-way extension of a matrix.



- Tensors are used for representing multidimensional data.

Preliminaries (cont.)

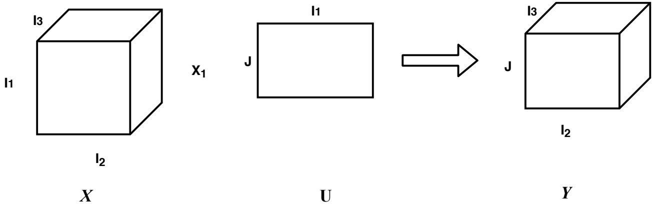
- **Tensor Matricization or Tensor Unfolding** is the process of reordering the elements of an N-order tensor into a matrix. The mode-n unfolding of $\mathcal{X}$,denoted by $\mathbf{X}_{(n)}$, arranges the mode-n fibers to be the columns of resulting matrix.



Preliminaries (cont.)

- **Tensor-Matrix Multiplication** The *n-mode (matrix) product* of a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_N}$ with a matrix $\mathbf{U} \in \mathbb{R}^{J \times I_n}$ is denoted by $\boldsymbol{\mathcal{X}} \times_n \mathbf{U}$ and is of size $I_1 \times \ldots I_{n-1} \times J \times I_{n+1} \times \ldots I_N$.

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} \times_n U \iff \mathbf{Y}_{(n)} = \mathbf{U}\mathbf{X}_{(n)}$$



Outline

1 Preliminaries

2 A Dual Framework for Tensor Completion
- Introduction
- Proposed Variant of Trace Norm
- Dual Formulation
- Optimization
- Experimental Results
- Conclusion

Introduction

- Tensors can be used to model multi-dimensional data.
- In practice, often the multidimensional datasets are incomplete.

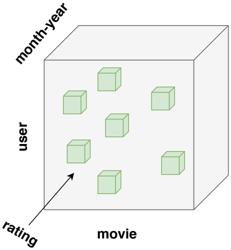


Figure : Partially observed tensor, green represents observed entries.

Introduction (cont.)

- **Tensor Completion**: The problem of imputing or predicting missing values in a partially observed tensor.
- Applications:
  - Image Completion
  - Recommendation Systems
  - Video Completion
  - Link prediction in Knowledge Graphs, Social Networks etc
  - . . .

Introduction (cont.)

- A common modeling assumption for tensor completion: the tensor is low-rank.
- Common approaches:
  - Factorization based
  - Trace-norm regularization based

Problem Formulation

Trace Norm based Fromulation:

$$\min_{\mathcal{W} \in \mathbb{R}^{n_1 \times n_2 \times \ldots n_K}} \|\mathcal{W}_\Omega - \mathcal{Y}_\Omega\|_F^2 + \frac{1}{\lambda} \|\mathcal{W}\|_*$$

where $\mathcal{Y}$ is a partially observed tensor, Ω is the sparsity pattern, $\|\cdot\|_*$ is the trace norm and $\lambda > 0$ is the regularization constant.

A rank-$(r_1, r2, \ldots, r_K)$ tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \ldots \times n_K}$ can be written as:

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \ldots \times_K \mathbf{U}_K,$$

where

- $\mathbf{U}_k \in \mathbb{R}^{n_k \times r_k}$ are left singular values from SVD of $\boldsymbol{\mathcal{X}}_{(k)}$.
- $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{r_1 \times r_2 \ldots r_K}$ is the core tensor.

A rank-$(r_1, r2, \ldots, r_K)$ tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \ldots \times n_K}$ can be written as:

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \ldots \times_K \mathbf{U}_K,$$

where

- $\mathbf{U}_k \in \mathbb{R}^{n_k \times r_k}$ are left singular values from SVD of $\boldsymbol{\mathcal{X}}_{(k)}$.
- $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{r_1 \times r_2 \ldots r_K}$ is the core tensor.

- The above decomposition is called Tucker decomposition.

- $\mathcal{G}$ is not diagonal in general.

- Not straightforward to generalize trace norm to tensors.

Trace Norm of Tensor

Strategies for tensor trace norm [Tomioka et al., 2010].

1. Tensor as a Matrix:

$$\|\mathcal{W}\|_* := \left\|\mathbf{W}_{(k)}\right\|_*, \textit{for} \quad \textit{some} \quad 1 \leq k \leq K$$

Trace Norm of Tensor

Strategies for tensor trace norm [Tomioka et al., 2010].

1. Tensor as a Matrix:

$$\|\mathcal{W}\|_* := \left\|\mathbf{W}_{(k)}\right\|_*, \textit{for} \quad \textit{some} \quad 1 \leq k \leq K$$

2. Overlapped Trace Norm:

$$\|\mathcal{W}\|_* := \sum_{k=1}^{K} \left\|\mathbf{W}_{(k)}\right\|_*$$

Trace Norm of Tensor

Strategies for tensor trace norm [Tomioka et al., 2010].

1. Tensor as a Matrix:

$$\|\mathcal{W}\|_* := \left\|\mathbf{W}_{(k)}\right\|_*, \textit{for} \quad \textit{some} \quad 1 \leq k \leq K$$

2. Overlapped Trace Norm:

$$\|\mathcal{W}\|_* := \sum_{k=1}^{K} \left\|\mathbf{W}_{(k)}\right\|_*$$

3. Latent Trace Norm:

$$\|\mathcal{W}\|_* := \inf_{\sum_{k=1}^{K} \mathcal{W}^{(k)} = \mathcal{W};\ \mathcal{W}^{(k)} \in \mathbb{R}^{n_1 \times \dots \times n_K}} \sum_{k=1}^{K} \|\mathcal{W}_{(k)}^{(k)}\|_*$$

Latent Trace Norm

- ✓ Latent Trace Norm is showed to produce better results [Tomioka and Suzuki, 2013].
- × But, Latent Trace Norm learns a skewed-mixture in sparse settings.

Outline

1. Preliminaries

2. A Dual Framework for Tensor Completion
   - Introduction
   - Proposed Variant of Trace Norm
   - Dual Formulation
   - Optimization
   - Experimental Results
   - Conclusion

Non-Sparse Trace Norm

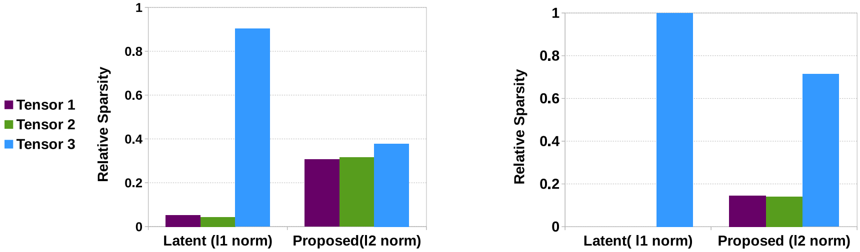
We propose the following variant of trace norm:

$$\min_{\mathcal{W}^{(k)} \in \mathbb{R}^{n_1 \times \ldots \times n_K}} \left\| \sum_k \mathcal{W}_\Omega^{(k)} - \mathcal{Y}_\Omega \right\|_F^2 + \sum_k \frac{1}{\lambda_k} \| \mathcal{W}_{(k)}^{(k)} \|_*^2, \tag{1}$$

where $\mathcal{W} = \sum_k \mathcal{W}^{(k)}$ is the learned tensor.

Non-Sparse Trace Norm (cont.)

$$\text{Relative sparsity} = \frac{\left\|\mathcal{W}^{(k)}\right\|_F}{\sum_k \left\|\mathcal{W}^{(k)}\right\|_F}$$



(a) Ribeira (b) Baboon

Figure : (a) & (b) Relative sparsity of each tensor in the mixture of tensors for Ribeira and Baboon datasets. Our proposed formulation learns a ℓ_2-norm based non-sparse combination of tensors

Outline

1 Preliminaries

2 A Dual Framework for Tensor Completion
- Introduction
- Proposed Variant of Trace Norm
- Dual Formulation
- Optimization
- Experimental Results
- Conclusion

Dual Formulation

Theorem

An equivalent minimax formulation of the problem (1) is

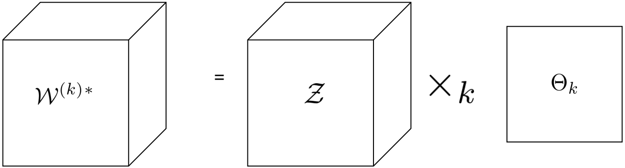
$$\min_{\Theta_1 \in \mathcal{P}^{n_1}, \ldots, \Theta_K \in \mathcal{P}^{n_K}} \; \max_{\boldsymbol{\mathcal{Z}} \in \mathcal{C}} \; \langle \boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{Y}}_{\Omega} \rangle - \frac{1}{4} \|\boldsymbol{\mathcal{Z}}\|_F^2 - \sum_k \frac{\lambda_k}{2} \langle \Theta_k, \boldsymbol{\mathcal{Z}}_{(k)} \boldsymbol{\mathcal{Z}}_{(k)}^\top \rangle, \quad (2)$$

where $\boldsymbol{\mathcal{Z}}$ is the dual tensor variable corresponding to the primal problem (1) and $\boldsymbol{\mathcal{Z}}_{(k)}$ is the mode-k unfolding of $\boldsymbol{\mathcal{Z}}$. The set $\mathcal{C} := \{\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_1 \times \ldots \times n_K} : \boldsymbol{\mathcal{Z}}_{(i_1, \ldots, i_K)} = 0, (i_1, \ldots, i_K) \notin \Omega\}$ constrains $\boldsymbol{\mathcal{Z}}$ to be a sparse tensor with $|\Omega|$ non-zero entries. Let $\{\Theta_1^*, \ldots, \Theta_K^*, \boldsymbol{\mathcal{Z}}^*\}$ be the optimal solution of (2). The optimal solution of (1) is given by $\boldsymbol{\mathcal{W}}^* = \sum_k \boldsymbol{\mathcal{W}}^{(k)*}$, where $\boldsymbol{\mathcal{W}}^{(k)*} = \lambda_k (\boldsymbol{\mathcal{Z}}^* \times_k \Theta_k^*)$ $\forall k$ and $\times_k$ denotes the tensor-matrix multiplication along mode k.

Dual Formulation (cont.)

Corollary

(Representer theorem) The optimal solution of the primal problem (1) admits a representation of the form: $\boldsymbol{\mathcal{W}}^{(k)*} = \lambda_k(\boldsymbol{\mathcal{Z}} \times_k \Theta_k)\ \forall k$, *where* $\boldsymbol{\mathcal{Z}} \in \mathcal{C}$ *and* Θ_k *are PSD with* $trace(\Theta_k) = 1$.



Dual Formulation (cont.)

- **Fixed-rank parameterization of Θ_k.** We propose to explicitly constrain the rank of Θ_k to r_k by $\Theta_k = \mathbf{U}_k \mathbf{U}_k^\top$, where $\mathbf{U}_k \in \mathcal{S}_{r_k}^{n_k}$ and $\mathcal{S}_r^n := \{\mathbf{U} \in \mathbb{R}^{n \times r} : \|\mathbf{U}\|_F = 1\}$.

- **Fixed-rank dual formulation**

$$\text{Problem} \quad \mathcal{D}: \quad \min_{u \in \mathcal{S}_{r_1}^{n_1} \times \ldots \times \mathcal{S}_{r_K}^{n_K}} \quad g(u),$$

where $u = (\mathbf{U}_1, \ldots, \mathbf{U}_K)$ and $g : \mathcal{S}_{r_1}^{n_1} \times \ldots \times \mathcal{S}_{r_K}^{n_K} \to \mathbb{R}$ is the function

$$g(u) \coloneqq \max_{\boldsymbol{\mathcal{Z}} \in \mathcal{C}} \langle \boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{Y}}_\Omega \rangle - \frac{1}{4} \|\boldsymbol{\mathcal{Z}}\|_F^2 - \sum_k \frac{\lambda_k}{2} \left\| \mathbf{U}_k^\top \boldsymbol{\mathcal{Z}}_{(k)} \right\|_F^2 .$$

Outline

1. Preliminaries

2. A Dual Framework for Tensor Completion
   - Introduction
   - Proposed Variant of Trace Norm
   - Dual Formulation
   - Optimization
   - Experimental Results
   - Conclusion

Riemannian Optimization Algorithm

- $\mathcal{S}_r^n$ is the *spectrahedron* manifold [Journée et al., 2010].
- Optimization on spectrahedron manifold is handled in the Riemannian optimization framework.
- We propose Riemannian TR algorithms (**TR-MM**) for $\mathcal{D}$.

Outline

1. Preliminaries

2. A Dual Framework for Tensor Completion
   - Introduction
   - Proposed Variant of Trace Norm
   - Dual Formulation
   - Optimization
   - Experimental Results
   - Conclusion

Datasets

Dataset	Dimensions	Task
Ribeira	$203 \times 268 \times 33$	Hyperspectral Image Completion
Tomato	$242 \times 320 \times 167$	Video Completion
Baboon	$256 \times 256 \times 3$	Image Completion
MovieLens10M	$71567 \times 10681 \times 731$	Recommendation
FB15k-237	$14541 \times 14541 \times 237$	Link Prediction
YouTube (subset)	$1509 \times 1509 \times 5$	Link Prediction
YouTube (full)	$15088 \times 15088 \times 5$	Link Prediction

Table : Datasets for Tensor Completion

Results

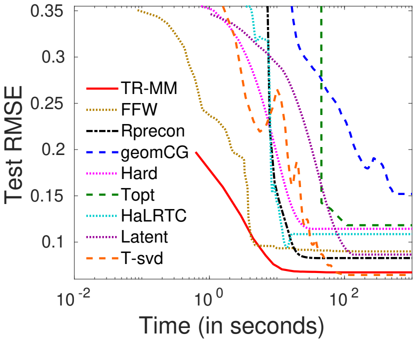
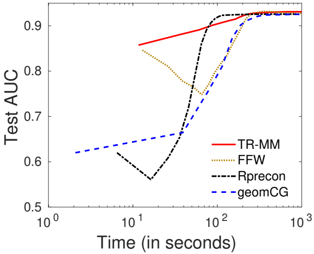
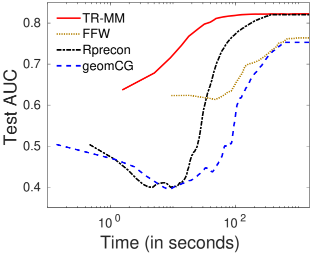


Figure : Evolution of test RMSE on Ribeira (Image Completion Dataset)

Results (cont.)



(a) YouTube (full)



(b) FB15K-237

Figure : Evolution of test AUC on link prediction datasets YouTube and FB15K-237.

Results (cont.)

	TR-MM	FFW	Rprecon	geomCG	Hard	Topt	HaLRTC	Latent	T-svd	BayesCP
RMSE reported										
Ribeira	0.067	0.088	0.083	0.156	0.114	0.127	0.095	0.087	**0.064**	0.154
Tomato	**0.041**	0.045	0.052	0.052	0.060	0.102	0.202	0.046	0.042	0.103
Baboon	**0.121**	0.133	0.128	0.128	0.126	0.130	0.247	0.459	0.146	0.159
ML10M	0.840	0.895	**0.831**	0.844	–	–	–	–	–	–
AUC reported										
YouTube (subset)	**0.957**	0.954	0.941	0.941	0.954	0.941	0.783	0.945	0.941	0.950
YouTube (full)	**0.932**	0.929	0.926	0.926	–	–	–	–	–	–
FB15k-237	**0.823**	0.764	0.821	0.785	–	–	–	–	–	–

Table : Generalization performance across several applications: hyperspectral-image/video/image completion, movie recommendation, and link prediction.

Outline

1 Preliminaries

2 A Dual Framework for Tensor Completion
- Introduction
- Proposed Variant of Trace Norm
- Dual Formulation
- Optimization
- Experimental Results
- Conclusion

Conclusion

- We propose a variant of Latent trace norm that learns a non-skewed mixture of tensors.
- We propose a dual framework for analyzing the formulation.
- The dual framework achieves better generalization performance.

Conclusion (cont.)

- Madhav Nimishakavi, Pratik Jawanpuria and Bamdev Mishra. A Dual Framework for Low-rank Tensor Completion. In *In Advances of Neural Information Processing Systems (NIPS), 2018.*

Bibliography

[Journée et al., 2010] Journée, M., Bach, F., Absil, P.-A., and Sepulchre, R. (2010).
Low-rank optimization on the cone of positive semidefinite matrices.
SIAM J. on Optimization, 20(5):2327–2351.

[Tomioka et al., 2010] Tomioka, R., Hayashi, K., and Kashima, H. (2010).
Estimation of low-rank tensors via convex optimization.
Technical report, arXiv preprint arXiv:1010.0789.

[Tomioka and Suzuki, 2013] Tomioka, R. and Suzuki, T. (2013).
Convex tensor decomposition via structured schatten norm regularization.
In *NIPS*.