

Structured Low-Rank Tensor Learning

Jayadev Naram
Tanmay Kumar Sinha
Pawan Kumar
CSTAR, IIT Hyderabad

JAYADEV.NARAM@RESEARCH.IIT.AC.IN
 TANMAY.KUMAR@RESEARCH.IIT.AC.IN
 PAWAN.KUMAR@IIT.AC.IN

Abstract

We consider the problem of learning low-rank tensors from partial observations with structural constraints, and propose a novel factorization of such tensors, which leads to a simpler optimization problem. The resulting problem is an optimization problem on manifolds. We develop first-order and second-order Riemannian optimization algorithms to solve it. The duality gap for the resulting problem is derived, and we experimentally verify the correctness of the proposed algorithm. We demonstrate the algorithm on nonnegative constraints and Hankel constraints.

1. Introduction

With the rise in the availability of multidimensional data such as colour images, video sequences, and hyperspectral images, tensor-based techniques have started gaining attention as traditional matrix-based methods cannot exploit the underlying structure present in higher-dimensional data. Many recent applications of tensor reconstruction techniques also enforce structural constraints such as nonnegativity ([10], [20]) or a Hankel structure ([17], [18]).

We propose a framework for dealing with tensor completion problems with general structural constraints. In particular, we consider the structured low-rank tensor learning problem of the following form:

$$\begin{aligned} \min_{\mathcal{W} \in \mathbb{R}^{n_1 \times \dots \times n_K}} \quad & C \|\mathcal{W}_\Omega - \mathcal{Y}_\Omega\|^2 + R(\mathcal{W}) \\ \text{subject to} \quad & A(\mathcal{W}) \geq 0, \end{aligned} \quad (1)$$

where $\mathcal{Y}_\Omega \in \mathbb{R}^{n_1 \times \dots \times n_K}$ is a partially observed tensor for indices given in the set Ω , $(\mathcal{W}_\Omega)_{i_1, \dots, i_K} = w_{i_1, \dots, i_K}$ for $(i_1, \dots, i_K) \in \Omega$, $C > 0$ denotes the cost parameter, $R : \mathbb{R}^{n_1 \times \dots \times n_K} \rightarrow \mathbb{R}$ is a regularizer that induces low-rank constraint on the tensor, and $A : \mathbb{R}^{n_1 \times \dots \times n_K} \rightarrow \mathbb{R}^n$ is a linear map that induces structural constraint on the tensor.

Following [5], we learn the tensor as a sum of K tensors, $\mathcal{W} = \mathcal{W}^{(1)} + \dots + \mathcal{W}^{(K)}$, and use the following low-rank regularizer:

$$R(\mathcal{W}) = \sum_{k=1}^K \frac{1}{\lambda_k} \left\| \mathcal{W}_k^{(k)} \right\|_*^2. \quad (2)$$

The main contributions of the paper are given below.

- We propose a novel factorization for modeling structured low-rank tensors through a partial dual problem of (1).

- We develop first-order and second-order Riemannian optimization algorithms that exploit proposed factorization’s inherent geometric structure.
- We compute the expression for the duality gap and verify the correctness of the proposed algorithm through experiments.
- We apply the proposed algorithm to the nonnegative constraint and the Hankel constraint.

2. Notation

We follow [2] for our tensor notation. We present a few important notions here. Tensors are denoted by uppercase calligraphic letters, e.g., \mathcal{W} . Matrices are denoted by uppercase letters, e.g., X . For a square matrix $X \in \mathbb{R}^{n \times n}$, we denote its trace by $tr(X)$. \mathbb{R}_+ denotes the interval $[0, \infty)$. The inner product of two tensors is defined as follows:

$$\langle \mathcal{W}, \mathcal{U} \rangle = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_K=1}^{n_K} w_{i_1, \dots, i_K} u_{i_1, \dots, i_K}.$$

A mode- k fiber of a tensor $\mathcal{W} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$, denoted by $w_{i_1, \dots, i_{k-1}, :, i_{k+1}, \dots, i_K}$, is a vector obtained by fixing all but k -th index of \mathcal{W} . The mode- k unfolding of a tensor $\mathcal{W} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ is a matrix $unfold_k(\mathcal{W}) = W_k \in \mathbb{R}^{n_k \times n_1 \cdots n_{k-1} n_{k+1} \cdots n_K}$ formed by arranging the mode- k fibers to be the columns of the resulting matrix. Similarly, we can define the k -mode folding operation ($fold_k$) as the inverse of the unfolding operation - it converts a given matrix to a tensor of a suitable order. The k -mode product of a tensor $\mathcal{W} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ with a matrix $U \in \mathbb{R}^{m \times n_k}$ is defined as follows:

$$(\mathcal{W} \times_k U)_{i_1, \dots, i_{k-1}, j, i_{k+1}, \dots, i_K} = \sum_{i_k}^{n_k} w_{i_1, \dots, i_K} u_{j, i_k}, \text{ i.e., } \mathcal{X} = \mathcal{W} \times_k U \iff X_k = U W_k.$$

3. Related Work

There are a number of well-known methods for tensor completion using different approaches for enforcing the low-rank constraint. Many methods use a generalization of the matrix trace norm regularizer to the tensor case. Two well-known regularizers are the overlapped trace norm regularizer ([7], [12] and [11]) and latent trace norm regularizer ([15]). The overlapped trace norm regularizer uses the regularizer $R(\mathcal{W}) = \sum_{k=1}^K \|W_k\|_*$. In the latent trace norm regularizer model, the tensor is modelled as the sum of K -tensors, $\mathcal{W} = \sum_{k=1}^K \mathcal{W}^{(k)} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$, and the regularizer is defined as

$$R(\mathcal{W}) = \inf_{\mathcal{W}^{(k)}, k \in \{1, \dots, K\}} \sum_{k=1}^K \|W_k^{(k)}\|_*.$$

Other methods for tensor completion use tensor decompositions like the Tucker decomposition and CP decomposition to learn the tensors. [8] and [9] formulate the tensor completion problem as an optimization problem on the Riemannian manifold of fixed multi-linear rank tensors.

A general formulation for low-rank matrix completion problems with structural constraints was developed in [13]. It provided a unified framework for dealing with general linear inequality and equality constraints.

Recently, tensor completion problems with structural constraints have started attracting attention. [10] and [20] develop nonnegative tensor completion algorithms for image and video reconstruction tasks. [14] falls under the current framework which considers the special case of nonnegative constraints in detail with vast experimentation. [17] models image completion with missing slices as a higher-order Hankel tensor completion problem. [19] uses low-rank Hankel tensor model for estimating traffic states from partial observations.

4. Dual Framework

Following [5], we construct a partial dual problem to the primal problem (1), incorporating the structural constraint into the formulation. We do this using the approach outlined in [13]. We generalize Theorem 1 of [5] to the case of tensors with structural constraints. The following lemma [6] is used in the development of the dual formulation.

Lemma 1 *For a matrix $X \in \mathbb{R}^{d \times T}$, the nuclear norm of X satisfies the following relation:*

$$\|X\|_*^2 = \min_{\Theta \in P^d, \text{range}(X) \subseteq \text{range}(\Theta)} \langle \Theta^\dagger X, X \rangle,$$

where $P^d = \{S \in \mathbb{R}^{d \times d} : S \succeq 0, \text{tr}(S) = 1\}$, $\text{range}(\Theta) = \{\Theta z : z \in \mathbb{R}^d\}$, Θ^\dagger denotes the pseudo-inverse of Θ . For a given X , the minimizer is $\bar{\Theta} = \sqrt{X X^T} / \text{tr}(\sqrt{X X^T})$.

Using the above lemma, we can write (1) as

$$\begin{aligned} \min_{\substack{\Theta_k \in P^{n_k}, \\ k \in \{1, \dots, K\}}} \min_{\substack{\mathcal{W}^{(k)}, \\ k \in \{1, \dots, K\}}} C \left\| \left(\sum_{k=1}^K \mathcal{W}^{(k)} \right)_\Omega - \mathcal{Y}_\Omega \right\|^2 + \sum_{k=1}^K \frac{1}{2\lambda_k} \langle \Theta_k^\dagger W_k^{(k)}, W_k^{(k)} \rangle \\ \text{subject to} \quad A(\mathcal{W}) \geq 0. \end{aligned} \quad (3)$$

Theorem 2 *The following minimax problem is equivalent to the problem (3):*

$$\min_{\substack{\Theta_k \in P^{n_k}, \\ k \in \{1, \dots, K\}}} \max_{\mathcal{Z} \in \mathcal{C}, s \in \mathbb{R}^n} \langle \mathcal{Z}, \mathcal{Y}_\Omega \rangle - \frac{1}{4C} \|\mathcal{Z}\|^2 - \sum_{k=1}^K \frac{\lambda_k}{2} \langle Z_k + (A^*(s))_k, \Theta_k [Z_k + (A^*(s))_k] \rangle, \quad (4)$$

where $\mathcal{C} = \{\mathcal{Z} \in \mathbb{R}^{n_1 \times \dots \times n_K} : \mathcal{Z} = \bar{\mathcal{Z}}_\Omega\}$. The optimal solution $\bar{\mathcal{W}}$ of (3) is related to the optimal solution $\{\bar{\Theta}_1, \dots, \bar{\Theta}_K, \bar{\mathcal{Z}}, \bar{s}\}$ of (4) by

$$\bar{\mathcal{W}} = \sum_{k=1}^K \lambda_k (\bar{\mathcal{Z}} + A^*(\bar{s})) \times_k \bar{\Theta}_k. \quad (5)$$

For proof, see Appendix B.

From (5) it can be seen that the structured and low-rank constraints on \mathcal{W} can be decomposed into structured constraints on s and low-rank constraints on Θ_k , which leads to a simpler optimization method.

5. Proposed Algorithm

Due to the low-rank constraint on \mathcal{W} , each Θ_k has a low rank. Therefore, a fixed-rank parameterized problem can be given by writing $\Theta_k = U_k U_k^T$. The problem (4) can then be written as follows.

$$\min_{U \in S_{r_1}^{n_1} \times \dots \times S_{r_K}^{n_K}} g(U), \quad (6)$$

where $U = (U_1, \dots, U_K)$, $S_r^n = \{U \in \mathbb{R}^{n \times r} : \|U\|_F = 1\}$, and

$$g(U) = \max_{\mathcal{Z} \in \mathcal{C}, s \in \mathbb{R}^n} \langle \mathcal{Z}, \mathcal{Y}_\Omega \rangle - \frac{1}{4C} \|\mathcal{Z}\|^2 - \sum_{k=1}^K \frac{\lambda_k}{2} \|U_k^T (Z_k + (A^*(s))_k)\|^2. \quad (7)$$

The optimization problem in (7) is strongly convex for a given U , while problem (6) is a non-convex problem in U .

The set $S_{r_1}^{n_1} \times \dots \times S_{r_K}^{n_K}$ is a Riemannian manifold ([4], [5]), and thus problem (6) is an optimization problem on a manifold. We solve it using either Riemannian conjugate gradient or Riemannian Trust-Region algorithm, depending on the structural constraint. The proposed algorithm is shown in Algo. 1. For more details on optimization on general manifolds, we refer the reader to [1] and [3].

The Riemannian optimization algorithm in Algo. 1 requires computing the Euclidean gradient and its directional derivative, which are given in the following lemma.

Lemma 3 *Let $\{\hat{\mathcal{Z}}, \hat{s}\}$ be the maximizer of the convex problem (7) at U . Then, the Euclidean gradient $\nabla g(U)$ is given by*

$$\nabla g(U) = -(\lambda_1 P_1, \dots, \lambda_K P_K),$$

where $P_k = (\hat{Z}_k + (A^*(\hat{s}))_k)(\hat{Z}_k + (A^*(\hat{s}))_k)^T U_k$. Let $V \in \mathbb{R}^{n_1 \times r_1} \times \dots \times \mathbb{R}^{n_K \times r_K}$ and \dot{Z}_k, \dot{s} denote the directional derivatives of Z_k and s along V respectively. Then, the directional derivative of ∇g at U along V is

$$D\nabla g(U)[V] = -(\lambda_1 Q_1, \dots, \lambda_K Q_K),$$

where $Q_k = (\hat{Z}_k + (A^*(\hat{s}))_k)(\dot{Z}_k + (A^*(\dot{s}))_k)(\hat{Z}_k + (A^*(\hat{s}))_k)^T U_k + 2 \text{sym}((\dot{Z}_k + (A^*(\dot{s}))_k)(\hat{Z}_k + (A^*(\hat{s}))_k)^T) U_k$ and $\text{sym}(X) = (X + X^T)/2$.

Remark 4 *It can be seen that computing $D\nabla g(U)[V]$ requires the terms \dot{Z} and \dot{s} . These terms can be computed by applying directional derivative along V on the first-order optimality conditions of the problem (7) at $\{\hat{\mathcal{Z}}, \hat{s}\}$.*

Algorithm 1: Proposed Algorithm for Structured Low-Rank Tensor Completion

Data: \mathcal{Y}_Ω , $\text{rank}=(r_1, \dots, r_K)$, ε , $(\lambda_1, \dots, \lambda_K)$

Result: $\hat{\mathcal{W}} = \sum_{k=1}^K \lambda_k (\hat{\mathcal{Z}} + A^*(\hat{s})) \times_k (U_k U_k^T)$

for $t = 1, 2, \dots$ **do**

Check Termination: if $\|\nabla g(U^{(t)})\| \leq \varepsilon$ then break;

Solve for $\hat{\mathcal{Z}}^{(t)}$ and $\hat{s}^{(t)}$ in (7);

Compute cost $g(U^{(t)})$, gradient $\nabla g(U^{(t)})$ and the directional derivative of $\nabla g(U^{(t)})$;

Update U : $U^{(t+1)} = \text{RiemannianCG-update}(U^{(t)})$ or
 $U^{(t+1)} = \text{RiemannianTR-update}(U^{(t)})$;

end

We have the following result regarding the optimality of the proposed algorithm. It is a generalization of Theorem 3 in [13].

Theorem 5 Let $\hat{U} = (U_1, \dots, U_K)$ be a feasible solution of problem (6) and $\{\hat{\mathcal{Z}}, \hat{s}\}$ be the maximizer of the convex problem (7) at $U = \hat{U}$. Let $\hat{A} = A^*(\hat{s})$, and σ_k be the maximum singular value of $\hat{Z}_k + \hat{A}_k$. Let $\hat{\Theta} = (\hat{\Theta}_1, \dots, \hat{\Theta}_K)$, where $\hat{\Theta}_k = \hat{U}_k \hat{U}_k^T$. Then, $\{\hat{\Theta}, \hat{\mathcal{Z}}, \hat{s}\}$ is a candidate solution for the partial dual problem (4) and we have the following expression for the duality gap Δ :

$$\Delta = \sum_{k=1}^K \frac{\lambda_k}{2} \left(\sigma_k^2 - \|\hat{U}_k^T (\hat{Z}_k + \hat{A}_k)\|^2 \right). \quad (8)$$

For proof see Appendix C.

6. Applications

We consider several popular applications for our proposed method. See [5] for the case where there are no structure constraints.

6.1. Nonnegative Tensor completion

The nonnegative tensor completion problem is

$$\begin{aligned} \min_{\mathcal{W} \in \mathbb{R}^{n_1 \times \dots \times n_K}} \quad & C \|\mathcal{W}_\Omega - \mathcal{Y}_\Omega\|^2 + \sum_{k=1}^K \frac{1}{\lambda_k} \left\| W_k^{(k)} \right\|_*^2 \\ \text{subject to} \quad & \mathcal{W} \geq 0. \end{aligned} \quad (9)$$

The fixed-rank dual problem of (9) is

$$\min_{U \in S_{r_1}^{n_1} \times \dots \times S_{r_K}^{n_K}} g(U), \quad (10)$$

$g(U)$ is given by

$$g(U) = \max_{\mathcal{S} \in \mathbb{R}_+^{n_1 \times \dots \times n_K}} \left(\max_{\mathcal{Z} \in \mathcal{C}} \langle \mathcal{Z}, \mathcal{Y}_\Omega \rangle - \frac{1}{4C} \|\mathcal{Z}\|^2 - \sum_{k=1}^K \frac{\lambda_k}{2} \|U_k^T \mathcal{Z}_k + U_k^T \mathcal{S}_k\|^2 \right). \quad (11)$$

where $U = (U_1, \dots, U_K)$ and $\mathbb{R}_+^{n_1 \times \dots \times n_K}$ is the set of all tensors of size $n_1 \times \dots \times n_K$ with non-negative entries.

The problem (11) can be solved alternatively for \mathcal{Z} and \mathcal{S} . Then the resulting problem in \mathcal{Z} is unconstrained least-squares problem, which can be solved using linear conjugate gradient algorithm, and over \mathcal{S} it is nonnegative least-squares problem which can be solved using [16]. The cost $g(U)$ and the gradient $\nabla g(U)$ can be easily computed following Lemma 3, hence, we employ Riemannian conjugate gradient to solve the outer optimization problem over U . The per-iteration complexity of the proposed algorithm is $O\left(T|\Omega| \sum_{k=1}^K r_k + \sum_{k=1}^K n_k r_k^2 + \sum_{k=1}^K r_k^3\right)$, where T is the sum of iterations of CG and NNLS solver for solving \mathcal{Z} and \mathcal{S} respectively.

6.2. Hankel Tensor Completion

The primal problem is given by

$$\min_{\mathcal{W} \in \mathbb{R}^{n_1 \times \dots \times n_K}} C \|\mathcal{W}_\Omega - \mathcal{Y}_\Omega\|^2 + \sum_{k=1}^K \frac{1}{\lambda_k} \left\| \mathcal{H}_{2k}(\mathcal{W}^{(k)}) \right\|_*^2, \quad (12)$$

where $\mathcal{H}_k = \text{unfold}_k \circ \mathcal{H}$ and \mathcal{H} is the k -th order Hankel transform such that for $\mathcal{W} \in \mathbb{R}^{n_1 \times \dots \times n_K}$, $\mathcal{H}(\mathcal{W}) \in \mathbb{R}^{\tau_1 \times n_1 - \tau_1 + 1 \times \dots \times \tau_K \times n_K - \tau_K + 1}$ is a $2K$ order tensor for some $\tau = (\tau_1, \dots, \tau_K)$, which is the duplication parameter of the Hankel transform (see [17]). The fixed-rank dual problem of (12) is

$$\min_{U \in S_{r_1}^{n_1 - \tau_1 + 1} \times \dots \times S_{r_K}^{n_K - \tau_K + 1}} g(U), \quad (13)$$

where $U = (U_1, \dots, U_K)$, and

$$g(U) = \max_{\mathcal{Z} \in \mathcal{C}, \mathcal{S}} \langle \mathcal{Z}, \mathcal{Y}_\Omega \rangle - \frac{1}{4C} \|\mathcal{Z}\|^2 - \sum_{k=1}^K \frac{\lambda_k}{2} \|U_k^T \mathcal{S}_k\|^2, \quad \text{subject to } \mathcal{H}^*(\mathcal{S}) = \mathcal{Z}. \quad (14)$$

The problem (14) can be solved using linear conjugate gradient algorithm over \mathcal{Z} and \mathcal{S} , and the equality constraint is ensured at each step by performing a projection step. The cost $g(U)$, the gradient $\nabla g(U)$, and the directional derivative $D\nabla g(U)[V]$ can be easily computed following Lemma 3 and Remark 4 above. We employ Riemannian trust-region algorithm to solve (13). The per-iteration complexity of the proposed algorithm is $O\left(TI|\Omega| \sum_{k=1}^K r_k + \sum_{k=1}^K n_k r_k^2 + \sum_{k=1}^K r_k^3\right)$, where T is the iterations of CG and $I|\Omega|$ is the number of non-zero entries of \mathcal{S} .

7. Toy Experiment

We consider a toy problem of size $100 \times 100 \times 3$ with 10% train data for both Nonnegative and Hankel Tensor Completion problems. The ranks (r_1, r_2, r_3) are chosen as $(10, 10, 3)$ for both the experiments, and the regularization constants λ_k 's are chosen according to [5]. The maximum iterations for Riemannian optimization problem was set to 200. The duplication parameter τ in the Hankel transform was set to $(10, 10, 1)$. The plots of the variation in gradient norm and relative duality gap with iterations for both problems is shown in Appendix A in Figures 1 and 2. We observe that in both the cases the plots of gradient norm, and the relative duality gap decreases rapidly with iterations.

References

- [1] Boumal, N.: An introduction to optimization on smooth manifolds, 2020. Accessed online from <http://www.nicolasboumal.net/book>
- [2] Kolda, T. G. and Bader, B. W.: Tensor decompositions and applications, SIAM Review, 51(3), pp. 455–500, 2009.
- [3] Absil, P.-A., Mahony, R. and Sepulchre, R.: Optimization Algorithms on Matrix Manifolds, Princeton University Press, Princeton, NJ, 2008. Accessed online from <https://press.princeton.edu/absil>

- [4] Journée, M., Bach, F., Absil, P.-A. and Sepulchre, R.: Low-rank optimization on the cone of positive semidefinite matrices, *SIAM Journal on Optimization*, 20(5), pp. 2327-2351, 2010.
- [5] Nimishakavi, M., Jawanpuria, P. and Mishra, B.: A Dual Framework for Trace Norm Regularized Low-rank Tensor Completion, *Conference on Neural Information Processing Systems (NeurIPS)*, 2018.
- [6] Argyriou, A., Evgeniou, T. and Pontil, M.: Multi-task feature learning, *NIPS*, 2006.
- [7] Liu, J., Musialski, P., Wonka, P., and Ye, J.: Tensor completion for estimating missing values in visual data, *IEEE transactions on pattern analysis and machine intelligence*, 35(1), pp. 208-220, 2012.
- [8] Kressner, D., Steinlechner, M. and Vandereycken, B.: Low-rank tensor completion by Riemannian optimization, *BIT Numerical Mathematics*, 54(2), pp. 447-468, 2014.
- [9] Kasai, H. and Mishra, B.: Low-rank tensor completion: a Riemannian manifold preconditioning approach, *International conference on machine learning (PMLR)*, pp. 1012-1021, 2016.
- [10] Xu, Y. and Yin, W.: A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion, *SIAM Journal on imaging sciences*, 6(3), pp. 1758-1789, 2013.
- [11] Signoretto, M., Dinh, Q. T., De Lathauwer, L., and Suykens, J. A.: Learning with tensors: a framework based on convex optimization and spectral regularization, *Machine Learning*, 94(3), pp. 303-351, 2014.
- [12] Cheng, H., Yu, Y., Zhang, X., Xing, E., and Schuurmans, D.: Scalable and sound low-rank tensor learning, *Artificial Intelligence and Statistics (PMLR)*, pp. 1114-1123, 2016.
- [13] Jawanpuria, P. and Mishra, B.: A unified framework for structured low-rank matrix learning, *ICML*, 2018.
- [14] Sinha, T.K., Naram, J. and Kumar, P.: Nonnegative Low-Rank Tensor Completion via Dual Formulation with Applications to Image and Video Completion, *Winter Conference on Applications of Computer Vision (WACV)*, 2022.
- [15] Guo, X., Yao, Q. and Kwok, J. T.: Efficient sparse low-rank tensor completion using the frank-wolfe algorithm, *AAAI*, 2017.
- [16] Kim, D., Sra, S. and Dhillon, I. S.: A non-monotonic method for large-scale non-negative least squares, *Optimization Methods and Software*, 28(5), pp. 1012–1039, 2013.
- [17] Yokota, T., Erem, B., Guler, S., Warfield, S. K. and Hontani, H.: Missing Slice Recovery for Tensors Using a Low-Rank Model in Embedded Space, *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8251-8259, 2018.
- [18] Ying, J., Lu, H., Wei, Q., Cai, J.F., Guo, D., Wu, J., Chen, Z., Qu, X.: Hankel Matrix Nuclear Norm Regularized Tensor Completion for N -dimensional Exponential Signals, *IEEE Transactions on Signal Processing*, 65(14), pp. 3702-3717, 2017.

- [19] Wang, X., Wu, Y., Zhuang, D., Sun, L.: Low-Rank Hankel Tensor Completion for Traffic Speed Estimation, arXiv preprint arXiv:2105.11335, 2021.
- [20] Chen, B., Sun, T., Zhou, Z., Zeng, Y. and Cao, L.: Nonnegative tensor completion via low-rank Tucker decomposition: model and algorithm, IEEE Access, 7, pp. 95903-95914, 2019.
- [21] Boumal, N. and Absil, P.-A.: Low-rank matrix completion via preconditioned optimization on the Grassmann manifold, Linear Algebra and its Applications, vol. 475, pp. 200-239, 2015.

Appendix A. Experimental plots

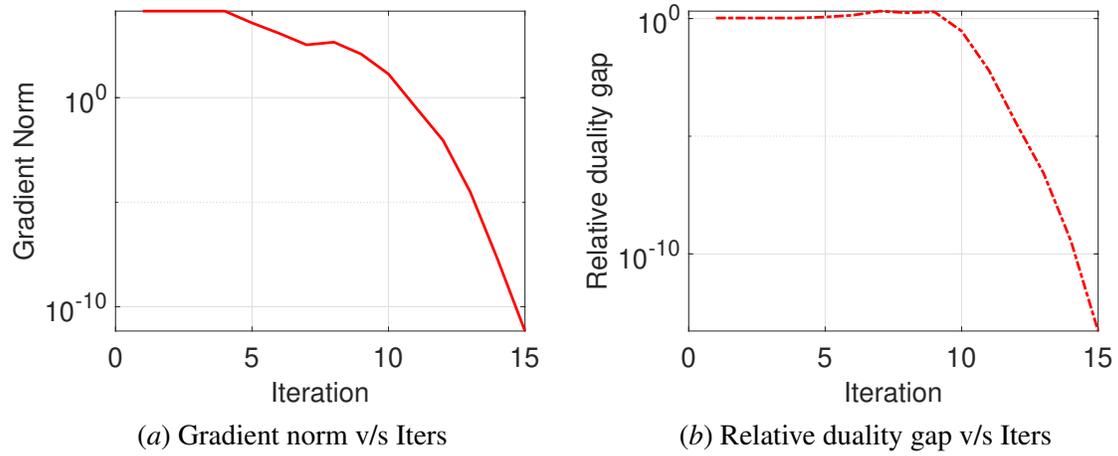


Figure 1: Variation of gradient norm and relative duality gap with iterations for Nonnegative Tensor Completion problem.

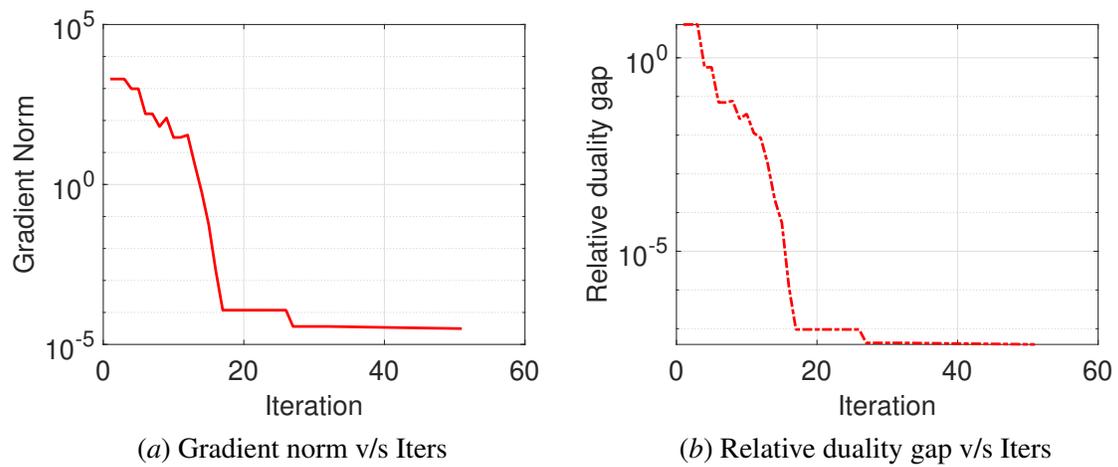


Figure 2: Variation of gradient norm and relative duality gap with iterations for Hankel Tensor Completion problem.

Appendix B. Proof of Dual Formulation (Proof of Theorem 2).

Proof Consider the following subproblem of the above problem

$$\begin{aligned} \min_{\substack{\mathcal{W}^{(k)}, \\ k \in \{1, \dots, K\}}} \quad & C \left\| \left(\sum_{k=1}^K \mathcal{W}^{(k)} \right)_{\Omega} - \mathcal{Y}_{\Omega} \right\|^2 + \sum_{k=1}^K \frac{1}{2\lambda_k} \langle \Theta_k^{\dagger} \mathcal{W}_k^{(k)}, \mathcal{W}_k^{(k)} \rangle. \\ \text{subject to} \quad & A(\mathcal{W}) \geq 0. \end{aligned} \quad (15)$$

We introduce auxiliary variables U_k with the associated constraints $U_k = \mathcal{W}_k^{(k)}$. Now, we construct the Lagrangian of the problem

$$\begin{aligned} \mathcal{L}(\mathcal{W}^{(1)}, \dots, \mathcal{W}^{(K)}, U_1, \dots, U_K, \Lambda_1, \dots, \Lambda_K, s) = & C \left\| \left(\sum_{k=1}^K \mathcal{W}^{(k)} \right)_{\Omega} - \mathcal{Y}_{\Omega} \right\|^2 + \\ & \sum_{k=1}^K \frac{1}{2\lambda_k} \langle \Theta_k^{\dagger} U_k, U_k \rangle + \sum_{k=1}^K \langle \Lambda_k, \mathcal{W}_k^{(k)} - U_k \rangle - \langle s, A(\mathcal{W}) \rangle, \end{aligned} \quad (16)$$

where $\Lambda_1, \dots, \Lambda_K, s$ are the Lagrange multipliers. The dual function of the above will be given by

$$q(\Theta_1, \dots, \Theta_K, \Lambda_1, \dots, \Lambda_K, s) = \min_{\substack{\mathcal{W}^{(k)}, U_k, \\ k \in \{1, \dots, K\}}} \mathcal{L}(\mathcal{W}^{(1)}, \dots, \mathcal{W}^{(K)}, U_1, \dots, U_K, \Lambda_1, \dots, \Lambda_K, s).$$

Taking the derivative of the Lagrangian with respect to $\mathcal{W}^{(k)}$ and equating to 0, we get

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathcal{W}^{(k)}} = 0 \implies & 2C \left[\left(\sum_{k=1}^K \mathcal{W}^{(k)} \right)_{\Omega} - \mathcal{Y}_{\Omega} \right] + \text{fold}_k(\Lambda_k) - A^*(s) = 0 \\ \implies \text{fold}_k(\Lambda_k) = & 2C \left[\mathcal{Y}_{\Omega} - \left(\sum_{k=1}^K \mathcal{W}^{(k)} \right)_{\Omega} \right] + A^*(s), \end{aligned} \quad (17)$$

where A^* is the adjoint of A . The right-hand side of (17) is independent of k , so $\text{fold}_k(\Lambda_k)$ is also independent of k . Then we have,

$$\text{fold}_k(\Lambda_k) = \mathcal{Z} + A^*(s), \text{ where } \mathcal{Z} = 2C \left[\mathcal{Y}_{\Omega} - \left(\sum_{k=1}^K \mathcal{W}^{(k)} \right)_{\Omega} \right]. \quad (18)$$

Next, we minimize w.r.t. U_k , to get

$$U_k = \lambda_k \Theta_k \Lambda_k. \quad (19)$$

Now we compute (16) term by term by using (18) and (19)

$$C \left\| \left(\sum_{k=1}^K \mathcal{W}^{(k)} \right)_{\Omega} - \mathcal{Y}_{\Omega} \right\|^2 = C \left(\frac{1}{4C^2} \|\mathcal{Z}\|^2 \right) = \frac{1}{4C} \|\mathcal{Z}\|^2, \quad (20)$$

$$\begin{aligned}
 \sum_{k=1}^K \frac{1}{2\lambda_k} \langle \Theta_k^\dagger U_k, U_k \rangle - \langle \Lambda_k, U_k \rangle &= \frac{1}{2\lambda_k} \langle \Theta_k^\dagger U_k - 2\lambda_k \Lambda_k, U_k \rangle \\
 &= \sum_{k=1}^K \frac{1}{2\lambda_k} \langle \Theta_k^\dagger \lambda_k \Theta_k [Z_k + (A^*(s))_k] - 2\lambda_k [Z_k + (A^*(s))_k], \lambda_k \Theta_k [Z_k + (A^*(s))_k] \rangle \\
 &= - \sum_{k=1}^K \frac{\lambda_k}{2} \langle Z_k + (A^*(s))_k, \Theta_k [Z_k + (A^*(s))_k] \rangle, \tag{21}
 \end{aligned}$$

$$\begin{aligned}
 \sum_{k=1}^K \langle \Lambda_k, W_k^{(k)} \rangle - \langle A^*(s), \mathcal{W} \rangle &= \sum_{k=1}^K \langle \text{fold}_k(\Lambda_k), \mathcal{W}^{(k)} \rangle - \langle A^*(s), \mathcal{W} \rangle \\
 &= \langle \mathcal{Z} + A^*(s), \sum_{k=1}^K \mathcal{W}^{(k)} \rangle - \langle A^*(s), \mathcal{W} \rangle \\
 &= \langle \mathcal{Z} + A^*(s), \mathcal{W} \rangle - \langle A^*(s), \mathcal{W} \rangle = \langle \mathcal{Z}, \mathcal{W} \rangle \stackrel{(*)}{=} \langle \mathcal{Z}, \mathcal{W}_\Omega \rangle \\
 &= \left\langle \mathcal{Z}, \left(\sum_{k=1}^K \mathcal{W}^{(k)} \right)_\Omega \right\rangle \\
 &= \langle \mathcal{Z}, \mathcal{Y}_\Omega - (1/2C)\mathcal{Z} \rangle = \langle \mathcal{Z}, \mathcal{Y}_\Omega \rangle - \frac{1}{2C} \|\mathcal{Z}\|^2, \tag{22}
 \end{aligned}$$

where $(*)$ is obtained by noting that $\mathcal{Z} = \mathcal{Z}_\Omega$ by definition in (18). Summing (20), (21), (22), we obtain the expression for the dual function as

$$q(\Theta_1, \dots, \Theta_K, s) = \langle \mathcal{Z}, \mathcal{Y}_\Omega \rangle - \frac{1}{4C} \|\mathcal{Z}\|^2 - \sum_{k=1}^K \frac{\lambda_k}{2} \langle Z_k + (A^*(s))_k, \Theta_k [Z_k + (A^*(s))_k] \rangle.$$

This gives the minimax problem (4). From (18) and (19), we can deduce the said relation between optimal points of primal and minimax problems. \blacksquare

Appendix C. Proof of Duality Gap (Proof of Theorem 5).

Proof Following [5], we rewrite problem (4) in the form

$$\min_{\Theta \in P^{n_1} \times \dots \times P^{n_K}} G_1(\Theta),$$

where

$$G_1(\Theta) = \max_{\mathcal{Z} \in \mathcal{C}, s \in \mathbb{R}^n} \langle \mathcal{Z}, \mathcal{Y}_\Omega \rangle - \frac{1}{4C} \|\mathcal{Z}\|^2 - \sum_{k=1}^K \frac{\lambda_k}{2} \langle Z_k + (A^*(s))_k, \Theta_k [Z_k + (A^*(s))_k] \rangle.$$

Now, using the min-max interchange, the max-min problem equivalent to 4 can be written as

$$\max_{\mathcal{Z} \in \mathcal{C}, s \in \mathbb{R}^n} G_2(\mathcal{Z}, s),$$

where

$$\begin{aligned}
 G_2(\mathcal{Z}, s) &= \min_{\Theta \in P^{n_1} \times \dots \times P^{n_K}} \langle \mathcal{Z}, \mathcal{Y}_\Omega \rangle - \frac{1}{4C} \|\mathcal{Z}\|^2 - \sum_{k=1}^K \frac{\lambda_k}{2} \langle Z_k + (A^*(s))_k, \Theta_k [Z_k + (A^*(s))_k] \rangle \\
 &= \langle \mathcal{Z}, \mathcal{Y}_\Omega \rangle - \frac{1}{4C} \|\mathcal{Z}\|^2 - \sum_{k=1}^K \frac{\lambda_k}{2} \max_{\Theta_k \in P^{n_k}} \langle Z_k + (A^*(s))_k, \Theta_k [Z_k + (A^*(s))_k] \rangle \\
 &= \langle \mathcal{Z}, \mathcal{Y}_\Omega \rangle - \frac{1}{4C} \|\mathcal{Z}\|^2 - \sum_{k=1}^K \frac{\lambda_k}{2} B_k(\mathcal{Z}, s),
 \end{aligned}$$

where

$$B_k(\mathcal{Z}, s) = \max_{\Theta_k \in P^{n_k}} \langle Z_k + (A^*(s))_k, \Theta_k [Z_k + (A^*(s))_k] \rangle.$$

We know that the value of $B_k(\mathcal{Z}, s)$ is equal to the largest eigenvalue of $(Z_k + (A^*(s))_k)(Z_k + (A^*(s))_k)^T$. Equivalently, if σ_k is the largest singular value of $Z_k + (A^*(s))_k$, then $B_k(\mathcal{Z}, s) = \sigma_k^2$. Now, the duality gap will be given by

$$\Delta = G_1(\hat{\Theta}) - G_2(\hat{\mathcal{Z}}, \hat{s}) = \sum_{k=1}^K \frac{\lambda_k}{2} \left(\sigma_k^2 - \|\hat{U}_k^T (\hat{Z}_k + \hat{A}_k)\|^2 \right),$$

where \hat{U} , \hat{Z} , and \hat{s} were defined in Theorem 5. ■