3A-05

Riemannian preconditioning for tensor completion

Hiroyuki Kasai*
The University of Electro-Communications

Bamdev Mishra
Amazon Development Centre India

1 Introduction

This paper addresses the problem of low-rank tensor completion when the rank is a priori known or estimated. Without loss of generality, we focus on 3-order tensors. Given a tensor $\boldsymbol{\mathcal{X}}^{n_1\times n_2\times n_3}$, whose entries $\boldsymbol{\mathcal{X}}^\star_{i_1,i_2,i_3}$ are only known for some indices $(i_1,i_2,i_3)\in\Omega$, where Ω is a subset of the complete set of indices $\{(i_1,i_2,i_3): i_d\in\{1,\ldots,n_d\}, d\in\{1,2,3\}\}$, the *fixed-rank tensor completion problem* is formulated as

$$\begin{array}{ll}\displaystyle\min_{\boldsymbol{\mathcal{X}}\in\mathbb{R}^{n_1\times n_2\times n_3}} & \frac{1}{|\Omega|}\|\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}})-\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}}^\star)\|_F^2\\ \text{subject to} & \mathrm{rank}(\boldsymbol{\mathcal{X}})=\mathbf{r},\end{array}\tag{1}$$

where the operator $\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}})_{i_1i_2i_3}=\boldsymbol{\mathcal{X}}_{i_1i_2i_3}$ if $(i_1,i_2,i_3)\in\Omega$ and $\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}})_{i_1i_2i_3}=0$ otherwise and (with a slight abuse of notation) $\|\cdot\|_F$ is the Frobenius norm. $\mathrm{rank}(\boldsymbol{\mathcal{X}})$ $(=\mathbf{r}=(r_1,r_2,r_3))$, called the *multilinear rank* of $\boldsymbol{\mathcal{X}}$, is the set of the ranks of for each of mode-d unfolding matrices. $r_d\ll n_d$ enforces a low-rank structure. The optimization problem (1) has many variants [1, 2, 3]. We exploits *Tucker decomposition* [4, Section 4] of a low-rank tensor $\boldsymbol{\mathcal{X}}$ to develop large-scale algorithms for (1), e.g., in [5, 6]. The present paper exploits both the *symmetry* present in Tucker decomposition and the *least-squares* structure of the cost function of (1) by using the concept of *preconditioning*. We build upon the recent work [7] that suggests to use *Riemannian preconditioning* with a *tailored metric* (inner product) in the Riemannian optimization framework on quotient manifolds [8, 9, 10]. Our proposed preconditioned nonlinear conjugate gradient algorithm is implemented in the Matlab toolbox Manopt [11] and it outperforms state-of-the-art methods. We also provide a *generic* Manopt factory (a manifold description Matlab file).

2 A new metric and geometry

The quotient and least-squares structures. The Tucker decomposition of a tensor $\boldsymbol{\mathcal{X}}\in\mathbb{R}^{n_1\times n_2\times n_3}$ of rank $\mathbf{r}$ $(=(r_1,r_2,r_3))$ is [4, Section 4.1] $\boldsymbol{\mathcal{X}}=\boldsymbol{\mathcal{G}}\times_1\mathbf{U}_1\times_2\mathbf{U}_2\times_3\mathbf{U}_3$, where $\mathbf{U}_d\in\mathrm{St}(r_d,n_d)$ for $d\in\{1,2,3\}$ belongs to the *Stiefel manifold* of matrices of size $n_d\times r_d$ with orthogonal columns and $\boldsymbol{\mathcal{G}}\in\mathbb{R}^{r_1\times r_2\times r_3}$. Tucker decomposition is *not unique* as $\boldsymbol{\mathcal{X}}$ remains unchanged under the transformation $(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})\mapsto(\mathbf{U}_1\mathbf{O}_1,\mathbf{U}_2\mathbf{O}_2,\mathbf{U}_3\mathbf{O}_3,\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)$ for all $\mathbf{O}_d\in\mathcal{O}(r_d)$, which is the set of orthogonal matrices of size of $r_d\times r_d$. We encode the transformation in an abstract search space of *equivalence classes*, defined as, $[(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})]:=\{(\mathbf{U}_1\mathbf{O}_1,\mathbf{U}_2\mathbf{O}_2,\mathbf{U}_3\mathbf{O}_3,\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T):\mathbf{O}_d\in\mathcal{O}(r_d)\}$. The set of equivalence classes is the quotient manifold [12, Theorem 9.16] $\mathcal{M}/\sim:=\mathcal{M}/(\mathcal{O}(r_1)\times\mathcal{O}(r_2)\times\mathcal{O}(r_3))$, where $\mathcal{M}$ is called the *total space* that is the product space $\mathcal{M}:=\mathrm{St}(r_1,n_1)\times\mathrm{St}(r_2,n_2)\times\mathrm{St}(r_3,n_3)\times\mathbb{R}^{r_1\times r_2\times r_3}$. Due to the invariance of the Tucker decomposition, the local minima of (1) in $\mathcal{M}$ are not isolated, but they become isolated on $\mathcal{M}/\sim$. Consequently, the problem (1) is an optimization problem on a quotient manifold [8, 9, 10] by endowing $\mathcal{M}/\sim$ with a Riemannian structure. Another structure that is present in (1) is the least-squares structure of the cost function. A way to exploit it is to endow the search space with a metric (inner product) induced by the Hessian of the cost function [13]. Since applying this approach [7, Section 5] directly for (1) is computationally costly, we consider a simplified cost function by assuming that Ω contains the full set of indices, i.e., we focus on $\|\boldsymbol{\mathcal{X}}-\boldsymbol{\mathcal{X}}^\star\|_F^2$. The block diagonal approximation of the Hessian of $\|\boldsymbol{\mathcal{X}}-\boldsymbol{\mathcal{X}}^\star\|_F^2$ in $(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})$ is $((\mathbf{G}_1\mathbf{G}_1^T)\otimes\mathbf{I}_{n_1},(\mathbf{G}_2\mathbf{G}_2^T)\otimes\mathbf{I}_{n_2},(\mathbf{G}_3\mathbf{G}_3^T)\otimes\mathbf{I}_{n_3},\mathbf{I}_{r_1r_2r_3})$, where $\mathbf{G}_d$ is the mode-d unfolding of $\boldsymbol{\mathcal{G}}$.

A novel Riemannian metric and its motivation. An element x in the total space $\mathcal{M}$ has the matrix representation $(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})$. Consequently, the tangent space $T_x\mathcal{M}$ is the Cartesian product of the tangent spaces of the individual manifolds, i.e., $T_x\mathcal{M}$ has the matrix characterization [10] $T_x\mathcal{M}=\{(\mathbf{Z}_{\mathbf{U}_1},\mathbf{Z}_{\mathbf{U}_2},\mathbf{Z}_{\mathbf{U}_3},\mathbf{Z}_{\boldsymbol{\mathcal{G}}})\in\mathbb{R}^{n_1\times r_1}\times\mathbb{R}^{n_2\times r_2}\times\mathbb{R}^{n_3\times r_3}\times\mathbb{R}^{r_1\times r_2\times r_3}:\mathbf{U}_d^T\mathbf{Z}_{\mathbf{U}_d}+\mathbf{Z}_{\mathbf{U}_d}^T\mathbf{U}_d=0,\text{ for } d\in\{1,2,3\}\}$. The earlier discussion on symmetry and least-squares structure leads to the novel metric $g_x: T_x\mathcal{M}\times T_x\mathcal{M}\to\mathbb{R}$

$$\begin{array}{ll}g_x(\xi_x,\eta_x) & =\langle\xi_{\mathbf{U}_1},\eta_{\mathbf{U}_1}(\mathbf{G}_1\mathbf{G}_1^T)\rangle+\langle\xi_{\mathbf{U}_2},\eta_{\mathbf{U}_2}(\mathbf{G}_2\mathbf{G}_2^T)\rangle\\ & +\langle\xi_{\mathbf{U}_3},\eta_{\mathbf{U}_3}(\mathbf{G}_3\mathbf{G}_3^T)\rangle+\langle\xi_{\boldsymbol{\mathcal{G}}},\eta_{\boldsymbol{\mathcal{G}}}\rangle,\end{array}\tag{2}$$

where $\xi_x,\eta_x\in T_x\mathcal{M}$ are tangent vectors with matrix characterizations, $(\xi_{\mathbf{U}_1},\xi_{\mathbf{U}_2},\xi_{\mathbf{U}_3},\xi_{\boldsymbol{\mathcal{G}}})$ and $(\eta_{\mathbf{U}_1},\eta_{\mathbf{U}_2},\eta_{\mathbf{U}_3},\eta_{\boldsymbol{\mathcal{G}}})$, respectively and $\langle\cdot,\cdot\rangle$ is the Euclidean inner product. As contrasts to the classical Euclidean metric, the metric (2) *scales* the level sets of the cost function on the search space that leads a preconditioning effect on the algorithms.

*This work was initiated while Bamdev Mishra was with the Department of Electrical Engineering and Computer Science, University of Liège, 4000 Liège, Belgium and was visiting the Department of Engineering (Control Group), University of Cambridge, Cambridge, UK. H. Kasai is (partly) supported by the Ministry of Internal Affairs and Communications, Japan, as the SCOPE Project (150201002). B. Mishra was supported as an FNRS research fellow (Belgian Fund for Scientific Research). The scientific responsibility rests with its authors.

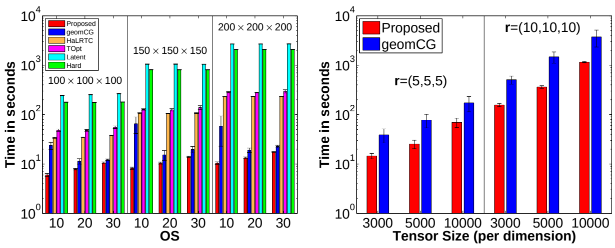
Table 1: Ingredients to implement an off-the-shelf conjugate gradient algorithm for (1).

Vertical tangent vectors in $\mathcal{V}_x$	$\{(\mathbf{U}_1\mathbf{\Omega}_1, \mathbf{U}_2\mathbf{\Omega}_2, \mathbf{U}_3\mathbf{\Omega}_3, -(\boldsymbol{\mathcal{G}}\times_1\mathbf{\Omega}_1 + \boldsymbol{\mathcal{G}}\times_2\mathbf{\Omega}_2 + \boldsymbol{\mathcal{G}}\times_3\mathbf{\Omega}_3)) : \mathbf{\Omega}_d \in \mathbb{R}^{r_d \times r_d}, \mathbf{\Omega}_d^T = -\mathbf{\Omega}_d, \text{for } d \in \{1,2,3\}\}$
Horizontal tangent vectors in $\mathcal{H}_x$	$\{(\zeta_{\mathbf{U}_1}, \zeta_{\mathbf{U}_2}, \zeta_{\mathbf{U}_3}, \zeta_{\boldsymbol{\mathcal{G}}}) \in T_x\mathcal{M} : (\mathbf{G}_d\mathbf{G}_d^T)\zeta_{\mathbf{U}_d}^T\mathbf{U}_d + \zeta_{\mathbf{G}_d}\mathbf{G}_d^T \text{ is symmetric, for } d \in \{1,2,3\}\}$
$\Psi(\cdot)$ projects an ambient vector $(\mathbf{Y}_{\mathbf{U}_1}, \mathbf{Y}_{\mathbf{U}_2}, \mathbf{Y}_{\mathbf{U}_3}, \mathbf{Y}_{\boldsymbol{\mathcal{G}}})$ onto $T_x\mathcal{M}$	$(\mathbf{Y}_{\mathbf{U}_1} - \mathbf{U}_1\mathbf{S}_{\mathbf{U}_1}(\mathbf{G}_1\mathbf{G}_1^T)^{-1}, \mathbf{Y}_{\mathbf{U}_2} - \mathbf{U}_2\mathbf{S}_{\mathbf{U}_2}(\mathbf{G}_2\mathbf{G}_2^T)^{-1}, \mathbf{Y}_{\mathbf{U}_3} - \mathbf{U}_3\mathbf{S}_{\mathbf{U}_3}(\mathbf{G}_3\mathbf{G}_3^T)^{-1}, \mathbf{Y}_{\boldsymbol{\mathcal{G}}})$, where $\mathbf{S}_{\mathbf{U}_d}$ for $d \in \{1,2,3\}$ are solutions to $\mathbf{S}_{\mathbf{U}_d}\mathbf{G}_d\mathbf{G}_d^T + \mathbf{G}_d\mathbf{G}_d^T\mathbf{S}_{\mathbf{U}_d} = \mathbf{G}_d\mathbf{G}_d^T(\mathbf{Y}_{\mathbf{U}_d}^T\mathbf{U}_d + \mathbf{U}_d^T\mathbf{Y}_{\mathbf{U}_d})\mathbf{G}_d\mathbf{G}_d^T$
$\Pi(\cdot)$ projects a tangent vector ξ onto $\mathcal{H}_x$	$(\xi_{\mathbf{U}_1} - \mathbf{U}_1\mathbf{\Omega}_1, \xi_{\mathbf{U}_2} - \mathbf{U}_2\mathbf{\Omega}_2, \xi_{\mathbf{U}_3} - \mathbf{U}_3\mathbf{\Omega}_3, \xi_{\boldsymbol{\mathcal{G}}} - (-(\boldsymbol{\mathcal{G}}\times_1\mathbf{\Omega}_1 + \boldsymbol{\mathcal{G}}\times_2\mathbf{\Omega}_2 + \boldsymbol{\mathcal{G}}\times_3\mathbf{\Omega}_3)))$, where $\mathbf{\Omega}_d$ are solutions to particular *coupled* Lyapunov equations.
$\mathrm{egrad}_x f$	$(\mathbf{S}_1(\mathbf{U}_3 \otimes \mathbf{U}_2)\mathbf{G}_1^T(\mathbf{G}_1\mathbf{G}_1^T)^{-1}, \mathbf{S}_2(\mathbf{U}_3 \otimes \mathbf{U}_1)\mathbf{G}_2^T(\mathbf{G}_2\mathbf{G}_2^T)^{-1}, \mathbf{S}_3(\mathbf{U}_2 \otimes \mathbf{U}_1)\mathbf{G}_3^T(\mathbf{G}_3\mathbf{G}_3^T)^{-1}, \boldsymbol{\mathcal{S}} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \times_3 \mathbf{U}_3^T) \times_3 \mathbf{U}_3^T)$, where $\boldsymbol{\mathcal{S}} = \frac{2}{\lvert\Omega\rvert}(\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{G}}\times_1\mathbf{U}_1\times_2\mathbf{U}_2\times_3\mathbf{U}_3) - \boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}}^\star))$.

A new geometry and the conjugate gradient. Based on this proposed Riemannian metric, the new geometry is finally formulated as Table 1. This is the ingredients to implement the Riemannian conjugate gradient algorithm [8, Section 8.3].

3 Numerical comparisons

We show numerical comparisons of our proposed algorithm with state-of-the-art algorithms that include TOpt [5] and geomCG [6], for comparisons with Tucker decomposition based algorithms, and HaLRTC [1], Latent [2], and Hard [3] as nuclear norm minimization algorithms. **Case 1** considers synthetic small-scale tensors of size $100 \times 100 \times 100$, $150 \times 150 \times 150$, and $200 \times 200 \times 200$ and rank $\mathbf{r} = (10, 10, 10)$ are considered. OS is $\{10, 20, 30\}$. Figure 1(a) shows that the convergence behavior of our proposed algorithm is either competitive or faster than the others. Next, **Case 2** considers large-scale tensors of size $3000 \times 3000 \times 3000$, $5000 \times 5000 \times 5000$, and $10000 \times 10000 \times 10000$ and ranks $\mathbf{r} = (5, 5, 5)$ and $(10, 10, 10)$. OS is 10. Our proposed algorithm outperforms geomCG in Figure 1(b).



(a) **Case 1:** small-scale tensors. (b) **Case 2:** large-scale tensors.

Fig. 1: Experiments results.

4 Conclusion

We have proposed a preconditioned nonlinear conjugate gradient algorithm for the tensor completion problem by exploiting the fundamental structures of symmetry, due to non-uniqueness of Tucker decomposition, and least-squares of the cost function. The full version of this paper is on [14].

References

[1] J. Liu, P. Musialski, P. Wonka, and J. Ye. Tensor completion for estimating missing values in visual data. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(1):208–220, 2013.

[2] R. Tomioka, K. Hayashi, and H. Kashima. Estimation of low-rank tensors via convex optimization. *arXiv*, 2011.

[3] M. Signoretto, Q. T. Dinh, L. D. Lathauwer, and J. A.K. Suykens. Learning with tensors: a framework based on convex optimization and spectral regularization. *Mach. Learn.*, 94(3):303–351, 2014.

[4] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.

[5] M. Filipović and A. Jukić. Tucker factorization with missing data with application to low- n -rank tensor completion. *Multidim. Syst. Sign. P.*, 2013.

[6] D. Kressner, M. Steinlechner, and B. Vandereycken. Low-rank tensor completion by Riemannian optimization. *BIT Numer. Math.*, 54(2):447–468, 2014.

[7] B. Mishra and R. Sepulchre. Riemannian preconditioning. *arXiv*, 2014.

[8] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds.* Princeton University Press, 2008.

[9] S. T. Smith. Optimization techniques on Riemannian manifold. In A. Bloch, editor, *Hamiltonian and Gradient Flows, Algorithms and Control*, volume 3, pages 113–136. Amer. Math. Soc., Providence, RI, 1994.

[10] A. Edelman, T.A. Arias, and S.T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2):303–353, 1998.

[11] N. Boumal and R. Sepulchre B. Mishra, P.A. Absil. Manopt: a matlab toolbox for manopt: a Matlab toolbox foroptimization on manifolds. *JMLR*, 15(1):1455–1459, 2014.

[12] J. M. Lee. *Introduction to smooth manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 2003.

[13] J. Nocedal and S. J. Wright. *Numerical Optimization.* Springer Series in Operations Research. Springer, 2006.

[14] Hiroyuki Kasai and Bamdev Mishra. Riemannian preconditioning for tensor completion. *arXiv*, 2015.