Scaled Tucker manifold and its application to large-scale machine learning

笠井 裕之[1,a)]　Bamdev Mishra[2]

概要：本稿では，低ランク・テンソル Tucker 分解のための新しい幾何空間 “Scaled Tucker Manifold” を提案する．一般的なテンソル回帰問題に対して，Scaled Tucker Manifold により，効率的な解法を確立することが可能となる．Scaled Tucker Manifol の導出にあたっては，Tucker 分解の対称構造と回帰問題の最小自乗構造に着目した新しいリーマン計量を提案し，幾何空間を定義する数々の構成要素を導出する．最後に，回帰問題の一形態である “テンソル補完問題” を例に取り挙げ，シミューレション実験から，Scaled Tucker Manifold に基づき導出した非線形共役勾配法アルゴリズムが，従来の最先端手法と比較してより良い性能を与えることを示す．

Scaled Tucker manifold and its application to large-scale machine learning

Hiroyuki Kasai[1,a)]　Bamdev Mishra[2]

Abstract: We propose a novel geometry for dealing with low-rank Tucker decomposition of tensors. The geometry of the scaled Tucker manifold readily allows to develop algorithms for a number of regression problems. Specifically, we propose a novel scaled Riemannian metric (an inner product) that suits well to least-squares cost. The simulation experiments on the tensor completion problem show that our proposed nonlinear conjugate gradient algorithm outperforms state-of-the-art algorithms.

1. Introduction

We address the optimization problem

$$\min_{\mathcal{W}\in\mathbb{R}^{n_1\times\cdots\times n_d}} f(\mathcal{W}) := \frac{1}{n}\sum_{i=1}^{n}\|y_i - \langle \mathcal{W}, \mathcal{X}_i\rangle\|^2, \quad (1)$$

where $\mathcal{W}$ and $\mathcal{X}$ are d-order tensors of size $\mathbb{R}^{n_1\times\ldots\times n_d}$, and y_i is a scalar. $\langle \mathcal{A}, \mathcal{B}\rangle$ is the sum of element-wise multiplications of $\mathcal{A}$ and $\mathcal{B}$, i.e., $\langle \mathcal{A}, \mathcal{B}\rangle = \mathrm{vec}(\mathcal{A})^T\mathrm{vec}(\mathcal{B})$, where $\mathrm{vec}(\cdot)$ is the *vectorization* of a tensor.

The problem formulation (1) has a number of applications. For example, if $y_{p,q,r}(= y_i) \in \Omega$ is each known observation, where the set Ω is a subset of the complete set of entries, and $\mathcal{X}_{p,q,r}(= \mathcal{X}_i) = e_p \circ e_q \circ e_r$, where $e_p \in \mathbb{R}^{n_1}$, $e_q \in \mathbb{R}^{n_2}$, and $e_r \in \mathbb{R}^{n_3}$ are canonical basis vectors, this casts into a *tensor completion* problem [1], where $n = |\Omega|$ is the number of known entries. If Ω is the complete set, i.e., $n = |\Omega| = n_1n_2n_3$, the problem is a *tensor decomposition* problem. The problem (1) is a *tensor regression* problem when $\mathcal{W}$ is a regression coefficient tensor and $\mathcal{X}_i$ and y_i are i-th sample pair, where n is the total number of samples [2].

For large n_d and d, the problem (1) is computationally challenging. Consequently, $\mathcal{W}$ is assumed to have a *low-rank* structure. Equivalently, we impose a new constraint $\mathrm{rank}(\mathcal{W}) = R$, where R is a smaller number [3]. The fixed-rank tensors belong to a smooth matrix manifold and we exploit the versatile framework of Riemannian optimization to develop efficient optimization algorithms [4]. This problem of interest is then

$$\min_{\mathcal{W}\in\mathcal{M}} f(\mathcal{W}) := \frac{1}{n}\sum_{i=1}^{n}\|y_i - \langle \mathcal{W}, \mathcal{X}_i\rangle\|^2, \quad (2)$$

where $\mathcal{M} := \{\mathcal{W} \in \mathbb{R}^{n_1\times\ldots\times n_d} : \mathrm{rank}(\mathcal{W}) = R\}$.

In this paper, we address fixed *multilinear rank* of *Tucker decomposition* of tensors [5]. Without loss of generality, we focus 3-order tensors. It should be noted the multilinear rank constraint forms a smooth manifold [1]. Specifically, we propose a *scaled quotient* geometry of the *Tucker manifold* by exploiting intrinsic *symmetries* of the Tucker format. This proposed geometry builds upon the recent work [6] that suggests to use *manifold preconditioning* with a *tailored metric* (inner product) in the *Riemannian optimization* framework on quotient manifolds [4]. More concretely, a novel *scaled* Riemannian metric or inner product is proposed that exploits *approximate* second-order information of the problems of the type (2), i.e., least-squares cost, as well as the intrinsic *structured symmetry* in the Tucker format. Concrete matrix formulas are derived which allow the use of the manifold optimization toolbox [7].

1 The University of Electro-Communications, 1-5-1, Chofugaoka, Chofu-shi, Tokyo, 182-8585, Japan
2 Amazon Development Centre India, Bangalore, Karnataka 560055, India
a) kasai@is.uec.ac.jp

The differences with respect to the works of [1], [8], which also propose the manifold structure of the Tucker format, are twofold. (i) They exploit the search space as an *embedded submanifold* of the Euclidean space, whereas we view it as a product of simpler search spaces with symmetries. Consequently, certain computations have straightforward interpretation. (ii) They work with the standard Euclidean metric, whereas we use a metric that is tuned to least-squares cost thereby inducing a *preconditioning* effect. This novel idea of using a scaled Riemannian metric leads to a scaled geometry of the Tucker manifold, which is the set of fixed multilinear rank tensors.

The effectiveness of the proposed geometry of the scaled Tucker manifold is shown on the tensor completion problem. We list all the optimization-related ingredient to develop a novel Riemannian nonlinear conjugate gradient algorithm. Our numerical experiments show that our proposed algorithm outperforms state-of-the-art algorithms. The full description the proposed geometry, the various derivations, the algorithm, and the experiments are in [9].

2. Tucker decomposition and a scaled Riemannian metric

We focus on two elements of the problem (2): the *symmetry* structure of the Tucker decomposition and *approximate* second-order information of the cost function.

The symmetry structure in Tucker decomposition. The Tucker decomposition of a tensor $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ of *multilinear* rank R $(=(r_1, r_2, r_3))$ is

$$\boldsymbol{\mathcal{W}} \quad = \quad \boldsymbol{\mathcal{G}} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3, \qquad (3)$$

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 $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ [5]. Here, $\boldsymbol{\mathcal{W}} \times_d \mathbf{V} \in \mathbb{R}^{n_1 \times \ldots n_{d-1} \times m \times n_{d+1} \times \ldots n_D}$ computes the *d-mode product* of a tensor $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{n_1 \times \ldots \times n_D}$ and a matrix $\mathbf{V} \in \mathbb{R}^{m \times n_d}$. Tucker decomposition (3) is *not unique* as $\boldsymbol{\mathcal{W}}$ remains unchanged under the transformation

$$\begin{array}{l}(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \boldsymbol{\mathcal{G}}) \mapsto \\ \quad (\mathbf{U}_1\mathbf{O}_1, \mathbf{U}_2\mathbf{O}_2, \mathbf{U}_3\mathbf{O}_3, \boldsymbol{\mathcal{G}} \times_1 \mathbf{O}_1^T \times_2 \mathbf{O}_2^T \times_3 \mathbf{O}_3^T)\end{array} \qquad (4)$$

for all $\mathbf{O}_d \in \mathcal{O}(r_d)$, which is the set of orthogonal matrices of size of $r_d \times r_d$. The classical remedy to remove this indeterminacy is to have additional structures on $\boldsymbol{\mathcal{G}}$ like sparsity or restricted orthogonal rotations [5], Section 4.3. In contrast, we encode the transformation (4) in an abstract search space of *equivalence classes*, defined as,

$$\begin{array}{rcl}[(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \boldsymbol{\mathcal{G}})] & := & \{(\mathbf{U}_1\mathbf{O}_1, \mathbf{U}_2\mathbf{O}_2, \mathbf{U}_3\mathbf{O}_3, \\ & & \boldsymbol{\mathcal{G}} \times_1 \mathbf{O}_1^T \times_2 \mathbf{O}_2^T \times_3 \mathbf{O}_3^T) : \mathbf{O}_d \in \mathcal{O}(r_d)\}.\end{array} \qquad (5)$$

The set of equivalence classes is the *quotient manifold* [10]

$$\mathcal{M}/\sim \quad := \quad \mathcal{M}/(\mathcal{O}(r_1) \times \mathcal{O}(r_2) \times \mathcal{O}(r_3)), \qquad (6)$$

where $\mathcal{M}$ is called the *total space* (computational space) that is the product space

$$\mathcal{M} := \mathrm{St}(r_1, n_1) \times \mathrm{St}(r_2, n_2) \times \mathrm{St}(r_3, n_3) \times \mathbb{R}^{r_1 \times r_2 \times r_3}. \qquad (7)$$

Due to the invariance (4), the local minima of (2) in $\mathcal{M}$ are not isolated, but they become isolated on $\mathcal{M}/\sim$. Consequently, the problem (2) is an optimization problem on a quotient manifold for which systematic procedures are proposed in [4]. A requirement is to endow $\mathcal{M}/\sim$ with a Riemannian structure, which conceptually translates (2) into an unconstrained optimization problem over the search space $\mathcal{M}/\sim$.

Least-squares structure of the cost function. In unconstrained optimization, the Newton method is interpreted as a *scaled* steepest descent method, where the search space is endowed with a metric (inner product) induced by the Hessian of the cost function. This induced metric (or its approximation) resolves convergence issues of first order optimization algorithms. Analogously, finding a good inner product for (2) is of profound consequence. Specifically for the case of *quadratic* optimization with rank constraint (matrix case), Mishra and Sepulchre [6] propose a family of Riemannian metrics from the Hessian of the cost function. Applying the metric tuning approach of [6] to the cost function (2) leads to a family of Riemannian metrics. To this end, we consider a simplified cost function of (2) by assuming $\boldsymbol{\mathcal{X}}_i = \boldsymbol{\mathcal{I}}$, so that the new geometry of scaled Tucker manifold is independent of a particular cost function. A good trade-off between computational cost and simplicity is by considering only the *block diagonal* elements of the Hessian of (2) with $\boldsymbol{\mathcal{X}}_i = \boldsymbol{\mathcal{I}}$. It should be noted that the cost function (2) is *convex and quadratic* in $\boldsymbol{\mathcal{W}}$. Consequently, it is also (strictly) convex and quadratic in the arguments $(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \boldsymbol{\mathcal{G}})$ individually. The block diagonal approximation of the Hessian of (2) with respect to $(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \boldsymbol{\mathcal{G}})$ is

$$((\mathbf{G}_1\mathbf{G}_1^T) \otimes \mathbf{I}_{n_1}, (\mathbf{G}_2\mathbf{G}_2^T) \otimes \mathbf{I}_{n_2}, (\mathbf{G}_3\mathbf{G}_3^T) \otimes \mathbf{I}_{n_3}, \mathbf{I}_{r_1 r_2 r_3}), \qquad (8)$$

where $\mathbf{G}_d$ is the mode-d unfolding of $\boldsymbol{\mathcal{G}}$ and is assumed to be full rank. $\otimes$ is the Kronecker product. The terms $\mathbf{G}_d\mathbf{G}_d^T$ for $d \in \{1,2,3\}$ are *positive definite* when $r_1 \leq r_2 r_3$, $r_2 \leq r_1 r_3$, and $r_3 \leq r_1 r_2$.

A scaled Riemannian metric. 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 of (7), i.e., $T_x\mathcal{M}$ has the matrix characterization

$$\begin{array}{rcl}T_x\mathcal{M} & = & \{(\mathbf{Z}_{\mathbf{U}_1}, \mathbf{Z}_{\mathbf{U}_2}, \mathbf{Z}_{\mathbf{U}_3}, \mathbf{Z}_{\boldsymbol{\mathcal{G}}}) \\ & & \in \mathbb{R}^{n_1 \times r_1} \times \mathbb{R}^{n_2 \times r_2} \times \mathbb{R}^{n_3 \times r_3} \times \mathbb{R}^{r_1 \times r_2 \times r_3} : \\ & & \mathbf{U}_d^T \mathbf{Z}_{\mathbf{U}_d} + \mathbf{Z}_{\mathbf{U}_d}^T \mathbf{U}_d = 0, \text{ for } d \in \{1,2,3\}\}.\end{array} \qquad (9)$$

From the earlier discussion on symmetry and least-squares structure, we propose the novel metric or inner product $g_x : T_x\mathcal{M} \times T_x\mathcal{M} \rightarrow \mathbb{R}$

$$\begin{array}{rcl}g_x(\xi_x, \eta_x) & = & \langle \xi_{\mathbf{U}_1}, \eta_{\mathbf{U}_1}(\mathbf{G}_1\mathbf{G}_1^T)\rangle + \langle \xi_{\mathbf{U}_2}, \eta_{\mathbf{U}_2}(\mathbf{G}_2\mathbf{G}_2^T)\rangle \\ & & + \langle \xi_{\mathbf{U}_3}, \eta_{\mathbf{U}_3}(\mathbf{G}_3\mathbf{G}_3^T)\rangle + \langle \xi_{\boldsymbol{\mathcal{G}}}, \eta_{\boldsymbol{\mathcal{G}}}\rangle,\end{array} \qquad (10)$$

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, , shown in (9). $\langle \cdot, \cdot \rangle$ is the Euclidean inner product. It should be emphasized that the proposed metric (10) is induced from (8).

We call $\mathcal{M}/\sim$, defined in (6), the *scaled Tucker manifold* as it results from Tucker decomposition endowed with the particular metric (10).

3. Geometry of scaled Tucker manifold

Each point on a quotient manifold represents an entire equivalence class of matrices in the total space. Abstract geometric objects on the quotient manifold $\mathcal{M}/\sim$ call for matrix representatives in the total space $\mathcal{M}$. Similarly, algorithms are run in the total space $\mathcal{M}$, but under appropriate compatibility between the Riemannian structure of $\mathcal{M}$ and the Riemannian structure of the quotient manifold

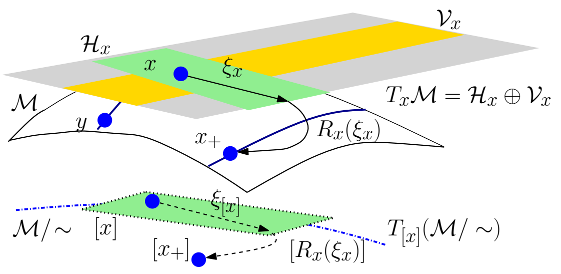


図 1 Riemannian optimization framework: geometric objects, shown in dotted lines, on quotient manifold $\mathcal{M}/\sim$ call for matrix representatives, shown in solid lines, in the total space $\mathcal{M}$ [9].

$\mathcal{M}/\sim$, they define algorithms on the quotient manifold. The key is endowing $\mathcal{M}/\sim$ with a Riemannian structure. Once this is the case, a constraint optimization problem, for example (2), is conceptually transformed into an unconstrained optimization over the Riemannian quotient manifold (6). Below we briefly show the development of various geometric objects that are required to optimize a smooth cost function on the quotient manifold (6) with first order methods, e.g., conjugate gradients.

Figure 1 illustrates a schematic view of optimization with equivalence classes, where the points x and y in $\mathcal{M}$ belong to the same equivalence class (shown in solid blue color) and they represent a single point $[x] := \{y \in \mathcal{M} : y \sim x\}$ on the quotient manifold $\mathcal{M}/\sim$. The abstract tangent space $T_{[x]}(\mathcal{M}/\sim)$ at $[x] \in \mathcal{M}/\sim$ has the matrix representation in $T_x\mathcal{M}$, but restricted to the directions that do not induce a displacement along the equivalence class $[x]$. This is realized by decomposing $T_x\mathcal{M}$ into two complementary subspaces, the vertical and horizontal subspaces. The vertical space $\mathcal{V}_x$ is the tangent space of the equivalence class $[x]$. On the other hand, the horizontal space $\mathcal{H}_x$ is the *orthogonal subspace* to $\mathcal{V}_x$ in the sense of the metric (10). Equivalently, $T_x\mathcal{M} = \mathcal{V}_x \oplus \mathcal{H}_x$.

The horizontal subspace $\mathcal{H}_x$ provides a valid matrix representation to the abstract tangent space $T_{[x]}(\mathcal{M}/\sim)$. An abstract tangent vector $\xi_{[x]} \in T_{[x]}(\mathcal{M}/\sim)$ at $[x]$ has a unique element $\xi_x \in \mathcal{H}_x$ that is called its *horizontal lift*.

From [4], endowed with the Riemannian metric (10), the quotient manifold $\mathcal{M}/\sim$ is a *Riemannian submersion* of $\mathcal{M}$. The submersion principle allows to work out concrete matrix representations of abstract object on $\mathcal{M}/\sim$, e.g., the gradient of a smooth cost function [4]. A key requirement for the Riemannian submersion to be valid is that the metric (10) should satisfy invariance properties.

This holds true for the proposed metric (10) as shown in the following proposition.

Proposition1 Let $(\xi_{\mathbf{U}_1}, \xi_{\mathbf{U}_2}, \xi_{\mathbf{U}_3}, \xi_{\boldsymbol{\mathcal{G}}})$ and $(\eta_{\mathbf{U}_1}, \eta_{\mathbf{U}_2}, \eta_{\mathbf{U}_3}, \eta_{\boldsymbol{\mathcal{G}}})$ be tangent vectors to the quotient manifold (6) at $(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \boldsymbol{\mathcal{G}})$, and $(\xi_{\mathbf{U}_1\mathbf{O}_1}, \xi_{\mathbf{U}_2\mathbf{O}_2}, \xi_{\mathbf{U}_3\mathbf{O}_3}, \xi_{\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T})$ and $(\eta_{\mathbf{U}_1\mathbf{O}_1}, \eta_{\mathbf{U}_2\mathbf{O}_2}, \eta_{\mathbf{U}_3\mathbf{O}_3}, \eta_{\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T})$ be tangent vectors to the quotient manifold (6) at $(\mathbf{U}_1\mathbf{O}_1, \mathbf{U}_2\mathbf{O}_2, \mathbf{U}_3\mathbf{O}_3, \boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)$. The metric (10) is invariant along the equivalence class (5), i.e.,

$$\begin{aligned} & g_{(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})}((\xi_{\mathbf{U}_1}, \xi_{\mathbf{U}_2}, \xi_{\mathbf{U}_3}, \xi_{\boldsymbol{\mathcal{G}}}), (\eta_{\mathbf{U}_1}, \eta_{\mathbf{U}_2}, \eta_{\mathbf{U}_3}, \eta_{\boldsymbol{\mathcal{G}}})) \\ & \quad = g_{(\mathbf{U}_1\mathbf{O}_1,\mathbf{U}_2\mathbf{O}_2,\mathbf{U}_3\mathbf{O}_3,\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)} \\ & \qquad ((\xi_{\mathbf{U}_1\mathbf{O}_1}, \xi_{\mathbf{U}_2\mathbf{O}_2}, \xi_{\mathbf{U}_3\mathbf{O}_3}, \xi_{\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T}), \\ & \qquad\quad (\eta_{\mathbf{U}_1\mathbf{O}_1}, \eta_{\mathbf{U}_2\mathbf{O}_2}, \eta_{\mathbf{U}_3\mathbf{O}_3}, \eta_{\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T})). \end{aligned} \tag{11}$$

Consequently, the Riemannian submersion principle allows to derive the optimization-related ingredients systematically. Table 1 lists the required ingredients.

4. Numerical comparisons on tensor completion

In this section, we focus on the problem of low-rank tensor completion when the rank is a priori known. We tackle the problem efficiently by exploiting the new geometry of Tucker manifold developed in Section 3.

We propose a Riemannian nonlinear conjugate gradient algorithm for the tensor completion problem that scales well to large-scale instances. Specifically, we use the conjugate gradient implementation of Manopt [7] with the ingredients described in Table 1. The convergence analysis of this method follows from [4], [11], [12]. The total computational cost per iteration of our proposed algorithm is $O(|\Omega|r_1r_2r_3)$, where $|\Omega|$ is the number of known entries.

We show numerical comparisons of our proposed algorithm with state-of-the-art algorithms that include TOpt [13] and geomCG [1], for comparisons with Tucker decomposition based algorithms, and HaLRTC [14], Latent [15], and Hard [16] as nuclear norm minimization algorithms. All simulations are performed in Matlab on a 2.6 GHz Intel Core i7 machine with 16 GB RAM. For specific operations with tensor unfoldings, we use the `mex` interfaces that are provided in geomCG. For large-scale instances, our algorithm is only compared with geomCG as other algorithms cannot handle these instances.

Since the dimension of the space of a tensor $\in \mathbb{R}^{n_1\times n_2\times n_3}$ of rank $R = (r_1, r_2, r_3)$ is $\dim(\mathcal{M}/\sim) = \sum_{d=1}^{3}(n_dr_d - r_d^2) + r_1r_2r_3$, we randomly and uniformly select known entries based on a multiple of the dimension, called the *over-sampling* (OS) ratio, to create the training set Ω. Algorithms (and problem instances) are initialized randomly, as in [1], and are stopped when either the mean square error (MSE) on the training set Ω is below 10^{-12} or the number of iterations exceeds 250. We also evaluate the MSE on a test set Γ, which is different from Ω. Five runs are performed in each scenario.

Case 1 considers synthetic small-scale tensors of size $100\times100\times100$, $150\times150\times150$, and $200\times200\times200$ and rank $R = (10, 10, 10)$ are considered. OS is $\{10, 20, 30\}$. Figure 2(a) shows that the convergence behavior of our proposed algorithm is either competitive or faster than the others.

Case 2 considers large-scale tensors of size $3000\times3000\times3000$, $5000\times5000\times5000$, and $10000\times10000\times10000$ and ranks $R = (5, 5, 5)$ and $(10, 10, 10)$. OS is 10. Our proposed algorithm outperforms geomCG in Figure 2(b).

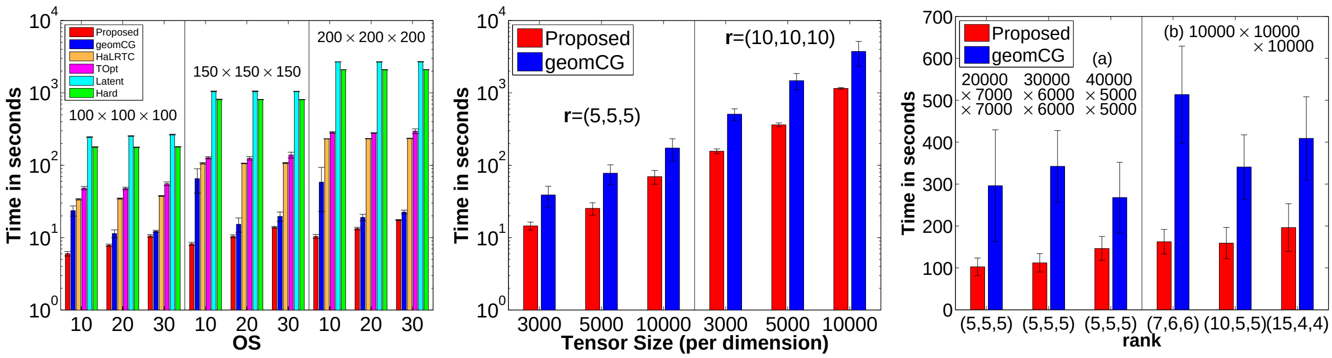
Case 3 considers instances where the dimensions and ranks along certain modes are different than others. Two cases are considered. Case (3.a) considers tensors size $20000\times7000\times7000$, $30000\times6000\times6000$, and $40000\times5000\times5000$ with rank $\mathbf{r} = (5, 5, 5)$. Case (3.b) considers a tensor of size $10000\times10000\times10000$ with ranks $(7, 6, 6)$, $(10, 5, 5)$, and $(15, 4, 4)$. In all the cases, the proposed algorithm converges faster than geomCG as in Figure 2(c).

5. Conclusion and future work

We proposed a geometry of Tucker manifold of tensor with a scaled Riemannian metric. The proposed metric exploits the symmetry structure of the Tucker decomposition and least-squares cost. The concrete matrix formulas for optimization were derived. Numerical comparisons on the tensor completion problem suggest that our proposed algorithm has a superior performance on different

表 1 Tucker manifold related optimization ingredients for (2) [9].

Matrix representation	$x = (\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \boldsymbol{\mathcal{G}})$
Computational space $\mathcal{M}$	$\mathrm{St}(r_1, n_1) \times \mathrm{St}(r_2, n_2) \times \mathrm{St}(r_3, n_3) \times \mathbb{R}^{r_1 \times r_2 \times r_3}$
Group action	$\{(\mathbf{U}_1\mathbf{O}_1, \mathbf{U}_2\mathbf{O}_2, \mathbf{U}_3\mathbf{O}_3, \boldsymbol{\mathcal{G}} \times_1 \mathbf{O}_1^T \times_2 \mathbf{O}_2^T \times_3 \mathbf{O}_3^T) : \mathbf{O}_d \in \mathcal{O}(r_d), \mathrm{for}\ d \in \{1,2,3\}\}$
Quotient space $\mathcal{M}/\sim$	$\mathrm{St}(r_1, n_1) \times \mathrm{St}(r_2, n_2) \times \mathrm{St}(r_3, n_3) \times \mathbb{R}^{r_1 \times r_2 \times r_3} / (\mathcal{O}(r_1) \times \mathcal{O}(r_2) \times \mathcal{O}(r_3))$
Ambient space	$\mathbb{R}^{n_1 \times r_1} \times \mathbb{R}^{n_2 \times r_2} \times \mathbb{R}^{n_3 \times r_3} \times \mathbb{R}^{r_1 \times r_2 \times r_3}$
Tangent vectors in $T_x\mathcal{M}$	$\{(\mathbf{Z}_{\mathbf{U}_1}, \mathbf{Z}_{\mathbf{U}_2}, \mathbf{Z}_{\mathbf{U}_3}, \mathbf{Z}_{\boldsymbol{\mathcal{G}}}) \in \mathbb{R}^{n_1 \times r_1} \times \mathbb{R}^{n_2 \times r_2} \times \mathbb{R}^{n_3 \times r_3} \times \mathbb{R}^{r_1 \times r_2 \times r_3}$ $: \mathbf{U}_d^T \mathbf{Z}_{\mathbf{U}_d} + \mathbf{Z}_{\mathbf{U}_d}^T \mathbf{U}_d = 0, \mathrm{for}\ d \in \{1,2,3\}\}$
Metric $g_x(\xi_x, \eta_x)$ for any $\xi_x, \eta_x \in T_x\mathcal{M}$	$\langle \xi_{\mathbf{U}_1}, \eta_{\mathbf{U}_1}(\mathbf{G}_1\mathbf{G}_1^T)\rangle + \langle \xi_{\mathbf{U}_2}, \eta_{\mathbf{U}_2}(\mathbf{G}_2\mathbf{G}_2^T)\rangle + \langle \xi_{\mathbf{U}_3}, \eta_{\mathbf{U}_3}(\mathbf{G}_3\mathbf{G}_3^T)\rangle + \langle \xi_{\boldsymbol{\mathcal{G}}}, \eta_{\boldsymbol{\mathcal{G}}}\rangle$
Vertical tangent vectors in $\mathcal{V}_x$	$\{(\mathbf{U}_1\mathbf{\Omega}_1, \mathbf{U}_2\mathbf{\Omega}_2, \mathbf{U}_3\mathbf{\Omega}_3, -(\boldsymbol{\mathcal{G}} \times_1 \mathbf{\Omega}_1 + \boldsymbol{\mathcal{G}} \times_2 \mathbf{\Omega}_2 + \boldsymbol{\mathcal{G}} \times_3 \mathbf{\Omega}_3)) : \mathbf{\Omega}_d \in \mathbb{R}^{r_d \times r_d}, \mathbf{\Omega}_d^T = -\mathbf{\Omega}_d$ $\mathrm{for}\ d \in \{1,2,3\}\}$
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 computed by solving Lyapunov equations.
$\Pi(\cdot)$ projects a tangent vector ξ onto $\mathcal{H}_x$	$(\xi_{\mathbf{U}_1} - \mathbf{U}_1\mathbf{\Omega}_1, \xi_{\mathbf{U}_2} - \mathbf{U}_2\mathbf{\Omega}_2, \xi_{\mathbf{U}_3} - \mathbf{U}_3\mathbf{\Omega}_3, \xi_{\boldsymbol{\mathcal{G}}} - (-(\boldsymbol{\mathcal{G}} \times_1 \mathbf{\Omega}_1 + \boldsymbol{\mathcal{G}} \times_2 \mathbf{\Omega}_2 + \boldsymbol{\mathcal{G}} \times_3 \mathbf{\Omega}_3)))$, $\mathbf{\Omega}_d$.
First order derivative of $f(x)$	$(\mathbf{S}_1(\mathbf{U}_3 \otimes \mathbf{U}_2)\mathbf{G}_1^T, \mathbf{S}_2(\mathbf{U}_3 \otimes \mathbf{U}_1)\mathbf{G}_2^T, \mathbf{S}_3(\mathbf{U}_2 \otimes \mathbf{U}_1)\mathbf{G}_3^T), \boldsymbol{\mathcal{S}} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \times_3 \mathbf{U}_3^T)$, where $\boldsymbol{\mathcal{S}} = \frac{2}{\lvert\Omega\rvert}(\boldsymbol{\mathcal{P}}_{\Omega}(\boldsymbol{\mathcal{G}} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3) - \boldsymbol{\mathcal{P}}_{\Omega}(\boldsymbol{\mathcal{X}}^{\star}))$.
Retraction $R_x(\xi_x)$	$(\mathrm{uf}(\mathbf{U}_1 + \xi_{\mathbf{U}_1}), \mathrm{uf}(\mathbf{U}_2 + \xi_{\mathbf{U}_2}), \mathrm{uf}(\mathbf{U}_3 + \xi_{\mathbf{U}_3}), \boldsymbol{\mathcal{G}} + \xi_{\boldsymbol{\mathcal{G}}})$
Horizontal lift of vector transport $\mathcal{T}_{\eta_{[x]}}\xi_{[x]}$	$\Pi_{R_x(\eta_x)}(\Psi_{R_x(\eta_x)}(\xi_x))$



(a) **Case 1:** small-scale tensors. (b) **Case 2:** large-scale tensors. (c) **Case 3:** asymmetric tensors.

図 2 Experiments on synthetic datasets [9].

benchmarks. As future work, we would experiment the developed tool on other regression problems.

参考文献

[1] Kressner, D., Steinlechner, M. and Vandereycken, B.: Low-rank tensor completion by Riemannian optimization, *BIT Numer. Math.*, Vol. 54, No. 2, pp. 447–468 (2014).

[2] Wimalawarne, K., Tomioka, R. and Sugiyama, M.: Theoretical and Experimental Analyses of Tensor-Based Regression and Classification, *Neural Comput.*, Vol. 28, No. 4, pp. 686–715. (2016).

[3] Grasedyck, L., Kressner, D. and Toble, C.: A literature survey of low-rank tensor approximation techniques, *GAMM-Mitteilungen*, Vol. 36, No. 1, pp. 53–78 (2013).

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

[5] Kolda, T. G. and Bader, B. W.: Tensor Decompositions and Applications, *SIAM Rev.*, Vol. 51, No. 3, pp. 455–500 (2009).

[6] Mishra, B. and Sepulchre, R.: Riemannian preconditioning, *SIAM J. Optim.*, Vol. 26, No. 1, pp. 635–660 (2016).

[7] Boumal, N., Mishra, B., Absil, P.-A. and Sepulchre, R.: Manopt: a Matlab toolbox for optimization on manifolds, *J. Mach. Learn. Res.*, Vol. 15, No. 1, pp. 1455–1459 (2014).

[8] Koch, O. and Lubich, C.: Dynamical Tensor Approximation, *SIAM. J. Matrix Anal.*, Vol. 31, No. 5, pp. 2360–2375 (2010).

[9] Kasai, H. and Mishra, B.: Low-rank tensor completion: a Riemannian manifold preconditioning approach, *ICML* (2016).

[10] Lee, J. M.: *Introduction to smooth manifolds*, Graduate Texts in Mathematics, Vol. 218, Springer-Verlag, New York, second edition (2003).

[11] Sato, H. and Iwai, T.: A new, globally convergent Riemannian conjugate gradient method, *Optimization*, Vol. 64, No. 4, pp. 1011–1031 (2015).

[12] Ring, W. and Wirth, B.: Optimization Methods on Riemannian Manifolds and Their Application to Shape Space, *SIAM J. Optim.*, Vol. 22, No. 2, pp. 596–627 (2012).

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

[14] Liu, J., Musialski, P., Wonka, P. and Ye, J.: Tensor Completion for Estimating Missing Values in Visual Data, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 35, No. 1, pp. 208–220 (2013).

[15] Tomioka, R., Hayashi, K. and Kashima, H.: Estimation of low-rank tensors via convex optimization, Technical report, arXiv preprint arXiv:1010.0789 (2011).

[16] Signoretto, M., Dinh, Q. T., Lathauwer, L. D. and Suykens, J. A. K.: Learning with tensors: a framework based on convex optimization and spectral regularization, *Mach. Learn.*, Vol. 94, No. 3, pp. 303–351 (2014).