ON THE RELAXATION OF ORTHOGONAL TENSOR RANK AND ITS NONCONVEX RIEMANNIAN OPTIMIZATION FOR TENSOR COMPLETION

Keisuke Ozawa

DENSO IT Laboratory

ABSTRACT

Natural extension of matrix rank has attracted interest toward a parsimonious representation and completion of a tensor with partial observation. In this paper, we focus on orthogonal tensor rank and discuss its nonconvex relaxation and minimization. Accordingly, we present a completion algorithm using the proximal alternating direction method of multipliers for three-way tensors, wherein we solve a minimization problem on the orthogonal group using the Riemannian subgradient descent. We also analyze the global convergence of the proposed algorithm. In a simulation experiment, we show that our algorithm could extract the parsimonious structure of a tensor with partial observation. We also demonstrate, against both synthetic and realistic data, a superior completion performance of our proposed algorithm to some recent methods.

Index Terms— Tensor completion, Orthogonal tensor rank, Nonconvex optimization, Riemannian optimization

1. INTRODUCTION

Tensor completion aims to estimate missing entries in corrupted multidimensional data. There, a tensor is assumed to be of a low rank: it should be composed of a small number of factors [1, 2]. A long-standing issue is that, unlike matrix, tensor rank is not uniquely defined. Inheriting the properties of matrix rank is of typical interest, since the extension could allow us to take over well-established matrix completion algorithms and to better describe certain categories of data.

The orthogonal tensor rank [3, 4] is notably a natural extension of matrix rank: it reduces to matrix rank in two dimensions; it is contravariant [2, 5] and has no entanglement over dimensions, which inevitably comes in when unfolding a tensor into matrices [6, 7]; its factorization reduces to matrix singular value decomposition (SVD) with a straightforward extension to higher-order factorizations (Definition 2.1). The orthogonal tensor rank is related to Tucker rank [8] when fixing the number of factors. However, optimizing based on Tucker rank is difficult in general when we do not know the number of factors. In contrast, the orthogonal tensor rank inherits well the properties of matrix rank, with which we try to break down the difficulty to relax the rank (Definition 3.1) in order to optimize without fixing the rank likewise for a matrix. A similar definition is partly used in [9], but it is defined with a fixed number of factors, and the purpose is to estimate CP or Tucker rank for succeeding completion. The orthogonal tensor rank is also a generalization of the transformed SVD (tSVD) [10, 11, 12] (Section 3.1), which has gained interest due to its high performance. However, it does not facilitate the parsimonious structure of data over all dimensions. A recent study [13] proposed learning the parsimonious structure of data, but it requires an over-complete dictionary. In this situation, developing a completion algorithm based on the orthogonal tensor rank and simultaneously seeking for a parsimonious representation of data still remains an issue.

In this study, we first consider the well-definedness of the orthogonal tensor rank and its relaxation to develop a completion algorithm, especially for three-way tensors. It requires the optimization of a nonconvex function on the orthogonal group and the tensor space. We tackle this using the proximal alternating direction method of multipliers (PADMM) [14], wherein a subproblem on the orthogonal group is solved with the Riemannian subgradient descent [15, 16]. We also discuss its global convergence based on our updates and a recent theoretical result [17]. Finally, using an empirical initialization, we experimentally show that our algorithm could find the parsimonious structure of synthetic data. We also compare the completion accuracy with some recent methods against synthetic data with some parsimonious structures and two realistic data with unknown structures.

2. BACKGROUND

Let the tensor product be $(\mathbf{a} \otimes \mathbf{b})_{ij} = a_i b_j$, and for a general $n_1 \times \cdots \times n_p$ tensor with a set $I := \{i_1, \cdots, i_p\}$, $\mathbf{F}_I = \mathbf{f}^{(1)}_{i_1} \otimes \cdots \otimes \mathbf{f}^{(p)}_{i_p} \in V^{(1)} \otimes \cdots \otimes V^{(p)}$ with $\{\mathbf{f}^{(d)}_i\}_{1 \leq i \leq \dim V^{(d)}}$, which spans an orthonormal basis set of the real vector space $V^{(d)}$ for $d = 1, \cdots, p$. According to Franc [3] and Kolda [4]:

Definition 2.1. The *orthogonal tensor rank* of $\mathcal{A}$ is

$$\operatorname{rank}(\mathcal{A}) = \min\# \left\{ \sigma^I \mid \sigma^I > 0, \mathcal{A} = \sum_I \sigma^I \mathbf{F}_I \right\}.$$

We show the well-definedness of the rank in a partly constructive manner, which we use to develop our algorithm. The well-definedness is not given in [4] and only the uniqueness w.r.t. least square fitting is discussed up to orthogonality [18].

Proposition 2.1. The orthogonal tensor rank of a $p(\geq 3)$-way tensor $\mathcal{A}$ is attainable.

Proof. Consider $p = 3$. Given $\{\mathbf{r}_k^{(3)}\}_{1\leq k\leq n_3}$ orthonormal, $\mathcal{A} = \sum_{sk} \sigma_{sk}\mathbf{u}_s^{(1,k)} \otimes \mathbf{v}_s^{(2,k)} \otimes \mathbf{r}_k^{(3)}$ for unique $\sigma_{sk} > 0$ and some orthonormal bases $\{\mathbf{u}_s^{(1,k)}\}_{1\leq s\leq M_k}$ and $\{\mathbf{v}_s^{(2,k)}\}_{1\leq s\leq M_k}$, where M_k is the rank of the k-th matrix along the 1^{st} and 2^{nd} dimensions. Since the matrix rank is lower-semicontinuous and the orthogonal group is compact, the infimum is attainable. For $p > 3$, since the direct product of compact groups is also compact, we inductively conclude the statement. □

3. RELAXATION TOWARD OPTIMIZATION

We relax the rank so that it can be optimized and also naturally reduces to the matrix nuclear norm for two-dimensional cases. It is beyond the scope of this paper whether the relaxation has any condition to minimize the rank exactly.

Definition 3.1. The *orthogonal tensor nuclear norm (OTNN)* of $\mathcal{A}$ is, using the factorization in Definition 2.1, defined as

$$\mathrm{OTNN}(\mathcal{A}) = \sum_I \sigma^I.$$

Although this is also well-defined, it is still difficult to obtain the minimum due to the intricate interaction between variables or bases. To resolve this, we introduce a function over variables to be optimized directly. In this study, we consider a three-way tensor for simplicity, but it can be inductively defined for higher-order tensors. The function is

$$\tau(\mathcal{X}, \mathbf{R}) = \sum_{sk} \sigma_{sk},\ \mathcal{X} = \sum_{sk} \sigma_{sk}\mathbf{u}_s^{(1,k)} \otimes \mathbf{v}_s^{(2,k)} \otimes \mathbf{r}_k^{(3)}, \quad (1)$$

where $R_{:k} = \mathbf{r}_k^{(3)}$ and this matrix $\mathbf{R} \in \mathrm{O}(n_3)$ from the orthogonal group spans a complete orthonormal basis set; $\sigma_{sk} \geq \sigma_{s'k} (s < s')$ holds, but we do not assume the order by k, since it does not affect the algorithm. It can be seen from the factorization in the proof of Proposition 2.1 that the value of τ is uniquely determined for a given $\mathbf{R}$.

We have some properties of OTNN and τ. First, although it is difficult to compute OTNN directly unlike the matrix nuclear norm, τ reduces to the sum of matrix nuclear norm over the index k given a fixed $\mathbf{R}$. This is inductively the same for higher-order tensors, but the implementation will be intricate.

Second, this is nonconvex, but there is some $\mathbf{R}^*$ that achieves $\tau(\mathcal{X}, \mathbf{R}^*) = \mathrm{OTNN}(\mathcal{X})$. Then, we mean to minimize τ over possible $\mathcal{X}$ and $\mathbf{R}$ to reduce the value of τ as much as possible within a nonconvex approach (Section 4).

Third, for three-way tensors, the factorization in OTNN is also a generalization of tSVD if $\mathbf{R}$ is fixed with the discrete Fourier transform (DFT) matrix. Thus, OTNN is a better generalization of the matrix nuclear norm and, at the expense of convexity, improves the degree of freedom for factorization compared with tSVD.

Finally, although it is difficult to see the global landscape of OTNN over $\{\mathbf{r}_k^{(3)}\}$, we have a permutation symmetry among bases when not fixing the order by k. Therefore, when seeking for a single basis, it is sufficient if the most nearest basis is rotated. Seeking for a subspace is likewise. This intuitively suggests that although minimizing OTNN is nonconvex and requires an elaborate initialization, there might exist multiple equivalent initializers coming from this symmetry.

4. TENSOR COMPLETION ALGORITHM

We consider a completion problem that minimizes the function τ instead of OTNN against a three-way tensor of $\mathbb{R}^N := \mathbb{R}^{n_1\times n_2\times n_3}$. We solve a nonconvex optimization problem

$$\min_{\mathcal{X}\in\mathbb{R}^N, \mathbf{R}\in\mathrm{O}(n_3)} \tau(\mathcal{X}, \mathbf{R}) + \iota_\Omega(\mathcal{X}), \quad (2)$$

where Ω has observed entries and ι_Ω is the indicator function on it. We seek for a structure of $\mathcal{X}$ that is as parsimonious as possible in terms of the OTNN. We solve Equation 2 with PADMM [14], upon which we discuss the global convergence by using a recent theoretical result in [17] (Section 4.2).

4.1. PADMM updates

PADMM [14] iteratively updates the Lagrangian with additional proximal terms and has been widely used in nonconvex optimization problems. The augmented Lagrangian is

$$\mathcal{L}(\mathcal{X}, \mathcal{Y}, \mathbf{R}; \Phi) = \tau(\mathcal{X}, \mathbf{R}) + \iota_\Omega(\mathcal{Y}) + \frac{\rho}{2}\left\|\mathcal{X} - \mathcal{Y} + \frac{\Phi}{\rho}\right\|_F^2, \quad (3)$$

where $\|\mathcal{A}\|_F^2 = \sum_{ijk} |\mathcal{A}_{ijk}|^2$ is the square of the Frobenius norms of the tensor $\mathcal{A}$, $\mathcal{Y}$ is the dual variable of $\mathcal{X}$, and Φ is the Lagrange multiplier. We denote the proximal terms between points x and y on the metric space M as $\Delta_M(x, y)$. Here, we use the induced metric (Frobenius norm) from the Euclidean space for tensors. We also use the induced metric from the Euclidean space for the orthogonal group $\mathrm{O}(n_3)$. Then, $\mathrm{O}(n_3)$ and the special orthogonal group $\mathrm{SO}(n_3)$ as its subgroup are endowed with the Riemannian (sub)gradient and the retraction on them, which will be used to update $\mathbf{R}$.

1) The update of $\mathcal{X}$ follows

$$\mathcal{X}^{(t+1)} = \operatorname*{arg\,min}_{\mathcal{X}\in\mathbb{R}^N} \tau(\mathcal{X}, \mathbf{R}^{(t)}) + \frac{\rho}{2}\left\|\mathcal{X} - \mathcal{Y}^{(t)} + \frac{\Phi^{(t)}}{\rho}\right\|_F^2 + \rho_{\mathcal{X}}\Delta_{\mathbb{R}^N}(\mathcal{X}, \mathcal{X}^{(t)}), \quad (4)$$

which is solved similarly as in [10, 11, 12]: with $\rho' = \rho + \rho_{\mathcal{X}}$; for the k-th frontal slice matrix of (with subscript $:: k$, where $:$ means all the entries for i's and j's of $\mathcal{X}_{ijk}$ appear in that dimension),

$$\mathcal{X}_{::k}^{(t+1)} = \mathcal{D}_{\rho'}\left[\frac{1}{\rho'}\left\{\rho\left(\mathcal{Y}_{::k}^{(t)} - \frac{\Phi_{::k}^{(t)}}{\rho}\right) + \rho_{\mathcal{X}}\mathcal{X}_{::k}^{(t)}\right\}\right]. \quad (5)$$

Here, we denote the singular value thresholding operator [19] as $\mathcal{D}_\mu(\mathbf{A}) = \mathbf{U}(\mathrm{diag}(\{\mathcal{S}_\tau(\sigma_i)\}_{1\le i\le r}))\mathbf{V}^\top$ with the reduced SVD as $\mathbf{A} = \mathbf{U}(\mathrm{diag}(\{\sigma_i\})_{1\le i\le r})\mathbf{V}^\top$. $\mathcal{S}_\tau(x) = \mathrm{sgn}(x)(|x| - \tau)_+$ is the soft-thresholding operator, where $y_+ = \max(y, 0)$, and it operates element-wise for a matrix.

2) The update of $\mathbf{R}$ follows

$$\mathbf{R}^{(t+1)} = \underset{\mathbf{R}\in\mathcal{N}(\mathbf{R}^{(t)})}{\arg\min}\ \tau(\mathcal{X}^{(t+1)}, \mathbf{R}) + \rho_{\mathbf{R}}\Delta_{\mathrm{O}(n_3)}(\mathbf{R}, \mathbf{R}^{(t)}), \quad (6)$$

where $\mathcal{N}(\mathbf{R}^{(t)})$ is a path-connected region of $\mathrm{O}(n_3)$ that includes $\mathbf{R}^{(t)}$ and one local minimum. We here restrict the path on $\mathrm{SO}(n_3)$. Although the region $\mathcal{N}(\mathbf{R}^{(t)})$ is not defined explicitly, we practically mean to achieve the local minimum via a line search algorithm with, e.g., the Armijo rule. Note that the search can only find a local minimum but not the global minimum even on the path-connected component. The cost function $f(\mathbf{R}) := \tau(\mathcal{X}^{(t+1)}, \mathbf{R}) + \frac{\rho_{\mathbf{R}}}{2}\left\|\mathbf{R} - \mathbf{R}^{(t)}\right\|_F^2$ is defined on $\mathrm{SO}(n_3)$ but can be extended to a function $\tilde{f}$ on the Euclid space $\mathbb{R}^{n_3\times n_3}$ with satisfying $\tilde{f}(\mathbf{R})|_{\mathrm{SO}(n_3)} = f(\mathbf{R})$. Using the subdifferential of the matrix nuclear norm [20], as a Euclidean subgradient of $\tilde{f}$ w.r.t. $\mathbf{R}$, we obtain

$$\left(\nabla\tilde{f}\right)_{ab} = \sum_{i,j,s}(\mathbf{u}_s^{(1,a)})_i(\mathbf{v}_s^{(2,a)})_j\mathcal{X}_{ijb} + \rho_{\mathbf{R}}(R_{ab} - R_{ab}^{(t)}), \quad (7)$$

where $\mathbf{u}$ and $\mathbf{v}$ are the left and right singular vectors, respectively, computed as $\mathcal{X} = \sum_{sa}\sigma_{sa}\mathbf{u}_s^{(1,a)}\otimes\mathbf{v}_s^{(2,a)}\otimes\mathbf{r}_a^{(3)}$ presented constructively in the proof of Proposition 2.1. Then, the (rescaled) Riemannian subgradient on the special orthogonal group (or the orthogonal group) is [21]

$$\mathrm{grad}f = \nabla\tilde{f} - \mathbf{R}(\nabla\tilde{f})^\top\mathbf{R}. \quad (8)$$

The line search on a Riemannian manifold requires a projection from the tangent bundle onto the base space. Instead of the exponential map, we use a computational friendly retraction that also satisfies the constraint on the manifold. From the induced Euclidean metric, we can use a retraction [21][1]

$$\mathcal{R}_{\mathbf{R}}(\boldsymbol{\xi}) = \mathrm{qf}(\mathbf{R} + \boldsymbol{\xi}), \quad (9)$$

where $\mathrm{qf}(\mathbf{X})$ is the orthogonal component from the QR decomposition of $\mathbf{X}$ with a positive definite counterpart, which are uniquely determined. In general, the Riemannian subgradient descent was introduced in [15, 16]. Although the subgradient is not necessarily a decreasing direction, our line search guarantees decreasing the cost or being stationary. The details for updating $\mathbf{R}$ is shown in Algorithm 2.

3) The update of $\mathcal{Y}$ follows

$$\mathcal{Y}^{(t+1)} = \underset{\mathcal{Y}\in\mathbb{R}^N}{\arg\min}\,\iota_\Omega(\mathcal{Y}) + \frac{\rho^{(t)}}{2}\left\|\mathcal{Y} - \left(\mathcal{X}^{(t+1)} + \frac{\Phi^{(t)}}{\rho^{(t)}}\right)\right\|_F^2$$
$$+ \rho_{\mathcal{Y}}\Delta_{\mathbb{R}^N}(\mathcal{Y}, \mathcal{Y}^{(t)}) = \mathcal{P}_{\Omega^{\mathrm{c}}}\left(\mathcal{Y}'^{(t)}\right) + \mathcal{P}_\Omega\left(\mathcal{M}\right), \quad (10)$$

[1]We assume for the summation the identification between the embedded space and the tangent space, both of linear.

where $\mathcal{Y}'^{(t)} = \left\{\rho\left(\mathcal{X}^{(t+1)} + \frac{\Phi^{(t)}}{\rho}\right) + \rho_{\mathcal{Y}}\mathcal{Y}^{(t)}\right\}/(\rho + \rho_{\mathcal{Y}})$, and $\mathcal{P}_\Omega(\mathcal{A})$ projects the tensor $\mathcal{A}$ onto $\Omega \subset \mathrm{dom}(\mathcal{A})$ with $[\mathcal{P}_\Omega(\mathcal{A})]_{ijk} = \mathcal{A}_{ijk}$ if $(i, j, k) \in \Omega$ and 0 otherwise; Ω^{c} is the complement of Ω.

The procedures in Algorithm 1 and 2 are as follows[2]:

Algorithm 1 Tensor completion via OTNN minimization

Input: Observed entries Ω in $\mathcal{X} \in \mathbb{R}^N$
Output: Completed tensor $\mathcal{X}^* \in \mathbb{R}^N$ and $\mathbf{R}^* \in \mathrm{SO}(n_3)$
Initialisation : $\mathcal{X} = \mathcal{M}$;
Parameters : tol, max iteration, ρ, $\rho_{\mathcal{X}(,\mathcal{Y},\mathbf{R})}$
while below max iteration or max diff $>$ tol **do**
  Update $\mathcal{X}$ using Equation 5
  Update $\mathbf{R}$ using Algorithm 2
  Update $\mathcal{Y}$ using Equation 10
  Update as $\Phi \leftarrow \Phi + \rho(\mathcal{X} - \mathcal{Y})$.
  Continue updates 2-5
end while=0

Algorithm 2 Line search on $\mathrm{SO}(n_3)$ with Armijo rule

Input: $\mathcal{X}^{(t+1)} \in \mathbb{R}^N$ and $\mathbf{R}^{(t)} \in \mathrm{SO}(n_3)$
Output: $\mathbf{R}^{(t+1)} \in \mathrm{SO}(n_3)$
Initialisation : $c, \alpha, \xi > 0$, $\alpha_{\min} > 0$
while $\alpha > \alpha_{\min}$ **do**
  Compute grad f using Equations 7 and 8.
  Compute $\mathbf{R}^{(t)\prime} = \mathcal{R}_{\mathbf{R}^{(t)}}(\mathbf{R}^{(t)} - \alpha\mathrm{grad}f)$
  if $f(\mathbf{R}^{(t)\prime}) \le f(\mathbf{R}^{(t)}) - c\alpha\|\mathrm{grad}f\|^2$ **then**
    $\mathbf{R}^{(t+1)} \leftarrow \mathbf{R}^{(t)\prime}$ and break
  else
    $\alpha \leftarrow \xi\alpha$
  end if
end while=0

4.2. Convergence analysis

We discuss the convergence of Algorithm 1 based on a theoretical result in [17]. We provide the outline below.

$\mathbb{I}$ and $\mathbb{O}$ are the identity and zero matrices, respectively. Let $A_1 = (\mathbb{I}, \mathbb{O})$, $A_2 = \mathbb{I}$, $\mathbf{x} = (\mathrm{vec}(\mathcal{X}); \mathrm{vec}(\mathbf{R}))$, $\mathbf{y} = \mathrm{vec}(\mathcal{Y})$, $f_1(\mathbf{x}) = \tau(\mathcal{X}, \mathbf{R})$ and $f_2(\mathbf{y}) = \iota_\Omega(\mathbf{y})$; f_1 and f_2 are lower semi-continuous, and our Lagrangian satisfies the Kurdyka-Lojasiewicz property. The proximal terms for $\mathcal{X}$, $\mathcal{Y}$, and $\mathbf{R}$ are all positive definite. The updates of $\mathcal{X}$ and $\mathcal{Y}$ are rigorous and satisfy the Fermat's rule. If the updates of $\mathbf{R}$ always achieve the local minimum or a stationary point for $t > {}^\exists T$, the sequence $\left\{\mathcal{X}^{(t)}, \mathcal{Y}^{(t)}, \mathbf{R}^{(t)}, \Phi^{(t)}\right\}_{1\le t<\infty}$ generated by Algorithm 1 converges to a stationary point, $\mathbf{x}$ being nonseparable though, similarly to showing Theorem 1 in [17].

If we use the Wolfe rule for Algorithm 2, it satisfies the above condition. Still, in practice, Algorithm 1 and 2 using the Armijo rule converged in all the experiments.

[2]The complexities per iteration are $\mathcal{O}(n_1n_2n_3\min(n_1, n_2))$ for (5), $\mathcal{O}(n_1n_2n_3)$ for (10), $\mathcal{O}(n_1n_2n_3)$ to update the Lagrange multiplier, each in Algorithm 1; $\mathcal{O}(n_1n_2n_3^2\min(n_1, n_2) + n_3^2)$ for (7), $\mathcal{O}(n_3^3 + n_3^2)$ for (8), and $\mathcal{O}(n_3^3)$ for (9), each in Algorithm 2, respectively.

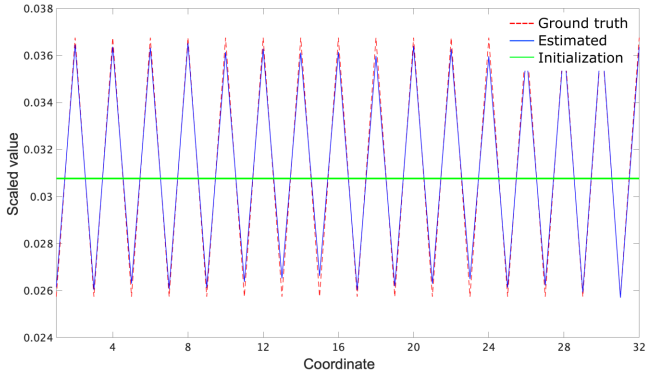

Fig. 1. Estimated basis (adjusted scale) and the initialization (adjusted offset for visualization) against the ground truth.

5. EXPERIMENT

We first conducted a simulation to see whether our algorithm finds the parsimonious structure of a tensor. Then, using both synthetic and realistic data, we compare its completion accuracy against those of recent methods: TMac [6], TRLRF [7], TREL-Tucker(TRELT) and TREL-CP(TRELC) [9], and tSVD [10, 11, 12]. We implemented tSVD in the same PADMM framework as ours. Recent studies [22, 12] suggest that methods using a discrete cosine (DC) transform outperform those using DFT. Based on this fact, we refer to our implementation as DCtSVD. Also, the DCT matrix was best for our initialization of $\mathbf{R}$; therefore we set $\mathbf{R}^{(0)}$ to a DCT matrix. We used authors' implementations for the others and set authors' recommended rank for TMac: three-way TR rank varying over $(2, 3, 4, 5, 6, 7, 8, 9, 10)$ and then tuned $\lambda \in (10^{-2}, 10^{-1}, 1, 10^{1}, 10^{2})$ for TRLRF, and $\lambda \in (10^{-1}, 1, 10^{1}, 10^{2}, 10^{3})$ for TRELT and TRELC; then we list the best results. Note that such parameter tuning related to rank is less practical. In contrast, our method only requires parameters for PADMM and the Armijo rule, and is easily adjusted. Throughout the experiments, we fixed them to $\rho = \rho_{\mathcal{X}} = \rho_{\mathcal{Y}} = \rho_{\Phi} = 10^{-1}$, $c = \alpha = 10^{-2}$, and $\xi = 1/2$. The other parameters are $\mathrm{tol} = 10^{-5}$ and 1000 of max iterations. We show the averages and the standard deviations of relative squared errors (RSEs), $\|\mathcal{X}_0 - \mathcal{X}^*\|_F^2 / \|\mathcal{X}_0\|_F^2$ against the ground truth tensor $\mathcal{X}_0$, for five trials.

5.1. Synthetic data

First, to see the behavior of our algorithm, we consider a tensor $\mathcal{X} = \sum_{s=1,\dots 8, k=1} \mathbf{f}_s^{(1)} \otimes \mathbf{f}_k^{(2)} \otimes \mathbf{r}_k^{(3)} \in \mathbb{R}^{32\times 32\times 32}$ with $\mathbf{f}$'s being random (not orthonormal) and $\mathbf{r}_1^{(3)}$ being of all one (the initialization) multiplied by $\mathrm{diag}\,(\mathbf{R}_{2\times 2}(\theta_1), \cdots, \mathbf{R}_{2\times 2}(\theta_{32}))$ with $\theta_{2m+1} = 10$ degree $(m = 0, \cdots, 15)$, that is, rotated over the axes in the 32 dimensional space. We have 0.5 of the sampling rate (SR). Fig. 1 shows the estimated basis $\mathbf{r}_1^{(3)}$ after 100 iterations against the ground truth. This suggests that our algorithm could find nearly the parsimonious structure.

Table 1. RSEs against synthetic data ($\mathbf{1}^{\mathrm{st}}$ and $\underline{2}^{\mathrm{nd}}$ bests).

	1-factor ($\times 10^{-3}$)		2-factors ($\times 10^{-2}$)		4-factors ($\times 10^{-2}$)	
SR	0.1	0.2	0.1	0.2	0.1	0.2
TMac	310±40	98±2.0	**17**±2.8	13±2.5	**19**±3.7	9.8±0.2
TRLRF	7.7±0.0	7.3±0.0	27±0.0	24±0.0	93±0.0	94±0.0
TRELT	250±12	54±10	38±1.1	18±0.8	31±0.9	16±0.6
TRELC	<u>3.6</u>±0.2	<u>1.6</u>±0.06	29±0.2	25±0.05	23±0.06	22±0.04
DCtSVD	230±9.0	19±8.0	<u>23</u>±0.6	<u>6.9</u>±0.3	20±0.4	<u>8.7</u>±0.2
OTNN	**1.4**±0.5	**0.65**±0.08	**15**±0.7	**4.1**±0.6	**18**±0.4	**8.0**±0.4

Table 2. RSEs against realistic data ($\mathbf{1}^{\mathrm{st}}$ and $\underline{2}^{\mathrm{nd}}$ bests).

	Flow Injection ($\times 10^{-2}$)			Chromatography ($\times 10^{-1}$)		
SR	0.1	0.2	0.4	0.2	0.4	0.6
TMac	24±1.0	9.9±0.5	3.3±0.3	5.3±0.0	4.1±0.0	1.6±0.5
TRLRF	**10**±2.5	3.1±0.3	1.2±0.1	5.6±0.0	5.0±0.0	5.2±0.0
TRELT	<u>14</u>±1.0	3.8±0.3	1.8±0.0	4.1±0.3	2.9±0.4	1.1±0.0
TRELC	45±0.8	42±0.1	40±0.4	4.6±0.2	3.8±0.0	3.4±0.0
DCtSVD	<u>14</u>±0.3	<u>1.9</u>±0.2	<u>0.26</u>±0.00	<u>3.8</u>±0.2	<u>1.7</u>±0.1	<u>0.75</u>±0.04
OTNN	**8.7**±0.2	**1.1**±0.0	**0.21**±0.00	**2.4**±0.2	**1.1**±0.0	**0.69**±0.06

Next, we compared the completion accuracy using random factors with varying the number of factors within $(1, 2, 4)$. We do not assume orthogonality to make the situation difficult for OTNN to fit. We tested with an SR of 0.1 and 0.2. In Table 1, our method has better results: TMac is comparable in some cases, DCtSVD the second, and TRELT partly becomes the second, considering the standard deviations. The results also suggest that the proposed method performs better for lower-rank tensors. This is intuitively understood that the best orthogonal bases could be found easily when the data spans across a subspace of a lower dimension.

5.2. Realistic data

We also compared the completion accuracy of the methods on the Flow Injection [23] and Chromatography [24] data. The dimensions are $(100 \times 89 \times 12)$ and $(20 \times 39 \times 15)$, respectively, where the final dimension is for the mixed components. So, we here our algorithm seek for a parsimonious factorization for components. We tested on the first fraction among 28 fractions from the Chromatography data. Since a small number of elements have no value due to saturation, we set them as unobserved and in evaluation. In Table 2, our method has better completion accuracy; TRLRF is comparable in one case, DCtSVD the second, and TRELT was partly the second.

6. CONCLUSION

In this study, we considered a tensor completion problem based on the orthogonal tensor rank. Against three-way tensors, we developed a nonconvex Riemannian optimization algorithm using PADMM and analyzed its global convergence. Experimental results suggest that the proposed algorithm is able to find the parsimonious structure of data, and performs overall better or comparable compared to recent methods. Possible future directions are to discuss better initialization and higher-order implementations.

7. REFERENCES

[1] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. PHAN, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145–163, 2015.

[2] P. Comon, "Tensors : A brief introduction," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 44–53, 2014.

[3] A. Franc, "Etude algébrique des multitableaux : apports de l'algèbre tensorielle," *These de Doctorat*, 1992.

[4] T. G. Kolda, "Orthogonal tensor decompositions," *SIAM Journal on Matrix Analysis and Applications*, vol. 23, no. 1, 2003.

[5] F. L. Hitchcock, "The expression of a tensor or a polyadic as a sum of products," *Journal of Mathematics and Physics*, vol. 6, no. 1, pp. 164–189, 1927.

[6] Y. Xu, R. Hao, W. Yin, and Z. Su, "Parallel matrix factorization for low-rank tensor completion," *Inverse Problems and Imaging*, vol. 9, no. 2, pp. 601–624, 2015.

[7] Y. Longhao, L. Chao, M. Danilo, C. Jianting, and Z. Qibin, "Tensor ring decomposition with rank minimization on latent space: An efficient approach for tensor completion," *AAAI*, 2018.

[8] L. R. Tucker, "Implications of factor analysis of three-way matrices for measurement of change," *In Problems in measuring change*, vol. 15, 1963.

[9] Q. Shi, H. Lu, and Y. MingCheung, "Tensor rank estimation and completion via cp-based nuclear norm," *ACM Conference on Information and Knowledge Management*, vol. 11, 2017.

[10] M. E. Kilmer, C. D. Martin, and L. Perrone, "A third-order generalization of the matrix svd as a product of third-order tensors," *Tech. Report Technical Report TR-2008-4, Tufts University, Department of Computer Science*, 2008.

[11] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel methods for multilinear data completion and de-noising based on tensor-svd," *IEEE Conference on Computer Vision and Pattern Recognition*, p. 3842–3849, 2014.

[12] C. Lu, X. Peng, and Y. Wei, "Low-rank tensor completion with a new tensor nuclear norm induced by invertible linear transforms," *IEEE Conference on Computer Vision and Pattern Recognition*, p. 5996–6004, 2019.

[13] T. X. Jiang, X. L. Zhao, H. Zhang, and M. Ng, "Dictionary learning with low-rank coding coefficients for tensor completion," *arXiv:2009.12507*, 2020.

[14] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-Lojasiewicz inequality," *Mathematics of Operations Research*, vol. 35, no. 2, pp. 438–457, 2010.

[15] G. Dirr and H. Lageman, "Nonsmooth riemannian optimization with applications to sphere packing and grasping," *Lecture Notes in Control and Information Sciences*, vol. 366, pp. 29–45, 2006.

[16] P. B. Borckmans, S. E. Selvan, N.Boumal, and P. A. Absil, "A riemannian subgradient algorithm for economic dispatch with valve-point effect," *Journal of Computational and Applied Mathematics*, vol. 255, pp. 848–866, 2013.

[17] M. Yashtini, "Multi-block nonconvex nonsmooth proximal admm: Convergence and rates under kurdyka-lojasiewicz property," *Journal of Optimization Theory and Application*, 2021.

[18] T. Zhang and G. H. Golub, "Rank-one approximationn to high order tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 23, no. 2, pp. 534–550, 2001.

[19] J. F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[20] G. A. Watson, "Characterization of the subdifferential of some matrix norms," *Linear Algebra and its Applications*, vol. 170, pp. 33–45, 1992.

[21] P. A. Absil, R. Mahony, and R. Sepulchre, "Optimization algorithms on matrix manifolds," *Princeton University Press*, 2008.

[22] E. Kernfeld, M. Kilmer, and S. Aeron, "Tensor–tensor products with invertible linear transforms," *Linear Algebra and its Applications*, vol. 485, pp. 545–570, 2015.

[23] L. Nørgaard and C. Ridder, "Rank annihilation factor analysis applied to flow injection analysis with photodiode-array detection," *Chemometrics and Intelligent Laboratory Systems*, vol. 23, no. 1, pp. 107–114, 1994.

[24] R. Bro, C. A. Andersson, and H. A. L. Kiers, "Parafac2-part ii. modeling chromatographic data with retention time shifts," *Journal of Chemometrics*, vol. 13, no. 3-4, pp. 295–309, 1999.