Random Projections for Low Multilinear Rank Tensors

Carmeliza Navasca and Deonnia N. Pompey

Abstract We proposed two randomized tensor algorithms for reducing multilinear ranks in the Tucker format. The basis of these randomized algorithms is from the randomized SVD of Halko, Martinsson and Tropp [9]. Here we provide randomized versions of the higher order SVD and higher order orthogonal iteration. Moreover, we provide a sharper probabilistic error bounds for the matrix low rank approximation. Thus, we can provide theoretical error bounds for the tensor case. In addition, these randomized algorithms are implemented on an MRI dataset.

1 Introduction

The problem of approximating a given matrix $M \in \mathbb{R}^{n\times m}$ with a low rank matrix $\hat{M} \in \mathbb{R}^{n\times k}$ is

$$\min_{\text{rank}(\hat{M})\leq k} \|M-\hat{M}\|_F^2$$

where $\|\cdot\|$ is the matrix Frobenius norm. The solution is

$$M^* = \text{argmin}_{\text{rank}(\hat{M})\leq k}\|M-\hat{M}\|_F^2 = U_k\Sigma_k V_k^T$$

where $U_k\Sigma_k V_k^T = \sum_{i=1}^k \sigma_i u_i \otimes v_i$ is the first k leading terms in the SVD. This is due to Eckart-Young [6]. Similarly, we ask this question for tensors. For a given kth order tensor $\mathscr{T}$,

Carmeliza Navasca
University of Alabama at Birmingham, Department of Mathematics, 1300 University Blvd, Birmingham, AL 35209-1170 e-mail: cnavasca@uab.edu

Deonnia N. Pompey
University of Alabama at Birmingham, Department of Mathematics, 1300 University Blvd, Birmingham, AL 35209-1170 e-mail: nicholep@uab.edu

$$\min \|\mathscr{T} - \hat{\mathscr{T}}\|_F^2 \tag{1}$$

where $\hat{\mathscr{T}}$ is subject to a rank constraint and $\|\mathscr{A}\|_F^2 = \mathrm{tr}\langle \mathscr{A}, \mathscr{A} \rangle$ is a Frobenius norm. For example, for a desired rank k in $\hat{\mathscr{T}} = \sum_{i=1}^{k} \sigma_i u_i^{(1)} \otimes u_i^{(2)} \otimes \cdots \otimes u_i^{(k)}$, does a minimizer in (1) exist? The answer for the tensor case is complicated. The best rank k approximant may exists, but it may not have orthogonal factors like in SVD of matrix[14]. Or the best k rank-one terms may not exist; this is called the degenerate case [5]. A minimizer in (1) is not always guaranteed to exist as opposed to the matrix case where the best rank k solution always exists. In fact, this problem is NP-hard [10]. These tensor questions are important in analyzing datasets with complex structure which appear across many disciplines, namely in, signal processing [2], PDEs [13, 19], geophysics [12], environmental sciences [16], brain connectome [23] and etc. Other interesting applications are in bioinformatics [20] and biomedical imaging [11]; see more applications in [15] and the references therein.

In this work, we propose some algorithms for approximation of a low multilinear rank $\hat{\mathscr{T}}$ from a given tensor $\mathscr{T} \in \mathscr{R}^{n_1 \times n_2 \times n_3}$; i.e

$$\min \|\mathscr{T} - \hat{\mathscr{T}}\|_F^2$$

where $\hat{\mathscr{T}} = \sum_{l=1}^{r_1} \sum_{m=1}^{r_2} \sum_{k=1}^{r_3} (\mathscr{S})_{lmn} (U_1)_{il}, (U_2)_{jm}, (U_3)_{kn}$ (Tucker format) with orthogonal matrices $U_1 \in \mathbb{R}^{n_1 \times r_1}$, $U_2 \in \mathbb{R}^{n_2 \times r_2}$ and $U_3 \in \mathbb{R}^{n_3 \times r_3}$. Here the core tensor is $\mathscr{S} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$. We say that the tensor $\hat{\mathscr{T}}$ has rank-(r_1, r_2, r_3). In these algorithms, we use random projections on matrices based on the work of Halko, Martinsson and Tropp [9] to approximate range space, rank and theoretical bounds. In addition, we prove a sharper probabilistic error bound found in [9] for the matrix case. Then, we apply this error bound for the tensor case. The tensor extension of these randomized projections was first proposed by Mahoney, Maggioni and Drineas [17].

In addition, we demonstrate the efficacy of the proposed randomized algorithms through MRI data compression. Numerical experiments are compared to tensor based methods for compression (e.g. quasi-newton methods on Grassmannian manifolds [22]). We restrict our comparison to tensor based methods for three-dimensional datasets even though there are well-known and powerful methods for compression based on wavelet analysis that has been successful for two-dimensional data. In the work of Wu et al. [26], they have shown that the higher-order SVD within the hierarchical tensor framework has some advantages over wavelet analysis for compression. The advantages are the following: (a) it can achieve far higher quality than wavelet transform at large compression ratios (b) the tensor framework facilitates progressive or partial data transmission and visualization; i.e. the receiver can quickly view the low resolution versions first and decide whether it is worthwhile to wait for higher resolution details.

2 Preliminaries

We denote the scalars in $\mathbb{R}$ with lower-case letters $(\alpha, \beta, \ldots)$ and the vectors with lower-case letters $(a, b, \ldots)$. The matrices are written as upper-case letters $(A, B, \ldots)$ and the symbol for tensors are calligraphic letters $(\mathscr{A}, \mathscr{B}, \ldots)$. The subscripts represent the following scalars: $(\mathscr{A})_{ijk} = a_{ijk}$, $(A)_{ij} = a_{ij}$, $(a)_i = a_i$. The superscripts indicate the length of the vector or the size of the matrices. For example, b^K is a vector with length K and $B^{N\times K}$ is a $N \times K$ matrix. In addition, the lower-case superscripts on a matrix indicate the mode in which it has been matricized. For example, R^n is the mode-n matricization of the tensor $\mathscr{R} \in \mathbb{R}^{I\times J\times K}$ for $n = 1, 2, 3$.

Definition 1. The Kronecker product of matrices A and B is defined as

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \ldots \\ a_{21}B & a_{22}B & \ldots \\ \vdots & \vdots & \ddots \end{bmatrix}.$$

Definition 2 (Mode-n vector). Given a tensor $\mathscr{T} \in \mathbb{R}^{I\times J\times K}$, there are three types of mode vectors, namely, mode-1, mode-2, and mode-3. There are $J \cdot K$ mode-1 vectors that are of length I which are obtained by fixing the indices (j,k) while varying i. Similarly, the mode-2 vector (mode-3 vector) is of length J (K) obtained from the tensor by varying j (k) with fixed (k,i) (i,j).

Definition 3 (Mode-n rank). The mode-n *rank* of a tensor $\mathscr{T}$ is the dimension of the subspace spanned by the mode-n vectors.

The order of a tensor refers to the cardinality of the index set. A matrix is a second-order tensor and a vector is a first-order tensor.

Definition 4 (rank-(L,M,N)). A third-order tensor $\mathscr{T} \in \mathbb{R}^{I\times J\times K}$ is rank-(L,M,N) if the mode-1 rank is L, the mode-2 rank is M and the mode-3 rank is N.

In the case when a third-order tensor has rank-$(1,1,1)$, it is simply called a *rank-1* tensor.

Definition 5 (Tucker mode product). Given a tensor $\mathscr{T} \in \mathbb{R}^{I\times J\times K}$ and the matrices $A \in \mathbb{R}^{\hat{I}\times I}$, $B \in \mathbb{R}^{\hat{J}\times J}$ and $C \in \mathbb{R}^{\hat{K}\times K}$, then the Tucker mode-$n$ products are as follows:

$$(\mathscr{T} \bullet_1 A)_{\hat{i},j,k} = \sum_{i=1}^{I} (\mathscr{T})_{ijk} a_{\hat{i}i},\ \forall \hat{i}, j, k \quad \text{(mode-1 product)}$$

$$(\mathscr{T} \bullet_2 B)_{\hat{j},i,k} = \sum_{j=1}^{J} (\mathscr{T})_{ijk} b_{\hat{j}j},\ \forall \hat{j}, i, k \quad \text{(mode-2 product)}$$

$$(\mathscr{T} \bullet_3 C)_{\hat{k},i,j} = \sum_{k=1}^{K} (\mathscr{T})_{ijk} c_{\hat{k}k},\ \forall \hat{k}, i, j \quad \text{(mode-3 product)}$$

In general, we have the following definition.

Definition 6. Given a kth order tensor $\mathscr{T} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_k}$ and matrices $U_i \in \mathbb{R}^{\hat{n}_l \times n_l}$, then the Tucker mode-$n_l$ is

$$(\mathscr{T} \bullet_l U_l)_{i_1 i_2 \cdots \hat{i}_l \cdots i_k} = \sum_{i_l} (\mathscr{T})_{i_1 i_2 \cdots i_l \cdots i_k} (U)_{\hat{i}_l i_l}$$

Definition 7 (Matricization). Matricization is the process of reordering the elements of an Nth order tensor into a matrix. The mode-n matricization of a tensor $\mathscr{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_k}$ is denoted by T_l and arranges the mode-l fibers to be the columns of the resulting matrix. The mode-l fiber, $t_{n_1 \cdots n_{l-1} : n_{l+1} \cdots n_k}$, is a vector obtained by fixing every index with the exception of the lth index.

If we use a map to express such matricization process for any Nth order tensor $\mathscr{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_k}$, that is, the tensor element $(n_1, n_2, \ldots, n_k)$ maps to matrix element (n_l, j), then there is a formula to calculate j:

$$j = 1 + \sum_{\substack{l=1 \\ l \neq k}}^{k} (n_l - 1) J_l \quad \text{with} \quad J_l = \prod_{\substack{m=1 \\ m \neq l}}^{l-1} n_m.$$

Then, given a third-order tensor $\mathscr{X} \in \mathbb{R}^{I \times J \times K}$, the mode-1, mode-2 and mode-3 matricizations of $\mathscr{X}$, respectively, are:

$$\begin{aligned} X_1 &= [x_{:11}, \ldots, x_{:J1}, x_{:12} \ldots, x_{:J2}, \ldots, x_{:1K}, \ldots, x_{:JK}], \\ X_2 &= [x_{1:1}, \ldots, x_{I:1}, x_{1:2} \ldots, x_{I:2}, \ldots, x_{1:K}, \ldots, x_{I:K}], \\ X_3 &= [x_{11:}, \ldots, x_{I1:}, x_{12:} \ldots, x_{I2:}, \ldots, x_{1J:}, \ldots, x_{IJ:}]. \end{aligned} \tag{2}$$

3 Theoretical Error Bounds

In this section, we introduce a randomization technique for calculating low rank matrices. It is well known from Eckart-Young Theorem [6] that the low rank k matrix approximation is attained from calculating the leading first k rank-one terms in the SVD; i.e.

$$\text{argmin}_{\text{rank}(\hat{A}) \leq k} \|A - \hat{A}\|_F = U_k \Sigma_k V_k^T.$$

Recent results [9] show that the randomized versions of classical numerical linear algebra techniques give fast, efficient and accurate algorithms. Here we build on the theoretical error bounds found in [9] which will be essential for the error bounds in the tensor case.

The goal is to create an algorithm for SVD with randomness. For simplicity, we fix a specific low rank k in mind. Given a matrix $A \in \mathbb{R}^{n \times m}$, a desired rank k, and an oversampling parameter p, we want to construct $Q \in \mathbb{R}^{n \times k+p}$ orthonormal columns such that

$$\|A - QQ^T A\| \approx \min_{\text{rank}(\hat{A}) \leq k} \|A - \hat{A}\|.$$

Essentially, we are constructing a random orthogonal project such that the residual $\|A - QQ^T A\| \leq \varepsilon$. A Gaussian test matrix Ω is used such that the columns of $A\Omega$ are orthonormalized; i.e. $A\Omega = QR$. See Table 1. Thus, Q captures the first k columns as the left singular vectors of A. Then, the matrix $B = Q^T A$ of size $k+p \times m$ is formed and factored into its SVD; ie. $B = \hat{U}\Sigma V^T$ which follows that $A = QB = Q\hat{U}\Sigma V^T = U\Sigma V^T$ where $U = Q\hat{U}$. SVD is performed on a smaller matrix B of size $k+p \times m$ as opposed to A of size $n \times m$. See Table 2 for the randomized SVD.

Table 1 Fixed low matrix rank [9].

Input: A, rank k, oversampling paramater p Output: $Q \in \mathbb{R}^{n \times k+p}$
Draw a random $n \times k+p$ test matrix Ω Form the matrix product $Y = A\,\Omega$ Compute a QR: $[Q,R]$=qr$(A\Omega)$

Table 2 Randomized SVD [9].

Input: A, orthogonal matrix Q Output: orthogonal matrices U, V and diagonal matrix Σ; i.e $A = U\Sigma V^T$
Form $B = Q^T A$ Compute a small SVD: $[\hat{U},\Sigma,V^T]$=svd(B) Set $U = Q\hat{U}$

Now we discuss a deterministic error bound for calculating the range space of **A**. Let A be an $m \times n$ that has an SVD of $A = U\ \Sigma\ V^T$; i.e.

$$A = U \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \tag{3}$$

where Σ_1 is an $k \times k$ matrix and Σ_2 is an $n-k \times n-k$ matrix. Let Ω be an $n \times l$ test matrix where $l \geq k$ in the coordinate system determined by the right polar decomposition of A via

$$\Omega_1 = V_1^T \Omega \quad \text{and} \quad \Omega_2 = V_2^T \Omega. \tag{4}$$

A deterministic bound is the following:

Theorem 1. *([9]) Let A be an $m \times n$ matrix with SVD A=$U\Sigma V^T$ (3). Let Ω be a test matrix and construct the matrix Y=$A\Omega$ with Ω_1 and Ω_2 are defined in 4. Assume Ω_1 has full row rank, then*

$$\|A - QQ^T A\|_F^2 \leq \|\Sigma_2\|_F^2 + \|\Sigma_2 \Omega_2 \Omega_1^\dagger\|_F^2.$$

Let V_s be an $n \times s$ matrix whose entries are i.i.d. $N(0,1)$ random variables and let $M_s = \frac{1}{s} V_s V_s^T$. The random matrix M_s is called the Wishart matrix.

Theorem 2. *([7]) Let $M_s = \frac{1}{s} V_s V_s^T$ where V_s is an $n \times s$ matrix with i.i.d. entries with $\sigma_{ij} \sim N(0,1)$. For $0 < \frac{n}{s} < \infty$, the largest singular values of M_s converges a.s. to $\left(1 + \sqrt{\frac{n}{s}}\right)$ as $s \to \infty$.*

Theorem 3. *([24]) Let $M_s = \frac{1}{s} V_s V_s^T$ where V_s is an $n \times s$ matrix with i.i.d. entries with $\sigma_{ij} \sim N(0,1)$. For $0 < \frac{n}{s} < 1$, the smallest singular values of M_s converges a.s. to $\left(1 - \sqrt{\frac{n}{s}}\right)$ as $s \to \infty$.*

Here we improve the theoretical bound in Theorem 1.

Theorem 4. *Let A be an $m \times n$ matrix with SVD A=$U\Sigma V^T$ (3). Let Ω be an $n \times l$ matrix with i.i.d. entries with $a_{ij} \sim N(0,1)$ and construct the matrix Y=A Ω with Ω_1 and Ω_2 are defined in (4). Assume Ω_1 has full row rank, then*

$$\|A - QQ^T A\|_F^2 \leq \left(1 + \frac{1 + \sqrt{\frac{n-k}{l}}}{1 - \sqrt{\frac{k}{l}}}\right) \|\Sigma_2\|_F^2$$

as $l \to \infty$ where $0 < \frac{n-k}{l} < \infty$ and $0 < \frac{k}{l} < 1$.

Proof. Theorem 1 states this bound:

$$\|A - QQ^T A\|_F^2 \leq \|\Sigma_2\|_F^2 + \|\Sigma_2 \Omega_2 \Omega_1^\dagger\|_F^2.$$

We now calculate a bound for $\|\Sigma_2 \Omega_2 \Omega_1^\dagger\|_F^2$. The following is true:

$$\|\Sigma_2 \Omega_2 \Omega_1^\dagger\|_F^2 = \mathrm{tr}((\Omega_1^\dagger)^T \Omega_2^T \Sigma_2^2 \Omega_2 \Omega_1^\dagger) = \mathrm{tr}(\Omega_2^T \Sigma_2^2 \Omega_2 \Omega_1^\dagger (\Omega_1^\dagger)^T).$$

Let $P_1 = \Omega_1^\dagger (\Omega_1^\dagger)^T$ and $P_2 = \Omega_2 \Omega_2^T$ with orthonormal bases $\{\phi_j\}$ and $\{\psi_k\}$ ([18]), respectively. Then

$$\begin{aligned}
\mathrm{tr}(\Omega_2^T \Sigma_2^2 \Omega_2 P_1) &= \mathrm{tr}\langle I, \Omega_2^T \Sigma_2^2 \Omega_2 P_1 \rangle = \sum_i \langle \phi_i, \Omega_2^T \Sigma_2^2 \Omega_2 P \phi_i \rangle \\
&\leq \sum_i \|P\|_2 \langle \phi_i, \Omega_2^T \Sigma_2^2 \Omega_2 \phi_i \rangle \leq \|P_1\|_2 \,\mathrm{tr}\langle I, \Omega_2^T \Sigma_2^2 \Omega_2 \rangle \\
&\leq \|P_1\|_2 \,\mathrm{tr}\langle I, \Sigma_2^2 \Omega_2 \Omega_2^T \rangle \leq \|P_1\|_2 \sum_i \langle \psi_k, \Sigma_2^2 P_2 \psi_i \rangle \\
&\leq \|P_1\|_2 \|P_2\|_2 \sum_i \langle \psi_k, \Sigma_2^2 P_2 \psi_i \rangle \leq \|P_1\|_2 \|P_2\|_2 \|\Sigma_2\|_F^2
\end{aligned}$$

Using Theorem 2, the largest singular value of P_2 converges almost surely to $\sqrt{l}\left(1+\sqrt{\frac{n-k}{l}}\right)$ as $l \to \infty$ for $0 < \frac{n-k}{l} < \infty$. Similarly, using Theorem 3, the largest singular value of P_1 is the reciprocal of the smallest singular value of $\Omega_1\Omega_1^T$; i.e. the largest singular value of P_1 converges almost surely to $\frac{1}{\sqrt{l}\left(1-\sqrt{\frac{k}{l}}\right)}$ as $l \to \infty$ for $0 < \frac{k}{l} < 1$. Thus,

$$\begin{aligned}\|A - QQ^TA\|_F^2 &\leq \|\Sigma_2\|_F^2 + \|\Sigma_2\Omega_2\Omega_1^\dagger\|_F^2 \\ &\leq (1 + \|P_1\|_2\|P_2\|_2)\|\Sigma_2\|_F^2 \\ &\leq \left(1 + \frac{1+\sqrt{\frac{n-k}{l}}}{1-\sqrt{\frac{k}{l}}}\right)\|\Sigma_2\|_F^2\end{aligned}$$

as $l \to \infty$ where $0 < \frac{n-k}{l} < \infty$ and $0 < \frac{k}{l} < 1$. □

This is a probabilistic bound which sharpens the result of [9] and [1]. The deterministic bound of [9] is an improvement on the result of [1]. In the next section, we will apply Theorem 4 for the error bounds in the tensor case.

4 Low Multilinear Rank Tensor Approximation

In this section, we will describe two low multilinear rank tensor approximations as well apply the probabilistics theoretical bounds. An extension of SVD to tensor is called the multilinear SVD [3] (or Higher-order SVD). Recall the Tucker mode-n_l (Definition 6). Given a kth order tensor $\mathscr{T} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_k}$ and matrices $U_i \in \mathbb{R}^{\hat{n}_l \times n_l}$ for $i = 1, \cdots,$, then the Tucker mode-n_l is

$$(\mathscr{T} \bullet_l U_l)_{i_1 i_2 \cdots \hat{i}_l \cdots i_k} = \sum_{i_l} \mathscr{T}_{i_1 i_2 \cdots i_l \cdots i_k} U_{\hat{i}_l i_l}$$

Definition 8 (Multilinear SVD). A third order tensor $\mathscr{T} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_k}$ can be factored into a product of a core third order tensor and three orthogonal matrices; i.e.

$$\mathscr{T} = \mathscr{S} \bullet_1 U_1 \bullet_2 U_2 \cdots \bullet_l U_l \cdots \bullet_k U_k$$

where $\mathscr{S} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_k}$ is the core tensor and $U_iU_i \in \mathbb{R}^{n_l \times n_l}$ are orthogonal matrices. The core tensor satisfies:

- an all-orthogonality constraint for each mode l: $\langle \mathscr{S}_{i_1 i_2 \cdots i_l = \alpha \cdots i_k}, \mathscr{S}_{i_1 i_2 \cdots i_l = \beta \cdots i_k} \rangle = (\sigma^l)^2 \delta_{\alpha,\beta}$ with $\alpha, \beta = 1, \cdots, n_l$
- ordering

$$\|S_{i_l=1}\|_F \geq \|S_{i_l=2}\|_F \geq \|S_{i_l=3}\|_F \geq \cdots \geq \|S_{i_l=n_l}\|_F$$

where $S_{i_l=\alpha} = \mathscr{S}_{i_1 i_2 \cdots i_l=\alpha \cdots i_k}$ and $\|S_{i_l=\alpha}\|_F = \sigma_\alpha^l$ for $\alpha = 1, \cdots, n_l$.

A kth order tensor has k set of singular values $\{\sigma_\alpha^l\}$ and singular vectors $(U_l)_\alpha$ for $\alpha = 1, \cdots, n_l$.

Definition 9 (Low Multilinear SVD). A given third order tensor $\mathscr{T} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_k}$ can be factored into a product of a core third order tensor and three orthogonal matrices; i.e.

$$\mathscr{T} = \mathscr{S} \bullet_1 U_1 \bullet_2 U_2 \cdots \bullet_l U_l \cdots \bullet_k U_k$$

where $\mathscr{S} \in \mathbb{R}^{r_1 \times r_2 \cdots \times r_k}$ is the core tensor and $U_i \in \mathbb{R}^{n_l \times r_l}$ are orthogonal matrices. The core tensor satisfies

- an all-orthogonality constraint for each mode l: $\langle \mathscr{S}_{i_1 i_2 \cdots i_l=\alpha \cdots i_k}, \mathscr{S}_{i_1 i_2 \cdots i_l=\beta \cdots i_k} \rangle = (\sigma^l)^2 \delta_{\alpha,\beta}$ with $\alpha, \beta = 1, \cdots, r_l$
- ordering

$$\|S_{i_l=1}\|_F \geq \|S_{i_l=2}\|_F \geq \|S_{i_l=3}\|_F \geq \cdots \geq \|S_{i_l=n_l}\|_F$$

where $S_{i_l=\alpha} = \mathscr{S}_{i_1 i_2 \cdots i_l=\alpha \cdots i_k}$ and $\|S_{i_l=\alpha}\|_F = \sigma_\alpha^l$ for $\alpha = 1, \cdots, r_l$.

A kth order tensor has k set of singular values $\{\sigma_\alpha^l\}$ and singular vectors $(U_l)_\alpha$ for $\alpha = 1, \cdots, r_l$.

There are methods for computing (low) multilinear SVD [25, 3, 4]. One approach is to flatten (Definition 2) the tensor and apply matrix SVD. A third order tensor $\mathscr{T}$ is matricized into $T_1 = U_1 S_1 (U_2 \otimes U_3)^T$, $T_2 = U_2 S_2 (U_3 \otimes U_1)^T$ and $T_3 = U_3 S_3 (U_1 \otimes U_2)^T$. To obtain orthogonal matrices U_i, each matrix T_i is decomposed into its SVD. The core tensor $\mathscr{S}$ is then reconstructed. Low rank orthogonal matrices are constructed by taking the first r_i columns of $U_i \in \mathbb{R}^{n_i \times r_i}$. The low rank core tensor $\mathscr{S} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ is built via $\mathscr{S} = \mathscr{T} \bullet_1 U_1^T \bullet_2 U_2^T \bullet_3 U_3^T$. Moreover, the randomized SVD in Table 1 which is based on random projections can applied to each T_i. See Table 3 and Figure 1. In the next section, we describe an improvement of this method.

Table 3 Randomized Multilinear SVD (or HOSVD)

Input: $\mathscr{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, rank $-(r_1, r_2, r_3)$, oversampling parameter p_1, p_2, p_3
Output: orthogonal matrices U_l and diagonal tensor $\mathscr{S}$ in $\mathscr{T} = \mathscr{S} \bullet_1 U_1 \bullet_2 U_2 \bullet_3 U_3$
Reshape $\mathscr{T}$ into matrices T_1, T_2, T_3
$U_1 \leftarrow randsvd(T_1, r_1, p_1)$
$U_2 \leftarrow randsvd(T_2, r_2, p_2)$
$U_3 \leftarrow randsvd(T_3, r_3, p_3)$
Form $\mathscr{S} = \mathscr{T} \bullet_1 U_1^T \bullet_2 U_2^T \bullet_3 U_3^T$

4.1 A Randomized Multilinear Orthogonal Iteration (MOI)

The HOOI method is based on minimizing the Frobenius norm squared of the residual:

$$\begin{aligned}\min_{U_1,U_2,U_3} \|\mathscr{T} - \mathscr{S} \bullet_1 U_1 \bullet_2 U_2 \bullet_3 U_3\|_F^2 &= \min_{U_1,U_2,U_3} \|\mathscr{T}\|_F^2 - 2\langle \mathscr{S}, \mathscr{S}\rangle + \|\mathscr{S}\|_F^2 \\ &= \max_{U_1,U_2,U_3} \|\mathscr{S}\|_F^2\end{aligned}$$

where $\mathscr{S} = \mathscr{T} \bullet_1 U_1^T \bullet_2 U_2^T \bullet_3 U_3^T$. In general, the orthogonal matrices are obtained through

$$\max_{U_1,\cdots,U_k} \|\mathscr{T} \bullet_1 U_1^T \bullet_2 U_2^T \cdots \bullet_k U_k^T\|_F^2$$

The implementation of this maximization problem is an alternating maximization of the matricized subproblems:

$$\max_{U_l} \|U_l^T Z\|_F^2 \tag{5}$$

where $Z = S_l(U_1^T \otimes U_2^T \cdots \otimes U_{l-1}^T \otimes U_{l+1}^T \cdots U_k)$. Orthogonal U_l is attained by setting the leading r_l singular vectors of Z [8]. Here we show an alternative proof in the spirit of Regalia [21] on how the maximum is attained in (5); i.e.

$$\max_{U_l} \|U_l^T Z\|_F^2$$

where $Z = S_l(U_1^T \otimes U_2^T \cdots \otimes U_{l-1}^T \otimes U_{l+1}^T \cdots U_k)$.

Assume Z has an SVD; i.e. Z=$X\Sigma Y^T$. We calculate

$$\begin{aligned}\|U_l^T Z\|_F^2 &= \mathrm{tr}\langle U_l^T Z, U_l^T Z\rangle = \mathrm{tr}(Z^T U U^T Z) = \mathrm{tr}(Y\Sigma X^T U U^T X \Sigma Y^T) = \mathrm{tr}(X^T U U^T X \Sigma^2) \\ &= \mathrm{tr}\langle X^T U U^T X, \Sigma^2\rangle = \mathrm{tr}\sum_{i,j}(X^T U U^T X)_{ij}(\Sigma^2)_{ji} = \sum_i (X^T U U^T X)_{ii}(\sigma(\Sigma^2))_i \\ &\leq \sum_i (\sigma(\Sigma^2))_i\end{aligned}$$

since $\|X^T U U^T X\|_2 \leq \|X^T\|_2\|U\|_2\|U^T\|_2\|X\|_2 \leq 1$. If follows that when $U = X$, the maximum of the objective function is attained.

Here we describe how randomness is introduced to the orthogonal iteration. Let $\mathscr{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Consider the following subproblems from the HOOI formulation: find orthogonal Q_i, $i = 1,2,3$ such that

$$\begin{aligned}\|T_1 - Q_1 Q_1^T T_1\|_F^2 &= \min_{\mathrm{rank}(U_1)\leq r_1} \|T_1 - U_1 S_1 (U_2 \otimes U_3)^T\|_F^2 \\ \|T_2 - Q_2 Q_2^T T_2\|_F^2 &= \min_{\mathrm{rank}(U_2)\leq r_2} \|T_2 - U_2 S_2 (U_3 \otimes U_1)^T\|_F^2 \\ \|T_3 - Q_3 Q_3^T T_3\|_F^2 &= \min_{\mathrm{rank}(U_3)\leq r_3} \|T_3 - U_3 S_3 (U_1 \otimes U_2)^T\|_F^2\end{aligned}$$

Thus, we can iteratively calculate $Q_i^{(k)}$ and $U_i^{(k)}$ until a stopping criteria is satisfied. Moreover, we calculate the theoretical bounds for each mode.

Theorem 5. *Let $T_l \in \mathbb{R}^{n_l \times n_1 \cdots n_{l-1} n_{l+1} \cdots n_k}$ be a matricization of a kth-order tensor with SVD T_l=$U_l \Sigma (U_1 \otimes \cdots U_{l-1} \otimes U_{l+1} \cdots \otimes U_k)^T$ (3). Let Ω_l be an $N_l \times L_l$ matrix with i.i.d. entries with $a_{ij} \sim N(0,1)$ and construct the matrix Y_l=$T_l\ \Omega$ with Ω_1 and Ω_2 are defined in (4) where $N = n_1 \cdots n_{l-1} n_{l+1} \cdots n_k$. Assume Ω_1 has full row rank, then*

$$\|T_l - Q_l Q_l^T T_l\|_F^2 \leq \left(1 + \frac{1 + \sqrt{\frac{N_l - r_l}{L_l}}}{1 - \sqrt{\frac{r_l}{L_l}}}\right) \|\Sigma_2\|_F^2$$

as $L_l \to \infty$ where $0 < \frac{N_l - r_l}{L_l} < \infty$ and $0 < \frac{r_l}{L_l} < 1$.

Remark 1. It follows that we can bound (1) but the theorem above; i.e.

$$\|\mathscr{T} - \hat{\mathscr{T}}\|_F^2 \leq \left(1 + \frac{1 + \sqrt{\frac{N_l - r_l}{L_l}}}{1 - \sqrt{\frac{r_l}{L_l}}}\right) \|\Sigma_2\|_F^2$$

for any mode l since $\|\mathscr{T} - \hat{\mathscr{T}}\|_F^2 = \|T_l - Q_l Q_l^T T_l\|_F^2$ by rearrangements of elements.

Table 4 Randomized Higher Order Orthogonal Iteration (HOOI)

Input: $\mathscr{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, rank $-(r_1, r_2, r_3)$, oversampling parameter p_1, p_2, p_3
Output: orthogonal matrices U_l and diagonal tensor $\mathscr{S}$ in $\mathscr{T} = \mathscr{S} \bullet_1 U_1 \bullet_2 U_2 \bullet_3 U_3$
Reshape $\mathscr{T}$ into matrices T_1, T_2, T_3
for $k = 1, \cdots, MAXit$
$Z_1 \leftarrow T_1(U_2^{k-1} \otimes U_3^{k-1})$
$U_1^k \leftarrow randsvd(Z_1, r_1, p_1)$
$Z_2 \leftarrow T_2(U_3^{k-1} \otimes U_1^k)$
$U_2^k \leftarrow randsvd(Z_2, r_2, p_2)$
$Z_3 \leftarrow T_3(U_1^k \otimes U_3^k)$
$U_3^k \leftarrow randsvd(Z_3, r_3, p_3)$
Form $\mathscr{S} = \mathscr{T} \bullet_1 U_1^T \bullet_2 U_2^T \bullet_3 U_3^T$
end

5 Numerical Examples

In the first two numerical experiments, the knee MRI dataset was obtained from OsiriX [27]. The first numerical experiment is the implementation of the randomized

HOSVD described in Table 3. The dataset of size $92 \times 92 \times 26$ is compressed with a core tensor of size $35 \times 35 \times 26$. Here for each orthogonal matrix U_i, we took the first r_i columns of U_i where $r_1 = 35, r_2 = 35, r_3 = 26$. The calculated errors are $\|S_1\|_F = 1333.3$, $\|S_2\|_F = 58.2$ and $\|S_3\|_F = 5.1$. See Figure 1. Here $l_1 = l_2 = 85$ and $p_1 = p_2 = 50$.

The second experiment is the implementation of the randomized HOOI in Table 4. The dataset of size $92 \times 92 \times 26$ is compressed with a core tensor of size $35 \times 35 \times 26$. Here for each orthogonal matrix U_i, we took the first r_i columns of U_i where $r_1 = 35, r_2 = 35, r_3 = 26$. The calculated errors are $\|S_1\|_F = 5.1$, $\|S_2\|_F = 5.1$ and $\|S_3\|_F = 5.1$ after two iterations. See Figure 2. Here $l_1 = l_2 = 85$ and $p_1 = p_2 = 50$.

The third experiment is a comparison study of multilinear rank reduction using HOOI, randomized HOOI and Quasi-Newton [22]. In Figure 3, we find a low multilinear rank of (3,3,3) from a tensor $\mathscr{A}$ of size $5 \times 5 \times 5$ while having a stopping criteria of a maximum number of iterations of 100 and the error norm, $\|U_3 A_3 (U_1 \otimes U_2)\|_F^2$, to be within 10^{-6}. HOOI maxed out at 100 iterations and Quasi-Newton required 99 iterations while Randomized HOOI needed 18 iterations. For

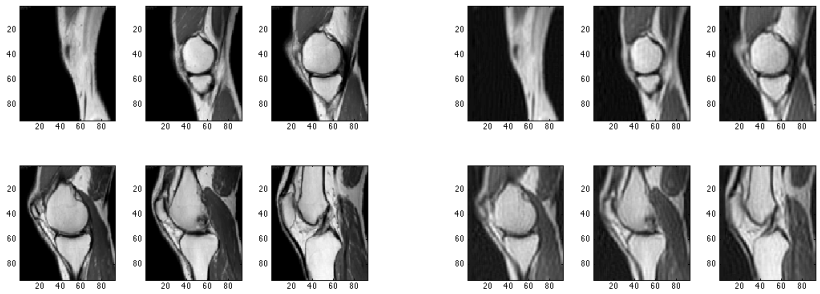

Fig. 1 Randomized HOSVD Example. (left) Six frames of the original data of size $92 \times 92 \times 26$. (right) Reconstructed data using Randomized HOSVD with low multilinear rank-$(35, 35, 26)$.

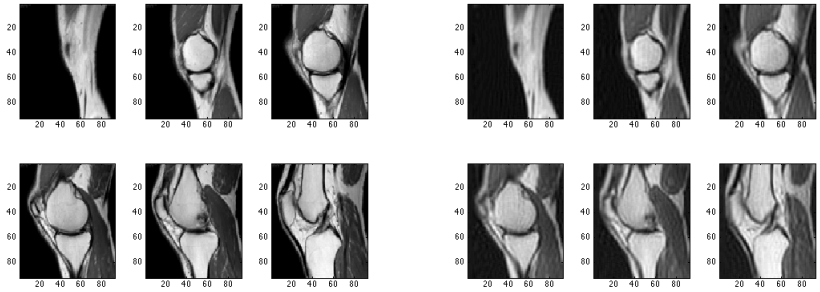

Fig. 2 Randomized HOOI Example. (left) Six frame of the original data of size $92 \times 92 \times 26$. (right) Reconstructed data using Randomized HOOI with low multilinear rank-$(35, 35, 26)$.

the randomized HOOI, we take $L_3 = 5$ (with an oversampling parameter of $p = 2$) and the desired rank of $r_3 = 3$.

Acknowledgements C.N. would like to thank Shannon Starr for pointing out some important references. We would also like to thank the reviewers for their valuable suggestions and comments.

References

1. Boutsidis, C. Mahoney, M. W. , and Drineas, P.: An improved approximation algorithm for the column subset selection problem, in Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), (2009), pp. 968-977.
2. Comon, P. and Jutten, C.: Handbook of Blind Source Separation: Independent component analysis and applications. Academic Press, 2010.
3. De Lathauwer, L. De Moor, B., Vandewalle, J.: A multilinear singular value decomposition. SIMAX, 21, (2000), pp. 1253-1278.
4. De Lathauwer, L., De Moor, B., Vandewalle, J.: On the best rank-1 and rank-$(R_1, R_2, ..., R_N)$ approximation of higher-order tensors. SIMAX, 21, (2000), pp. 1324-1342.
5. V. De Silva and L.-H. Lim, "Tensor rank and the ill-posedness of the best low-rank approximation problem," SIAM Journal on Matrix Analysis and Applications, 30 (2008), no. 3, pp. 1084–1127.

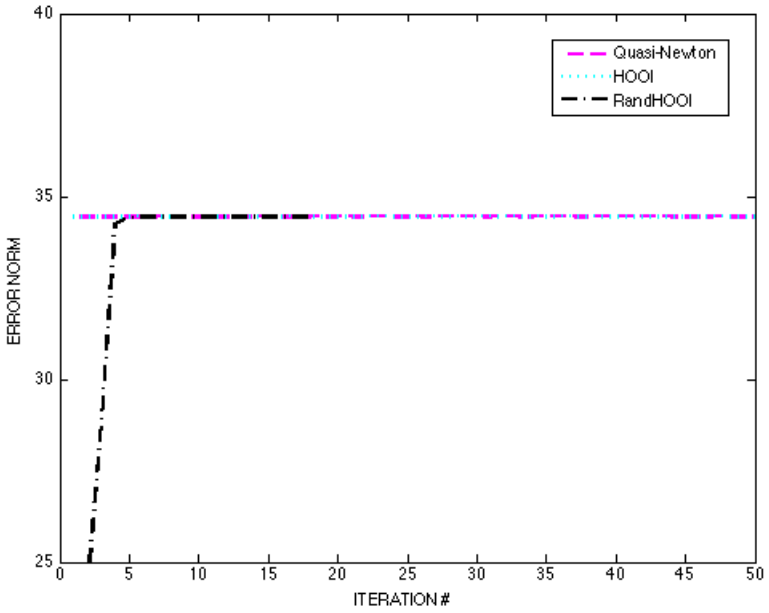

Fig. 3 Comparison of HOOI, Randomized HOOI and Quasi-Newton Methods.

6. Eckart, C. and Young, G.: The approximation of one matrix by another of lower rank. Psychometrika, 1 (3) (1936), pp. 211-218.
7. Geman, S.: A limit theorem for the norm of random matrices. Ann. Probab., 8 (1980), pp. 252-261.
8. Golub, G. and Van Loan, C.: Matrix Computations. Johns Hopkins University Press, 1996.
9. Halko, N., Martinsson, P.G. and Tropp, J.A.: Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. SIREV, 53, (2) (2011), pp. 217-288.
10. Hillar, C. J. and Lim, L.-H.: Most tensor problems are NP-hard. J. ACM 60, 6, Article 45 (2013), (39 pages).
11. Jiao, F., Gur, Y., Johnson, C.R., Joshi, S.: Detection of crossing white matter fibers with high-order tensors and rank-k decompositions, In Proceedings of the International Conference on Information Processing in Medical Imaging (IPMI 2011), Lecture Notes in Computer Science (LNCS), Vol. 6801, pp. 538–549. 2011. DOI: 10.1007/978-3-642-22092-044
12. Kreimer, N. and Sacchi, M. D.:A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation. Geophysics 77(3), V113-V122 (2012)
13. Kressner, D. and Tobler, C.: Low-rank tensor Krylov subspace methods for parametrized linear systems. SIMAX, 32(4) (2011), pp.1288-1316.
14. Kolda,T. G.: Orthogonal tensor decompositions. SIMAX, 23 (2001), pp. 243-255.
15. Kolda, T. and Bader, B.: Tensor Decompositions and Applications. SIREV, 51(3) (2009), pp. 455-500.
16. Li, N., Hopke, P., Pramod, K., Cliff, S., Zhao, Y., and Navasca, C.: Source apportionment of time- and size-resolved ambient particulate matter. Chemometrics and Intelligent Laboratory Systems, 129 (2013), pp. 15-20.
17. M. W. Mahoney, M. Maggioni, and P. Drineas, "Tensor-CUR decompositions for tensor-based data," SIAM Journal on Matrix Analysis and Applications, 30 (2008), no. 2, pp. 957–987.
18. Metha, Madan L.: Random Matrices. Elsevier, Amsterdam (2004)
19. Brazell, M., Li, N., Navasca, C. and Tamon, C.: Solving multilinear systems via tensor inversion. SIMAX, 34-2 (2013), pp. 542-570.
20. Ponnapalli, S.P., Saunders, M.A., Van Loan, C.F. and Alter, O.: A Higher-Order Generalized Singular Value Decomposition for Comparison of Global mRNA Expression from Multiple Organisms. Public Library of Science (PLoS) One 6 (12), article e28072 (December 2011); doi: 10.1371/journal.pone.0028072.
21. Regalia, P.A.: Monotonically convergent algorithms for symmetric tensor approximation. Linear Algebra and its Applications, 438 (2) (2013), pp. 875-890.
22. Savas, B. and L.-H. Lim: Quasi-Newton methods on Grassmannians and multilinear approximations of tensors. SIAM Journal on Scientific Computing, 32 (2010), no. 6, pp. 3352–3393.
23. Schultz, T. and Seidel, H.-P.: Estimating Crossing Fibers: A Tensor Decomposition Approach. IEEE Transactions on Visualization and Computer Graphics, 14 (6) (2008), pp. 1635-1642.
24. Silverstein, J.: The smallest eigenvalue of a large dimensional Wishart matrix. Ann. Probab., 13 (4) (1985), pp. 1364-1368.
25. Tucker, L.R.: Some mathematical notes on three-mode factor analysis. Psychometrika, 31, (1966), pp. 279-311.
26. Wu, Q., Xia, T., Chen, C., Lin, H.-Y.S., Wang, H. and Yu, Yizhou.: Hierarchical Tensor Approximation of Multidimensional Visual Data. IEEE Transactions on Visualization and Computer Graphics, 14 (1) (2008), pp. 186-199.
27. DICOM sample image sets repository. http://www.osirix-viewer.com, http://pubimage.hcuge.ch:8080/.