

A Randomized Block Krylov Method for Tensor Train Approximation

Gaohang Yu^{1*}, Jinhong Feng¹, Zhongming Chen¹, Xiaohao Cai²,
Liqun Qi^{1,3}

¹Department of Mathematics, Hangzhou Dianzi University, China.

²School of Electronics and Computer Science, University of Southampton, Southampton, UK.

³Huawei Theory Research Lab, Hong Kong, China.

*Corresponding author(s). E-mail(s): maghyu@163.com;

Contributing authors: fengjinhong0502@163.com; zmchen@hdu.edu.cn;
x.cai@soton.ac.uk; liqun.qi@polyu.edu.hk;

Abstract

Tensor train decomposition is a powerful tool to tackle high-dimensional large-scale tensor data and is not suffering from the curse of dimensionality. It is based on the low-rank approximation of auxiliary unfolding matrices. To accelerate the calculation of the auxiliary unfolding matrices, some randomized algorithms have been proposed; however, they are not suitable for noisy data. The randomized block Krylov method is capable of dealing with heavy-tailed noisy data in low-rank approximation of matrices. In this paper, we propose a randomized algorithm for low-rank tensor train approximation of large-scale tensors based on randomized block Krylov subspace iteration and provide theoretical guarantees. Extensive numerical experiments on synthetic and real-world tensor data demonstrate the great performance of the proposed algorithm.

Keywords: Tensor, tensor train decomposition, randomized algorithm, randomized block Krylov subspace iteration

MSC Classification: 68W20 , 15A18 , 15A69

1 Introduction

In many practical applications, such as machine learning, compressed sensing and data mining, data analysis generally plays the key role. Many large-scale datasets can be naturally expressed as tensors, often exhibiting low-rank structures. As a result, the low-rank tensor approximation has become a powerful tool for tensor data analysis [1, 2, 5, 9, 17, 35].

The main models for low-rank tensor approximation include the CANDECOMP/-PARAFAC (CP) decomposition [5], Tucker decomposition [17], T-product[16], tensor train (TT) decomposition [21], and tensor ring (TR) decomposition [37], etc. CP decomposition is a useful way to represent a large tensor as the sum of a series of rank-1 tensors. Unfortunately, it is not reliable due to the difficulty of finding the optimal CP rank. Tucker decomposition is more stable than CP decomposition, but it is affected by the curse of dimensionality. In comparison, TT decomposition is not affected by the curse of dimensionality and is more reliable. This paper will focus exclusively on TT decomposition, which is becoming increasingly popular due to its stability and efficiency.

TT decomposition is based on the low-rank approximation of auxiliary unfolding matrices. As the computation of the singular value decomposition (SVD) of large-scale auxiliary unfolding matrices is a time-consuming task, various solutions have been proposed in literature. One of these solutions is the randomized algorithms [3, 13, 14, 27–29], which have been shown to be effective in computing low-rank approximations of large-scale matrices. Randomized algorithms can significantly speed up decomposition and produce highly accurate results.

In recent years, numerous researchers have proposed randomized tensor decomposition methods [2, 7, 8, 12, 18, 22, 23, 25, 30–34, 36], showing the well-established effectiveness of randomized algorithms in various tensor decompositions. Huber et al. [15] proposed randomized TT decomposition based on a simple randomized SVD (TT-rSVD), which is much faster than the classical deterministic TT algorithm (i.e., TT-SVD) [21]. To address unknown exact TT rank situations, Che et al. [6] proposed an adaptive randomized TT decomposition. To take advantage of parallelization, Shi et al. [24] proposed parallelizable TT decompositions, including PTT2 and PTT2-onepass (which requires only one pass through the raw data). Daas et al. proposed parallel algorithms for TT arithmetic [11] and proposed randomized algorithms for rounding in the TT format [10]. Kressner et al. [18] proposed streaming TT approximation using two-sided sketching to make high-dimensional data streamable and easy to implement in parallel. Li et al. [19] proposed a fast TT for sparse tensors.

The randomized algorithm projects the original matrix onto a subspace using a random matrix, and power iteration can improve the accuracy of the approximation by reducing the tail energy [13, 26, 29]. However, the simultaneous power method only retains the highest-order term in the Krylov subspace, which limits its applicability and prevents it from handling the problem of under-sized singular value gaps. To address this issue, Musco et al. [20] proposed the randomized block Krylov subspace iteration method (rBKI), which performs better in experiments and does not depend on singular value gaps. Moreover, the approximation can be as accurate as the truncated SVD for rapidly decaying singular values. Recently, Qiu et al. [23] proposed a randomized

Tucker decomposition based on block Krylov iteration and proved its validity on noisy data.

In TT decomposition, direct use of the random algorithm to approximate the auxiliary unfolding matrices will lead to excessive tail energy. The accumulated error will be larger than that of TT-SVD. Given the advantage of randomized methods in reducing tail energy, in this paper, we propose a method based on randomized block Krylov methods for TT decomposition. Combining randomized block Krylov strategies, our proposed method can achieve more accurate TT approximation than existing standard randomized algorithms. Moreover, our method inherits the ability of the randomized block Krylov subspace iteration in matrices to handle noisy data and can achieve near-optimal approximated relative errors. Extensive experiments in synthetic and real-world data demonstrate the excellent performance of the proposed randomized method for low-rank TT approximation.

The rest of the paper is organized as follows. Section 2 briefly introduces symbols, definitions, tensor basics, and recalls the related work. In Section 3, we present our main method and analyze its probabilistic error bound. Extensive numerical results and comparisons validating the great performance of our proposed method are given in Section 4. We conclude in Section 5.

Table 1 Main notations.

Symbols	Notation
x	scalar
X	matrix
\mathcal{X}	tensor
$\mathcal{X}^{(n)}$	mode- n unfolding matrix of \mathcal{X}
\times_n	mode- n product of tensor and matrix
\times_n^m	mode- (n, m) product of tensor and tensor
X^\top	transpose of X
X^\dagger	pseudo-inverse of X
Ξ	expectation

2 Preliminary

2.1 Notations and basic operations

The main symbols used in this paper are shown in Table 1. The mode- n product of tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ by a matrix $B \in \mathbb{R}^{J_n \times I_n}$ is a tensor $\mathcal{C} \in \mathbb{R}^{I_1 \times \dots \times I_{n-1} \times J_n \times I_{n+1} \times \dots \times I_N}$, denoted by $\mathcal{C} = \mathcal{A} \times_n B$. The tensor-tensor product of two tensors $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ and $\mathcal{B} \in \mathbb{R}^{J_1 \times J_2 \times \dots \times J_M}$ with common modes $I_n = J_m$ produces an $(M + N - 2)$ -th order tensor $\mathcal{C} \in \mathbb{R}^{I_1 \times \dots \times I_{n-1} \times I_{n+1} \times \dots \times I_N \times J_1 \times \dots \times J_{m-1} \times J_{m+1} \times \dots \times J_M}$, i.e.,

$$\mathcal{C} = \mathcal{A} \times_n^m \mathcal{B}.$$

The Frobenius norm of a tensor \mathcal{A} is given by $\|\mathcal{A}\|_F = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}$.

Definition 1. (*Matricization* [15]) The α -Matricization is defined as

$$\hat{M}_\alpha : \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N} \rightarrow \mathbb{R}^{m_\alpha \times m_\beta}, \quad (1)$$

where $m_\alpha = I_1 \times I_2 \times \cdots \times I_{i-1}$ and $m_\beta = I_i \times I_{i+1} \times \cdots \times I_N$.

Definition 2. (*TT decomposition* [21]) Given a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, the TT decomposition of \mathcal{A} with $\text{rank}_{\text{TT}} = \mathbf{r} = (r_1, r_2, \dots, r_{N-1})$ is expressed as

$$\mathcal{A} \approx \mathcal{Q}_1 \times_3^1 \mathcal{Q}_2 \times_3^1 \cdots \times_3^1 \mathcal{Q}_N, \quad (2)$$

where $\mathcal{Q}_n \in \mathbb{R}^{r_{n-1} \times I_n \times r_n}$ is a factor tensor for $n = 1, 2, \dots, N$ and $r_0 = r_N = 1$.

Definition 3. (*Tail energy*) The j -th tail energy of a matrix X is defined as

$$\tau_j^2(X) = \min_{\text{rank}(Y) < j} \|X - Y\|_F^2 = \sum_{i \geq j} \sigma_i^2(X), \quad (3)$$

where each σ_i is a singular value.

2.2 TT-SVD

The classical TT decomposition (i.e., TT-SVD) was proposed by Osledets [21]. It was implemented based on the truncated SVD of the auxiliary unfolding matrix using precision to control the truncation parameters. The TT-SVD algorithm is given in Algorithm 1.

Algorithm 1 TT-SVD [21]

Require: Tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and the prescribed accuracy ϵ .
Ensure: Cores $\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_N$ of the TT approximation \mathcal{B} to \mathcal{A} in TT-format satisfying $\|\mathcal{A} - \mathcal{B}\|_F \leq \epsilon \|\mathcal{A}\|_F$.

- 1: Initialization: truncation parameter $\delta = \frac{\epsilon}{\sqrt{N-1}} \|\mathcal{A}\|_F$, $C = \mathcal{A}^{(1)}$, $r_0 = 1$.
- 2: **for** $n = 1, 2, \dots, N-1$ **do**
- 3: $C = \text{reshape}(C, [r_{n-1} I_n, I_{n+1} I_{n+2} \cdots I_N])$.
- 4: Compute the δ -truncated SVD: $C = USV + E$, $\|E\|_F \leq \delta$, $r_n = \text{rank}(S)$.
- 5: Form the factor tensor $\mathcal{Q}_n = \text{reshape}(U, [r_{n-1}, I_n, r_n])$.
- 6: Update $C = SV^\top$.
- 7: **end for**
- 8: Form the last factor tensor with $r_N = 1$, $\mathcal{Q}_N = \text{reshape}(C, [r_{N-1}, I_N, r_N])$.
- 9: **return** $\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_N$.

Let $I = \max\{I_1, \dots, I_N\}$ and $r = \max\{r_1, \dots, r_{N-1}\}$. The computational complexity of TT-SVD is dominated by the $(N-1)$ matrix SVD, i.e., $\mathcal{O}(I^{N+1} + r^2 \sum_{i=1}^{N-3} I^{N+1-i} + rI^3)$, which is still of exponential order. Unfortunately, tensors with e.g. sparse structure do not bring significant improvement in efficiency because the structure information is usually lost after the first SVD. As shown in the following Theorem 1, errors in TT decomposition are accumulated.

Theorem 1 (Theorem 4.3 [21]). Suppose that the unfoldings A_k of the tensor \mathcal{A} satisfy

$$A_k = R_k + E_k, \quad \text{rank}(R_k) = r_k, \quad \|E_k\|_F = \epsilon_k, \quad k = 1, 2, \dots, N-1. \quad (4)$$

Then TT-SVD computes a tensor \mathcal{B} with $\text{rank}_{\text{TT}} = (r_1, r_2, \dots, r_{N-1})$ in the TT-format and

$$\|\mathcal{A} - \mathcal{B}\|_F \leq \sqrt{\sum_{k=1}^{N-1} \epsilon_k^2}. \quad (5)$$

2.3 Randomized TT-SVD

Although TT decomposition can alleviate the curse of dimensionality, the cost of calculating the auxiliary matrix is still of exponential order, mainly due to the SVD of large-scale matrices. Randomized algorithms can significantly accelerate this process. Huber et al. [15] proposed a randomized algorithm for TT decomposition, i.e., TT-rSVD, see Algorithm 2. The matlab function `qr(Y, 0)` produces the “economy size” QR decomposition.

Algorithm 2 Randomized TT decomposition (TT-rSVD)[15]

Input: Tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, target rank $\mathbf{r} = [r_1, r_2, \dots, r_{N-1}]$, oversampling parameter p , and $r_0 = 1$.

Output: $\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_N$.

- 1: **for** $n = 1, 2, \dots, N-1$ **do**
 - 2: $A = \text{reshape}(\mathcal{A}, [r_{n-1} I_n, I_{n+1} I_{n+2} \dots I_N])$.
 - 3: Create random Gaussian matrix $\Omega \in R^{I_{n+1} I_{n+2} \dots I_N \times (r_n+p)}$.
 - 4: Calculate $Y = A\Omega$.
 - 5: Calculate $[Q, \sim] = \text{qr}(Y, 0)$.
 - 6: $Q = Q(:, 1 : r_n)$.
 - 7: Form the factor tensor $\mathcal{Q}_n = \text{reshape}(Q, [r_{n-1}, I_n, r_n])$.
 - 8: Update $\mathcal{A} = \mathcal{A} \times_1 Q^\top \in R^{r_n \times I_{n+1} \times \dots \times I_N}$.
 - 9: **end for**
 - 10: Form the last factor tensor with $r_N = 1$, $\mathcal{Q}_N = \text{reshape}(\mathcal{A}, [r_{N-1}, I_N, r_N])$.
 - 11: **return** $\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_N$.
-

However, when using randomized methods to calculate TT decomposition, the size of the random matrix determines the generated error, and the error of each step will accumulate. The accumulation of errors can cause considerable tail energy and result in the loss of too much original information. To address this problem, oversampling techniques can be employed in randomized algorithms. This technique involves using a larger random matrix to capture more information in the data matrix and then truncating it to improve the accuracy of randomized methods. The error bound of TT-rSVD is given in the following Theorem 2.

Theorem 2 (See [15]). *Given $\mathcal{A} \in R^{I_1 \times I_2 \times \dots \times I_N}$ and $s = r + p$ with $p \geq 4$. For every $u, t \geq 1$, the error bound of TT-rSVD reads*

$$\|\mathcal{A} - P_{2,\dots,N}(\mathcal{A})\| \leq \sqrt{N-1}\eta(r,p) \min_{\text{rank}_{\text{TT}}(\mathcal{B}) \leq r} \|\mathcal{A} - \mathcal{B}\|, \quad (6)$$

with probability at least $(1 - 5t^{-p} - 2e^{-u^2/2})^{N-1}$, where $P_{2,\dots,N}(\mathcal{A}) = \mathcal{Q}_1 \times_3^1 \mathcal{Q}_2 \times_3^1 \dots \times_3^1 \mathcal{Q}_N$ and η is given by

$$\eta = 1 + t\sqrt{\frac{12r}{p}} + ut\frac{e\sqrt{s}}{p+1}. \quad (7)$$

The subspace power iteration technology [13] is an effective method to enhance the accuracy of randomized methods. Replacing A with $(AA^\top)^q A$ can make the singular values of the original matrix decay faster, resulting in smaller tail energy and improved accuracy of randomized methods as shown in the following theorem.

Theorem 3 (Corollary 10.10 in [14]). *Define $B = (AA^\top)^q A$. For a nonnegative integer q and the oversampling parameter p , generate a random matrix Ω , construct the sample matrix $Z = B\Omega$, and calculate $[Q_Z, \sim] = \text{qr}(Z, 0)$. Let $P_Z = Q_Z Q_Z^\top$, and then*

$$\mathbb{E} \| (I - P_Z) A \| \leq \left[\left(1 + \sqrt{\frac{r}{p-1}} \right) \delta_{r+1}^{2q+1} + \frac{e\sqrt{s}}{p} \left(\sum_{j>r} \delta_j^{2(2q+1)} \right)^{\frac{1}{2}} \right]^{\frac{1}{2q+1}}. \quad (8)$$

Musco et al. [20] proposed an improved version of the simultaneous power iteration, a simple randomized block Krylov method, which gives the same guarantee in just $\mathcal{O}(\frac{1}{\sqrt{\epsilon}})$ iterations (recall that ϵ is the prescribed accuracy) and performs substantially better experimentally. In the following section, we propose our new randomized method based on the block Krylov iteration for TT decomposition.

3 Proposed Block Krylov Iteration for TT Approximation

We emphasize again that one big disadvantage of TT-rSVD is its limited accuracy and inability in tackling noisy data. To address this issue, we propose a new method called TT-rBKI based on the block Krylov subspace iteration for TT approximation.

Our TT-rBKI method is summarized in Algorithm 3. Comparing Step 6 of Algorithm 3 with Step 4 of Algorithm 2 (i.e., TT-rSVD), we can see that TT-rSVD directly utilizes a Gaussian random matrix to sketch the unfolding matrix A , while our TT-rBKI takes advantage of the Krylov subspace to create a larger random projection matrix U based on the data. This means that the TT-rBKI method produces a larger sketch Y through a data-related random projection operator. It is evident that the sketch obtained by this way ensures that it contains more principal components of the data than the sketch obtained by the TT-rSVD method. This plays a key role in improving the accuracy and stability of the method for the TT approximation. The error bound of the proposed TT-rBKI method is provided in Theorem 4 below.

Algorithm 3 Randomized TT approximation based on block Krylov subspace iteration (TT-rBKI)

Input: Tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, target rank $\mathbf{r} = [r_1, r_2, \dots, r_{N-1}]$, power iteration parameter q , oversampling parameter p , and $r_0 = 1$.

Output: $\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_N$.

- 1: **for** $n = 1, 2, \dots, N-1$ **do**
- 2: $A = \text{reshape}(\mathcal{A}, [r_{n-1} I_n, I_{n+1} I_{n+2} \cdots I_N])$.
- 3: Create random Gaussian matrix $\Omega \in \mathbb{R}^{I_{n+1} I_{n+2} \cdots I_N \times (r_n+p)}$.
- 4: Construct Krylov space $K = [A^\top A\Omega, (A^\top A)^2\Omega, \dots, (A^\top A)^q\Omega]$.
- 5: Calculate $[U, \sim] = \text{qr}(K, 0)$, where $U \in \mathbb{R}^{I_{n+1} I_{n+2} \cdots I_N \times q(r_n+p)}$.
- 6: Calculate $Y = AU \in \mathbb{R}^{r_{n-1} I_n \times q(r_n+p)}$.
- 7: Set $(Q, \sim) = \text{qr}(Y, 0)$.
- 8: $Q = Q(:, 1 : r_n)$.
- 9: Form the factor tensor $\mathcal{Q}_n = \text{reshape}(Q, [r_{n-1}, I_n, r_n])$.
- 10: Update $\mathcal{A} = \mathcal{A} \times_1 Q^\top \in \mathbb{R}^{r_n \times I_{n+1} \times \cdots \times I_N}$.
- 11: **end for**
- 12: Form the last factor tensor with $r_N = 1$, $\mathcal{Q}_N = \text{reshape}(\mathcal{A}, [r_{N-1}, I_N, r_N])$.
- 13: **return** $\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_N$.

Theorem 4 (Error bound of TT-rBKI). *Given $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, $s = r + p$ with $p \geq 4$, and nonnegative integer q . For every $u \geq 1$, the error bound of TT-rBKI is*

$$\|\mathcal{A} - P_{2,\dots,N}(\mathcal{A})\| \leq \sqrt{N-1}\eta(r, p, q) \min_{\text{rank}_{\text{TT}}(\mathcal{B}) \leq r} \|\mathcal{A} - \mathcal{B}\|, \quad (9)$$

with probability at least $(1 - 5t^{-p} - 2e^{-u^2/2})^{N-1}$, where the parameter η is given as

$$\eta = \left[\left(1 + \sqrt{\frac{r}{p-1}} + \frac{e\sqrt{s}}{p} \right) \right]^{\frac{1}{2q+1}}. \quad (10)$$

Proof. For syntactical convenience, we define $B_i = (\hat{M}_{1,\dots,i-1}(A))$ and use orthogonality from right to left. Since $P_{2,\dots,N}(X)$ is an orthogonal projector, we have

$$\begin{aligned} \|\mathcal{A} - P_{2,\dots,N}(\mathcal{A})\|^2 &= \|\mathcal{A}\|^2 - \|P_{2,\dots,N}(\mathcal{A})\|^2 \\ &= \|\mathcal{A}\|^2 - \langle B_2, B_2 \rangle \\ &= \|\mathcal{A}\|^2 - \langle Q_2 B_3, Q_2 B_3 \rangle \\ &= \|\mathcal{A}\|^2 - \langle B_3, Q_2^\top Q_2 B_3 \rangle \\ &= \|\mathcal{A}\|^2 - \langle B_3, B_3 - (I - Q_2^\top Q_2)B_3 \rangle \\ &= \|\mathcal{A}\|^2 - \|B_3\|^2 + \langle B_3, (I - Q_2^\top Q_2)B_3 \rangle \\ &= \|\mathcal{A}\|^2 - \|B_3\|^2 + \|(I - Q_2^\top Q_2)B_3\|^2. \end{aligned} \quad (11)$$

Then, iteratively, we can obtain

$$\begin{aligned}\|\mathcal{A} - P_{2,\dots,N}(\mathcal{A})\|^2 &= \|\mathcal{A}\|^2 - \|B_{N+1}\|^2 + \sum_{i=2}^N \|(I - Q_i^\top Q_i)B_{i+1}\|^2 \\ &= \sum_{i=2}^N \|(I - Q_i^\top Q_i)B_{i+1}\|^2,\end{aligned}\tag{12}$$

where the last step is obtained by using the fact that B_{N+1} has the same norm as \mathcal{A} . Since Q_i is obtained by the block Krylov iteration, which has the same error bound as the simultaneous iteration, by using Theorem 3, we have

$$\begin{aligned}&\|(I - Q_i^\top Q_i)B_{i+1}\|^2 \\ &\leq \left[\left[\left(1 + \sqrt{\frac{r}{p-1}} \right) \delta_{j+1}^{2q+1}(B_{i+1}) + \frac{e\sqrt{s}}{p} \left(\sum_{j>r} \delta_j^{2(2q+1)}(B_{i+1}) \right)^{\frac{1}{2}} \right]^{\frac{1}{2q+1}} \right]^2 \\ &\leq \left[\left[\left(1 + \sqrt{\frac{r}{p-1}} + \frac{e\sqrt{s}}{p} \right) \left(\sum_{j>r} \delta_j^{2(2q+1)}(B_{i+1}) \right)^{\frac{1}{2}} \right]^{\frac{1}{2q+1}} \right]^2 \\ &\leq \eta^2 \sum_{j>r} \delta_j^2(B_{i+1}).\end{aligned}\tag{13}$$

As shown by Halko [14], the application of an orthogonal projection can only decrease the singular values. It follows that

$$\begin{aligned}\|(I - Q_i^\top Q_i)B_{i+1}\|^2 &\leq \eta^2 \sum_{j>r} \delta_j^2(B_{i+1}) \\ &\leq \eta^2 \min_{\text{rank}_{\text{TT}}(\mathcal{B}) \leq r} \|\mathcal{A} - \mathcal{B}\|^2.\end{aligned}\tag{14}$$

By combining Eq. (12) and Eq. (14), this completes the proof. \square

For comparison purpose, we also provide a simple version of TT-rBKI with basic subspace power iteration, called TT-rSI, which is summarized in Algorithm 4. The TT-rSI method has a lower computational complexity than the TT-rBKI method (see Table 2), inasmuch as it only uses the information associated with the highest power in the Krylov subspace. On the other hand, this also makes it less capable of dealing with noisy data than the TT-rBKI method, as demonstrated by extensive numerical experiments in Section 4.

As shown in Table 2, the computational complexity of the proposed TT-rBKI method and the TT-rSI method is $\mathcal{O}(I^N s^2 q^2 + \sum_{i=1}^{N-1} r I^{N-i} s^2 q^2)$ and $\mathcal{O}(I^N s q + \sum_{i=1}^{N-1} r I^{N-i} s q)$, respectively. Based on the work in [15] and [23], we know that the computational complexity of TT-rSVD is $\mathcal{O}(NsI^N)$, which is significantly reduced compared to TT-SVD. Among these three randomized algorithms (i.e., TT-rSVD,

Table 2 Computational complexity of methods TT-SVD, TT-rSVD, TT-rSI and TT-rBKI.

Method	Computational Complexity
TT-SVD	$\mathcal{O}(I^{N+1} + \sum_{i=1}^{N-1} r^2 I^{N-i})$
TT-rSVD	$\mathcal{O}(NsI^N)$
TT-rSI	$\mathcal{O}(I^N sq + \sum_{i=1}^{N-1} rI^{N-i} sq)$
TT-rBKI	$\mathcal{O}(I^N s^2 q^2 + \sum_{i=1}^{N-1} rI^{N-i} s^2 q^2)$

TT-rSI and TT-rBKI), TT-rBKI theoretically has slightly higher computational complexity. The extensive numerical experiments in Section 4 below will demonstrate that the proposed TT-rBKI can achieve the best quality for TT approximation.

Algorithm 4 Randomized TT approximation with subspace power iteration (TT-rSI)

Input: Tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, iteration q , target rank $\mathbf{r} = [r_1, r_2, \dots, r_{N-1}]$, oversampling parameter p , and $r_0 = 1$.

Output: $\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_N$.

- 1: **for** $n = 1, 2, \dots, N - 1$ **do**
 - 2: $A = \text{reshape}(\mathcal{A}, [r_{n-1} I_n, I_{n+1} I_{n+2} \cdots I_N])$.
 - 3: Create random Gaussian matrix $\Omega \in R^{I_{n+1} I_{n+2} \cdots I_N \times (r_n + p)}$.
 - 4: $Y = A\Omega$.
 - 5: Calculate $[Q_0, \sim] = \text{qr}(Y, 0)$.
 - 6: **for** $j = 1, 2, \dots, q$ **do**
 - 7: $\hat{Y}_j = A^\top Q_{j-1}$.
 - 8: $(\hat{Q}_j, \sim) = \text{qr}(\hat{Y}_j, 0)$.
 - 9: $Y_j = A\hat{Q}_j$.
 - 10: $(Q_j, \sim) = \text{qr}(Y_j, 0)$.
 - 11: **end for**
 - 12: $Q = Q(:, 1 : r_n)$.
 - 13: Form factor tensor $\mathcal{Q}_n = \text{reshape}(Q, [r_{n-1}, I_n, r_n])$.
 - 14: Update $\mathcal{A} = \mathcal{A} \times_1 Q^\top \in R^{r_n \times I_{n+1} \times \cdots \times I_N}$.
 - 15: **end for**
 - 16: Form the last factor tensor with $r_N = 1$, $\mathcal{Q}_N = \text{reshape}(\mathcal{A}, [r_{N-1}, I_N, r_N])$.
 - 17: **return** $\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_N$.
-

4 Numerical Experiments

In this section, we test the proposed TT-rBKI algorithm on synthetic and real-world data with and without noise, and make comparison with the related state-of-the-art methods, i.e., TT-SVD, TT-rSVD and TT-rSI.

The quality of the reconstructed tensor is measured by the peak signal-to-noise ratio (PSNR) say ρ_{psnr} and relative error say ϵ_{err} . For tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ and its

low-rank TT approximation $\hat{\mathcal{A}}$, the PSNR is defined as

$$\rho_{\text{psnr}} = 10 \cdot \log_{10} \frac{I_1 I_2 I_3 \|\hat{\mathcal{A}}\|_\infty^2}{\|\mathcal{A} - \hat{\mathcal{A}}\|_F^2}.$$

The relative error of the low-rank reconstruction tensor is defined as

$$\epsilon_{\text{err}} = \|\mathcal{A} - \hat{\mathcal{A}}\|_F / \|\mathcal{A}\|_F.$$

All the codes are implemented in MATLAB with the Tensor Toolbox [4] and TT-Toolbox-master. The calculations are run on a laptop with AMD 5800H CPU (2.50GHz) and 16GB RAM.

4.1 Experiments on noise-free data

We first test and compare different methods on synthetic and real-world data for the noise-free case.

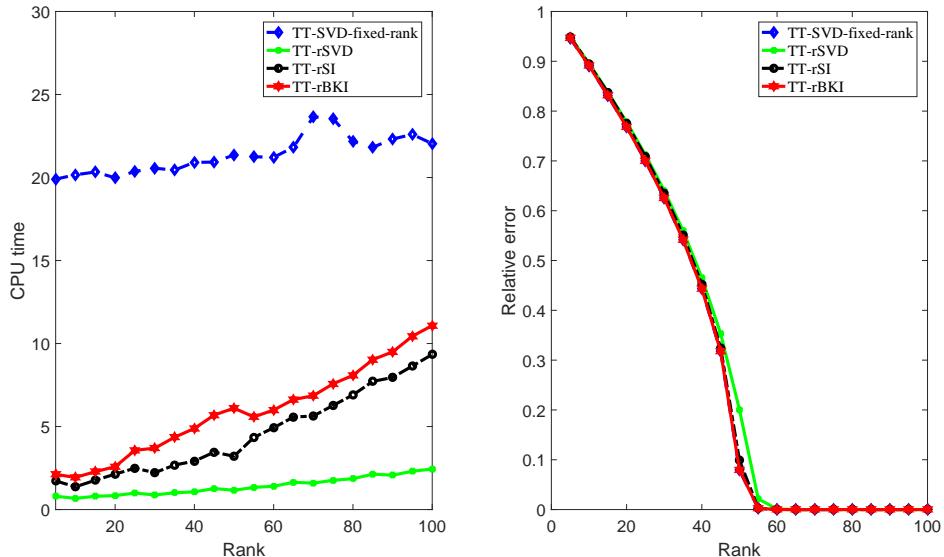


Fig. 1 Results comparison on a synthetic tensor $\mathcal{A} \in \mathbb{R}^{500 \times 500 \times 500}$ with exponentially decaying spectrum in terms of CPU time (left) and relative error (right). Rank is change from 5 to 100 and the oversampling parameter is fixed at 2.

4.1.1 Spectrum decaying tensor

We now test the performance of different algorithms on a synthetic tensor $\mathcal{A} \in \mathbb{R}^{500 \times 500 \times 500}$ with an exponentially decaying spectrum. The diagonal tensor with j -th

frontal slice has the form of

$$\mathcal{A}^{(j)} = \text{diag}(\underbrace{1, 1, \dots, 1}_{\min(T,j)}, 10^{-D}, 10^{-2D}, \dots, 10^{-(n-\min(T,j))D}).$$

We fix $T = 50$ and $D = 1$. The results for this synthetic data are given in Figure 1. Figure 1 shows that the relative error of TT-SVD and TT-rBKI is very similar, but TT-rBKI takes much less CPU time than TT-SVD. Moreover, TT-rSI and TT-rSVD cost less CPU time than TT-rBKI yet TT-rSVD achieves slightly worse accuracy than TT-rBKI when the rank is between 30 and 50.

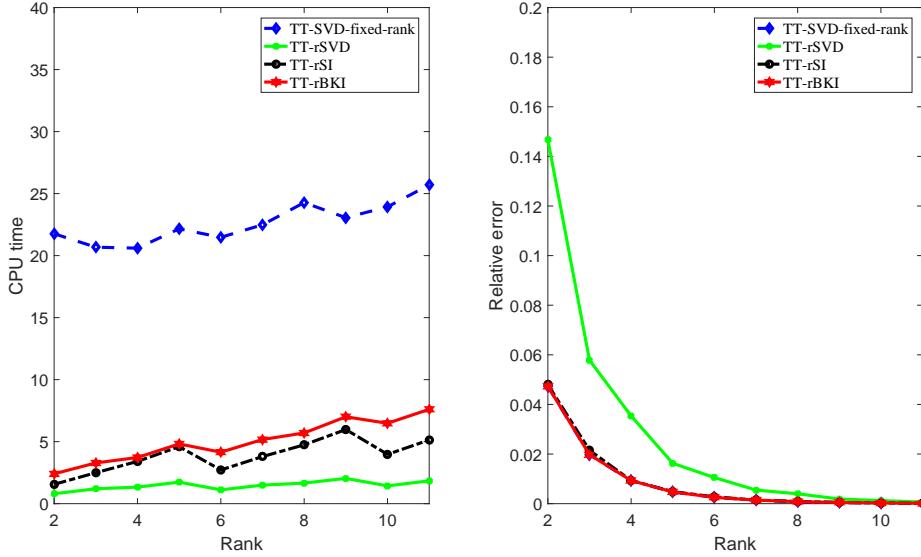


Fig. 2 Results comparison on the power function data with size of $45 \times 45 \times 45 \times 45 \times 45$ in terms of the CPU time (left) and relative error (right). The target rank $[r, r, r, r]$ is changed from 2 to 11 and oversampling parameter p is fixed at 2.

4.1.2 Power function tensor

A five-order power function tensor data can be generated by the following function

$$x_{i_1 i_2, \dots, i_n} = \frac{1}{\sqrt[n]{i_1^h + i_2^h + \dots + i_n^h}}, \quad n = 1, 2, \dots, N. \quad (15)$$

In this synthetic data experiment, we fixed $N = 5$, $h = 5$, and the size of the dataset is $[45, 45, 45, 45, 45]$, while increasing the target rank $[r, r, r, r]$.

The results of the compared methods for this generated synthetic data are displayed in Figure 2. The right panel of the figure reveals that the relative error achieved by

the TT-SVD, TT-rSI and TT-rBKI methods is very similar and much better than that of TT-rSVD. The left panel of Figure 2 reveals that TT-rSI requires less CPU time in seconds than TT-rBKI, while achieving almost the same accuracy as TT-SVD. Moreover, Figure 2 demonstrates that TT-rSVD is the fastest, yet its accuracy is much lower than the other methods, including our TT-rBKI.

4.1.3 Real-world data

In this subsection, we test different algorithms on three real-world tensors. The first real-world tensor is a gray video¹, named hall-quif, with a size of $144 \times 176 \times 150$. The fixed TT-rank $[r_1, r_2]$ is chosen from $r_1 = r_2 = 10$ to $r_1 = r_2 = 20$. The second real-world tensor is derived from hyperspectral data² of size $1096 \times 715 \times 102$. The rank is chosen from $r_1 = r_2 = 5$ to $r_1 = r_2 = 50$. The third real-world tensor is a color video³, i.e., a movie clip, with the size of $480 \times 848 \times 3 \times 147$. We permute the third dimension to the fourth dimension so that the number of frames can be transferred to the third dimension. The rank $[r_1, r_2, r_3]$ is chosen from $r_1 = r_2 = 5$ to $r_1 = r_2 = 50$ with $r_3 = 3$. The oversampling parameter was fixed at 5 for all three datasets.

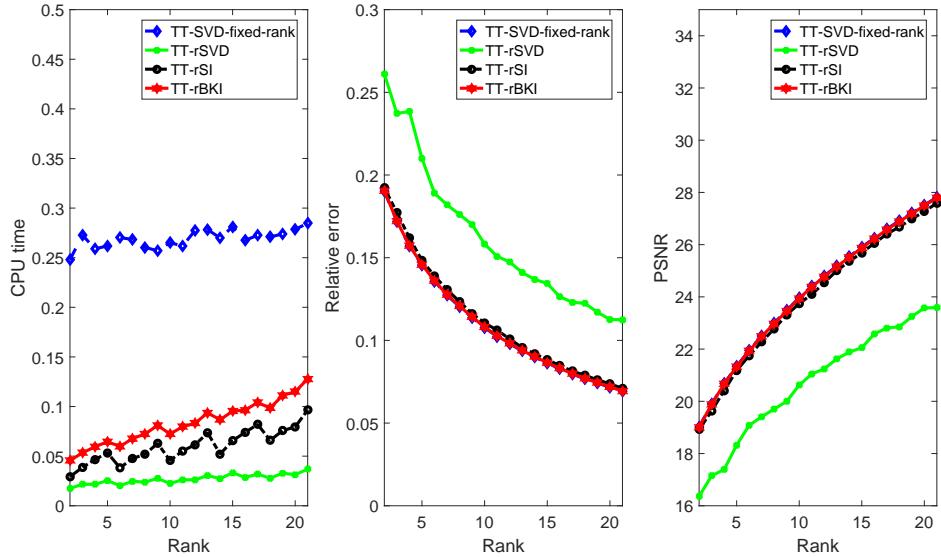


Fig. 3 Results comparison on the gray video clip with size of $144 \times 176 \times 150$ in terms of the CPU time (left), relative error (middle) and PSNR (right).

The comparison between different methods is again conducted in terms of the relative error, PSNR and CPU time. The experimental results on the gray video clip are shown in Figures 3 and 4. The experimental results on the hyperspectral tensor

¹<http://trace.eas.asu.edu/yuv/index.html>

²https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

³<https://github.com/a494626340/data>

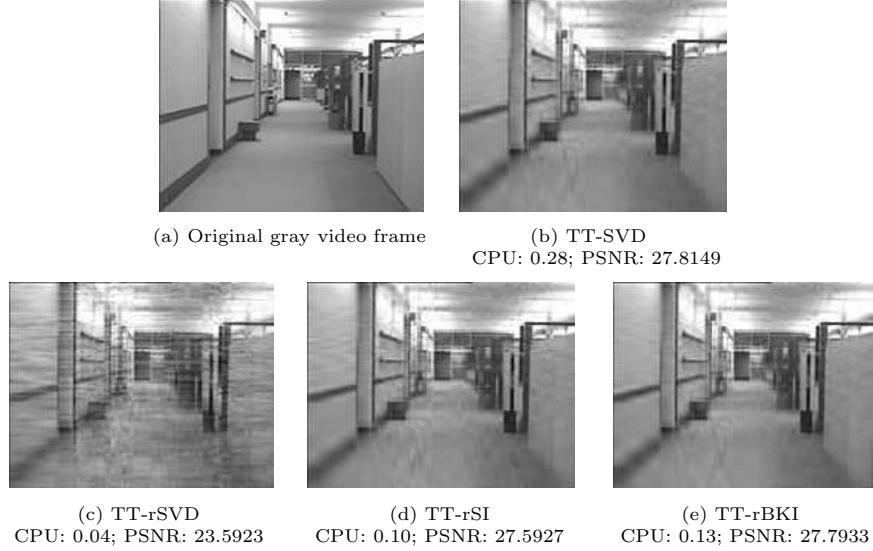


Fig. 4 Low-rank approximation of one frame of the gray video clip by different methods.

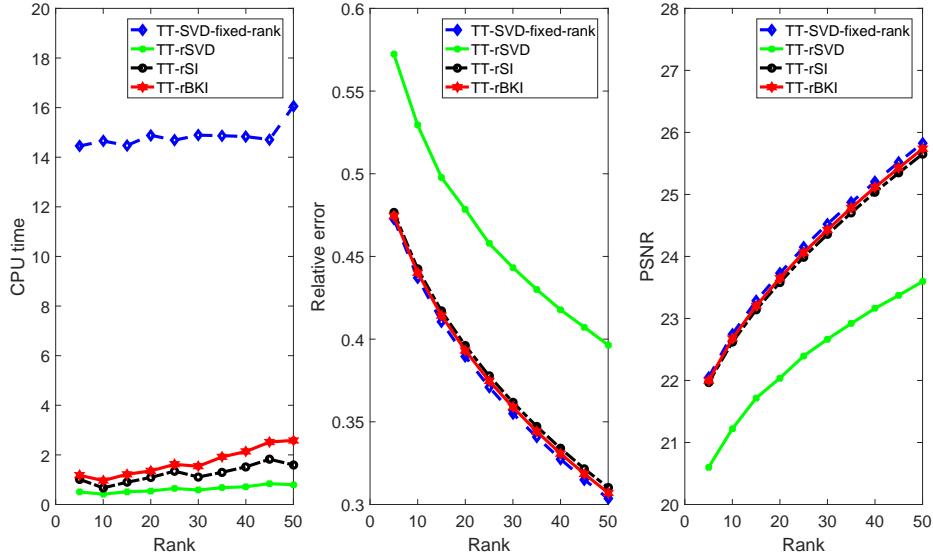


Fig. 5 Results comparison on the hyperspectral tensor with size of $1096 \times 715 \times 102$ in terms of the CPU time (left), relative error (middle) and PSNR (right).

are shown in Figure 5. The experimental results on the color video clip are shown in Figures 6 and 7.

As we can see from Figures 3–7, TT-SVD takes more time on progressively larger data sets (i.e., tensor with larger size), while the time required by the randomized

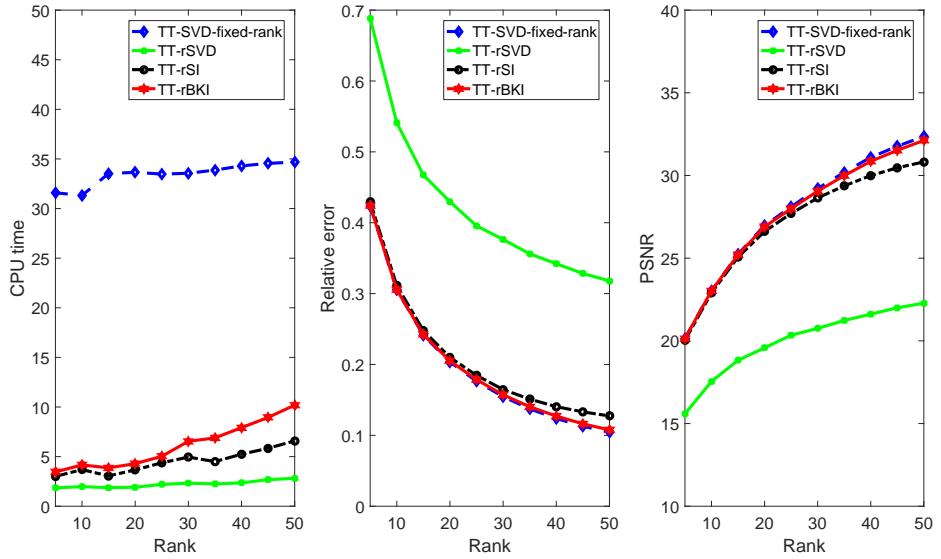


Fig. 6 Results comparison on the color video clip with size of $480 \times 848 \times 3 \times 147$ in terms of CPU time (left), relative error (middle) and PSNR (right).



Fig. 7 Low-rank approximation of one frame of the color video clip by different methods.

algorithms (i.e., TT-rSVD, TT-rSI and TT-rBKI) is much less and does not increase significantly. TT-rSVD, although it requires the shortest time, is not sufficiently accurate. TT-rBKI takes slightly more time, but it achieves results that are closer to those of TT-SVD and are better than those of TT-rSVD and TT-rSI in terms of relative error. On the whole, the proposed TT-rBKI method benefits from the superiority of

the randomized method and subspace iterations in obtaining nearly the same accuracy as that of TT-SVD by taking far less time.

4.2 Experiments on noisy data

We now test and compare different methods on real-world and synthetic data with Gaussian white noise. The built-in MATLAB function `awgn` is used to add noise. The difference from noise-free data is that the noisy data are usually heavy-tailed.

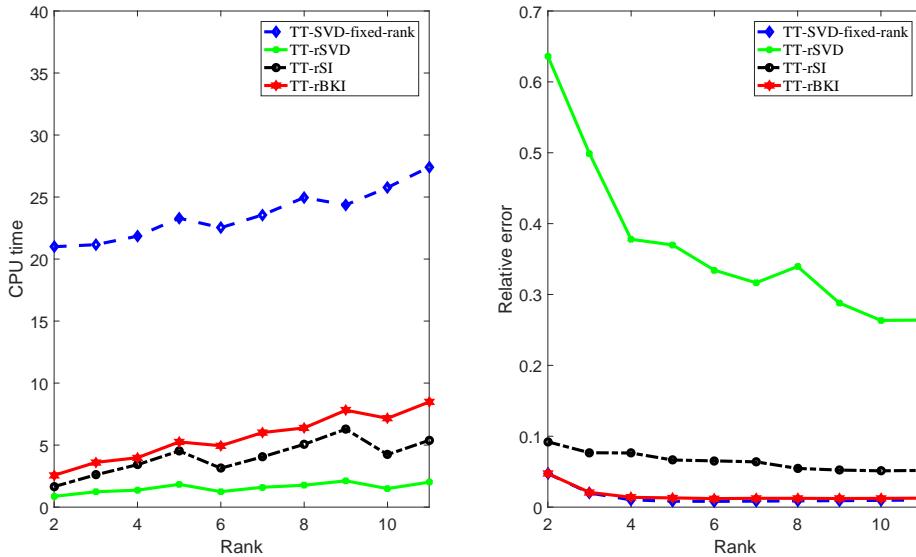


Fig. 8 Results comparison on the noisy power function data with size of $45 \times 45 \times 45 \times 45 \times 45$ in terms of the CPU time (left) and relative error (right). The target rank $[r, r, r, r]$ is changed from 2 to 11 and the noise level is fixed at 5dB.

4.2.1 Synthetic data

We first conduct test to evaluate the impact of noise on the performance of randomized algorithms using synthetic data. The data of a five-order power function defined in Eq. (15) is generated with $N = 5$ and $h = 5$. In this synthetic data experiment, we fix the noise level (i.e., SNR) at 5dB and keep increasing the target rank $[r, r, r, r]$. To illustrate the effect of high noise on individual methods, we fix the target rank and continuously attenuate the noise level. The results are presented in Figures 8 and 9. It shows that in the presence of noise, the proposed TT-rBKI achieves the best reconstruction quality among the randomized methods and requires much less time compared to TT-SVD. In comparison, TT-rSVD and TT-rSI achieve poor approximation results; and the higher the noise level, the worse the approximation.

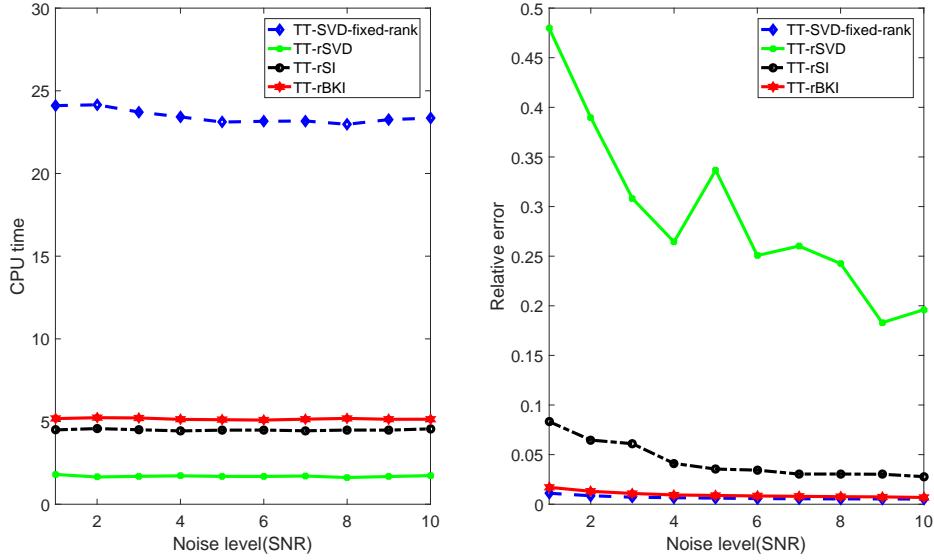


Fig. 9 Results comparison on the noisy power function data with size of $45 \times 45 \times 45 \times 45 \times 45$ and decreased noise level in terms of the CPU time (left) and relative error (right). The target rank $[r, r, r, r]$ is fixed at $[5, 5, 5, 5]$ and the oversampling parameter is fixed at 2.

4.2.2 Real-world data

Finally, we test all the methods on the real-world data with noise. We choose the same video clips used in Section 4.1.3, i.e., the hall-quiv gray video clip and the color movie clip, with the noise level fixed at 5dB and 2dB, respectively. The oversampling parameter was fixed at 5.

The experimental results are presented in Figures 10–15. In the video clips, a frame-showcase is also given in Figures 10 and 13; it is evident that TT-rSVD performs worse than other methods in noisy data, as the approximation it produces is unclear. All the results in Figures 11 and 14 also indicate that the proposed TT-rBKI is superior to TT-rSI, as it has less tail energy loss in its approximation, although it takes slightly more time. Furthermore, it is evident that the reconstruction quality of TT-rBKI, which requires much less time than TT-SVD, is closest to that of TT-SVD. Overall, the proposed TT-rBKI method performs the best among the randomized algorithms, with approximate accuracy comparable to TT-SVD but only requiring 1/5 of the CPU time used by TT-SVD. Figures 12 and 15 give the results of the comparison of all methods at different noise levels while fixing the rank. Consistent results are obtained. In particular, the higher the noise level (i.e., the smaller the SNR), the more obvious the difference. On the whole, the proposed TT-rBKI achieves the best approximate accuracy among the randomized algorithms in all cases.

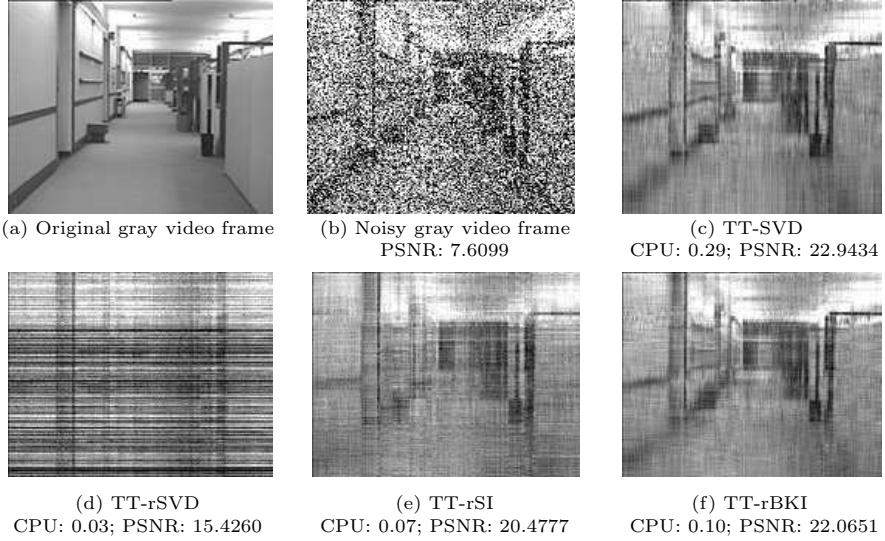


Fig. 10 One frame of the reconstructed noisy gray video clip with size of $144 \times 176 \times 150$ by different methods. The noise level is fixed at 5dB and the rank is 13.

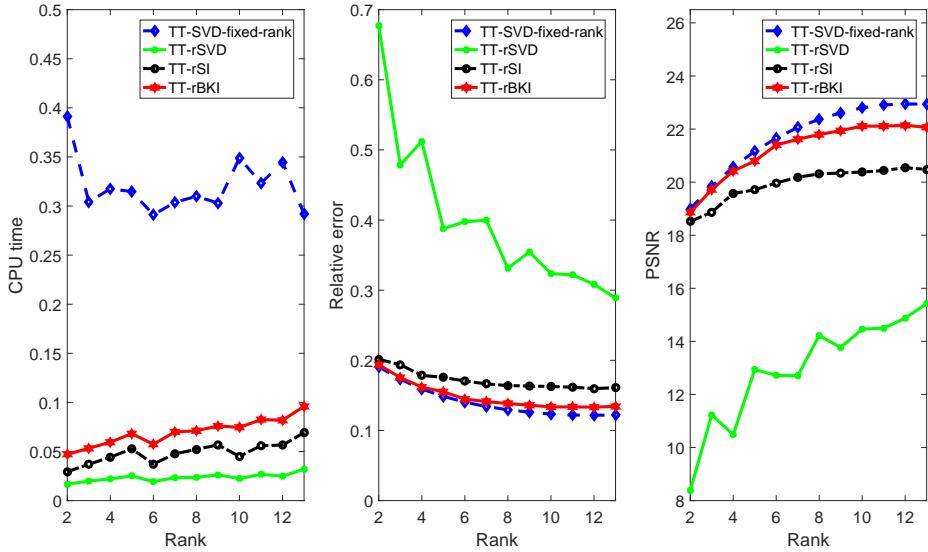


Fig. 11 Results comparison on the noisy gray video clip with the size of $144 \times 176 \times 150$ by different methods in terms of the CPU time (left), relative error (middle) and PSNR (right). The noise level is fixed at 5dB.

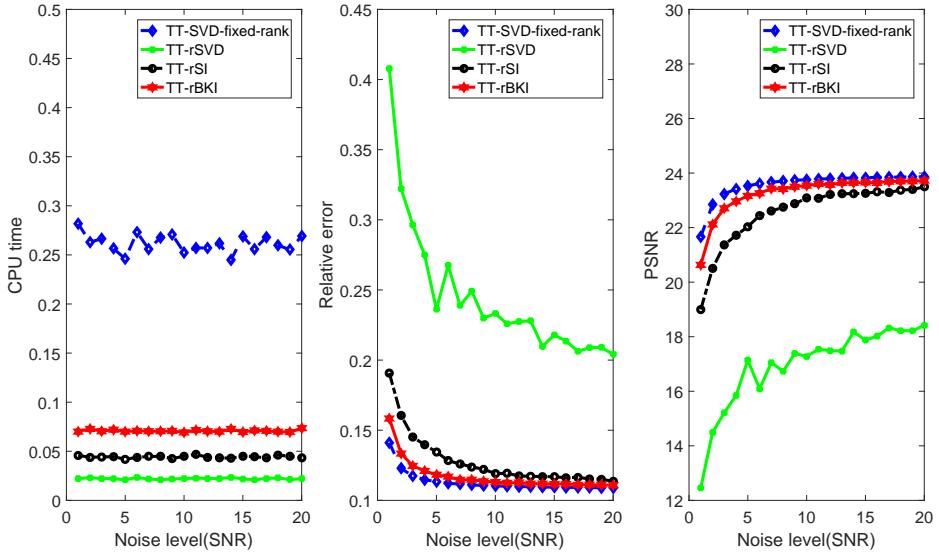


Fig. 12 Results comparison on the noisy gray video clip with size of $144 \times 176 \times 150$ and decreased noise level by different methods in terms of the CPU time (left), relative error (middle) and PSNR (right). The noise level is changed from 1 to 20 and the rank is fixed at 10.



Fig. 13 One frame of the reconstructed noisy color video clip with size of $480 \times 848 \times 3 \times 147$ by different methods. The noise level is fixed at 2dB.

5 Conclusion

In this paper, we proposed a randomized block Krylov subspace iteration, referred to as TT-rBKI, for TT approximation and validated its accuracy with a proof of its error bound. We conducted numerous experiments and comparisons on synthetic and

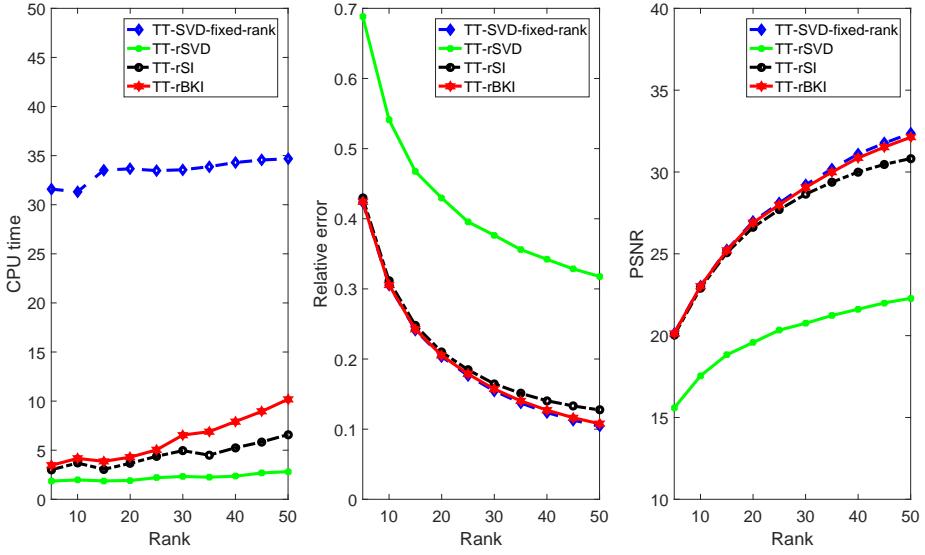


Fig. 14 Results comparison on the noisy color video clip with size of $480 \times 848 \times 3 \times 147$ by different methods in terms of the CPU time (left), relative error (middle) and PSNR (right). The noise level is fixed at 2dB.

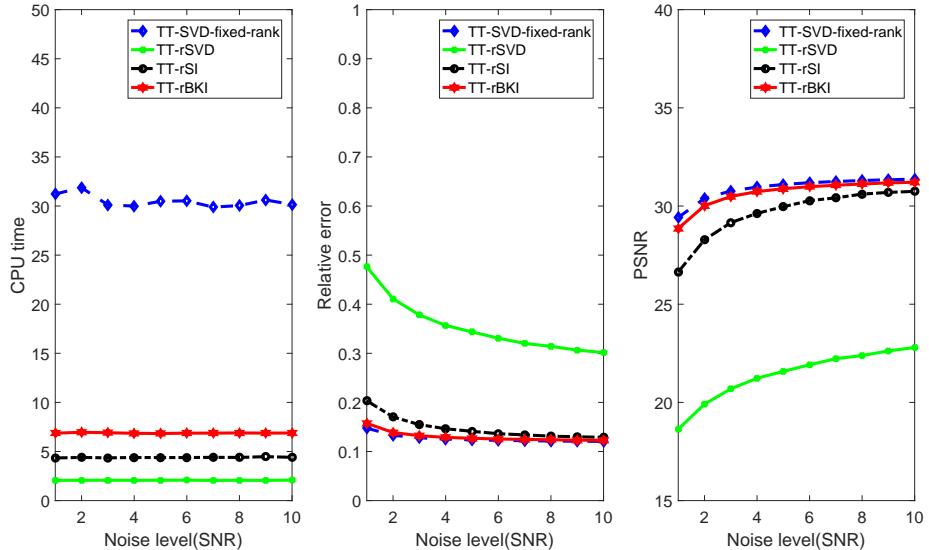


Fig. 15 Results comparison on the noisy color video clip with size of $480 \times 848 \times 3 \times 147$ and decreased noise level (SNR) by different methods (left), relative error (middle) and PSNR (right). The noise level is changed from 1 to 10 and the target rank is fixed at [40,40,3].

real-world data with and without noise to demonstrate its effectiveness. The numerical results showed that the block Krylov subspace iteration can indeed significantly improve the accuracy of the randomized TT approximation. Furthermore, compared to other existing randomized algorithms, the proposed TT-rBKI is more reliable, particularly when dealing with noisy tensor data.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (No. 12071104) and Natural Science Foundation of Zhejiang Province (No. LY22A010012, No. LD19A010002).

Ethics Declarations

We declare that we have no commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Data Availability

All underlying data sets can be made available upon request or are publicly available. MATLAB codes used in this paper are available upon request to the authors.

References

- [1] Ahmadi-Asl S, Cichocki A, Phan A H, et al. Randomized algorithms for fast computation of low rank tensor ring model[J]. Machine Learning: Science and Technology, 2020, 2(1): 011001.
- [2] Ahmadi-Asl S, Abukhovich S, Asante-Mensah M G, et al. Randomized algorithms for computation of Tucker decomposition and higher order SVD (HOSVD)[J]. IEEE Access, 2021, 9: 28684-28706.
- [3] Batselier K, Yu W, Daniel L, et al. Computing low-rank approximations of large-scale matrices with the tensor network randomized SVD[J]. SIAM Journal on Matrix Analysis and Applications, 2018, 39(3): 1221-1244.
- [4] Brett B, Tamara K, et al., Tensor Toolbox for MATLAB, Version 3.5, www.tensortoolbox.org, February 25, 2023.
- [5] Comon P., Tensor decompositions, in Mathematics in Signal Processing V, J. G. McWhirter, I. K. Proudler, Eds., pp. 1–24. Clarendon Press, Oxford, UK, 2002.
- [6] Che M, Wei Y. Randomized algorithms for the approximations of Tucker and the tensor train decompositions[J]. Advances in Computational Mathematics, 2019, 45(1): 395-428.

- [7] Che M, Wei Y, Yan H. The computation of low multilinear rank approximations of tensors via power scheme and random projection. *SIAM Journal on Matrix Analysis Applications*. 41(2), 605-636 (2020)
- [8] Che M, Wei Y, Yan H. An efficient randomized algorithm for computing the approximate Tucker decomposition[J]. *Journal of Scientific Computing*, 2021, 88(2): 32.
- [9] Chen Z, Batselier K, Suykens K, et al. Parallelized tensor train learning of polynomial classifiers [J]. *IEEE transactions on neural networks and learning systems*, 2017, 29(10): 4621–4632.
- [10] Daas A, Ballard G, Cazeaux P, et al. Randomized algorithms for rounding in the tensor-train format [J]. *SIAM Journal on Scientific Computing*, 2023, 45(1): A74–A95.
- [11] Daas H A, Ballard G, Benner P. Parallel algorithms for tensor train arithmetic[J]. *SIAM Journal on Scientific Computing*, 2022, 44(1): C25-C53.
- [12] Dong W, Yu G, Qi L, et al. Practical Sketching Algorithms for Low-Rank Tucker Approximation of Large Tensors[J]. *Journal of Scientific Computing*, 2023, 95(2): 52.
- [13] Gu M. Subspace iteration randomization and singular value problems[J]. *SIAM Journal on Scientific Computing*, 2015, 37(3): A1139-A1173.
- [14] Halko N, Martinsson P G, Tropp J A. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions[J]. *SIAM review*, 2011, 53(2): 217-288.
- [15] Huber B, Schneider R, Wolf S. A randomized tensor train singular value decomposition[C]//Compressed Sensing and its Applications: Second International MATHEON Conference 2015. Springer International Publishing, 2017: 261-290.
- [16] M. Kilmer and C.D. Martin, Factorization strategies for third-order tensors, *Linear Algebra and Its Applications* 435 (2011) 641-658.
- [17] Kolda T G, Bader B W. Tensor decompositions and applications[J]. *SIAM review*, 2009, 51(3): 455-500.
- [18] Kressner D, Vandeheycken B, and Voorhaar R. Streaming Tensor Train Approximation [J]. arXiv e-prints, 2022.
- [19] Li L, Yu W, and Batselier K. Faster tensor train decomposition for sparse data [J]. *Journal of Computational and Applied Mathematics*, 2022, 405: 113972.
- [20] Musco C, Musco C. Randomized block krylov methods for stronger and faster approximate singular value decomposition[J]. *Advances in neural information*

processing systems, 2015, 28.

- [21] Oseledets, I. V.: Tensor-train decomposition. SIAM Journal on Scientific Computing, 2011, 33(5): 2295–2317.
- [22] Qi L, Yu G. T-singular values and t-sketching for third order tensors[J]. arXiv preprint arXiv:2103.00976, 2021.
- [23] Qiu Y, Sun W, Zhou G, et al. Towards Efficient and Accurate Approximation: Tensor Decomposition Based on Randomized Block Krylov Iteration [J]. arXiv preprint arXiv:2211.14828, 2022.
- [24] Shi T, Ruth M, and Townsend A. Parallel algorithms for computing the tensor-train decomposition [J]. arXiv e-prints, 2021.
- [25] Sun Y, Guo Y, Luo C, et al. Low-rank tucker approximation of a tensor from streaming data[J]. SIAM Journal on Mathematics of Data Science, 2020, 2(4): 1123–1150.
- [26] Tropp A. Analysis of randomized block Krylov methods, Technical Report No. 2018-02, 2021.
- [27] Tropp A, Yurtsever A, Udell M, et al. Practical sketching algorithms for low-rank matrix approximation [J]. SIAM Journal on Matrix Analysis and Applications, 2017, 38(4): 1454–1485.
- [28] Tropp J A, Yurtsever A, Udell M, et al. Streaming low-rank matrix approximation with an application to scientific simulation[J]. SIAM Journal on Scientific Computing, 2019, 41(4): A2430-A2463.
- [29] Wang S, Zhang Z, Zhang T. Improved analyses of the randomized power method and block Lanczos method[J]. arXiv preprint arXiv:1508.06429, 2015.
- [30] Wolf A. Low rank tensor decompositions for high dimensional data approximation, recovery and prediction[D]. Technische Universitat Berlin, 2019.
- [31] Yu Y, Li H. Practical sketching-based randomized tensor ring decomposition[J]. arXiv preprint arXiv:2209.05647, 2022.
- [32] Yu Y, Li H, Zhou J. Block-Randomized Stochastic Methods for Tensor Ring Decomposition[J]. arXiv preprint arXiv:2303.16492, 2023.
- [33] Yuan L, Li C, Cao J, et al. Randomized tensor ring decomposition and its application to large-scale data reconstruction[C]//ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019: 2127-2131.

- [34] Zhang J, Saibaba A K, Kilmer M E, et al. A randomized tensor singular value decomposition based on the t-product[J]. Numerical Linear Algebra with Applications, 2018, 25(5): e2179.
- [35] Zhang H, Zhao X L, Jiang T X, et al. Multiscale feature tensor train rank minimization for multidimensional image recovery[J]. IEEE Transactions on Cybernetics, 2021, 52(12): 13395-13410.
- [36] Zhang H, Huang T Z, Zhao X L. A fast sketching-based algorithm for rank-(L, L, 1) block term decomposition[J]. Applied Mathematics Letters, 2022, 133: 108252.
- [37] Zhao Q, Zhou G, Xie S, Zhang L, Cichocki A, Tensor Ring Decomposition, arXiv:1606.05535, 2016.