SIAM J. MATRIX ANAL. APPL.
Vol. 30, No. 3, pp. 939–956

TUCKER DIMENSIONALITY REDUCTION OF THREE-DIMENSIONAL ARRAYS IN LINEAR TIME*

I. V. OSELEDETS†, D. V. SAVOSTIANOV†, AND E. E. TYRTYSHNIKOV†

Abstract. We consider Tucker-like approximations with an $r \times r \times r$ core tensor for three-dimensional $n \times n \times n$ arrays in the case of $r \ll n$ and possibly very large n (up to 10^4–10^6). As the approximation contains only $\mathcal{O}(rn + r^3)$ parameters, it is natural to ask if it can be computed using only a small amount of entries of the given array. A similar question for matrices (two-dimensional tensors) was asked and positively answered in [S. A. Goreinov, E. E. Tyrtyshnikov, and N. L. Zamarashkin, *A theory of pseudo-skeleton approximations*, Linear Algebra Appl., 261 (1997), pp. 1–21]. In the present paper we extend the positive answer to the case of three-dimensional tensors. More specifically, it is shown that if the tensor admits a good Tucker approximation for some (small) rank r, then this approximation can be computed using only $\mathcal{O}(nr)$ entries with $\mathcal{O}(nr^3)$ complexity.

Key words. multidimensional arrays, Tucker decomposition, tensor approximations, low-rank approximations, skeleton decompositions, dimensionality reduction, data compression, large-scale matrices, data-sparse methods

AMS subject classifications. 15A69, 15A18, 65F30

DOI. 10.1137/060655894

1. Introduction.

Multidimensional arrays of data appear in many different applications. One can mention signal processing, statistics [3, 1, 4], chemometrics [5], face recognition [7], and solving multidimensional integral and differential equations [6] (a very comprehensive list of references on the subject can be found on the Three-Mode Company's Web site, [2]). These arrays often cannot be handled by standard methods because of their huge sizes: we cannot solve linear systems or calculate required decompositions due to speed or memory restrictions. The obvious solution is to perform a sort of *dimensionality reduction*: an initial "large" array is transformed to a smaller array for which we can use standard methods. However, such a reduction by conventional approaches may be computationally still too expensive. In this paper we suggest a way to make it not only feasible but even quite fast. We will focus only on three-dimensional arrays mostly to simplify the presentation and note that our results can be generalized to more dimensions.

The most useful method to reduce dimension is based on the celebrated *Tucker decomposition* [22] and solves the following problem: *given a three-dimensional array (tensor)* $\mathcal{A} = [a_{ijk}]$, $i = 1, \dots, n_1$, $j = 1, \dots, n_2$, $k = 1, \dots, n_3$, *compute its approximation*

$$a_{ijk} = \sum_{i'=1}^{r_1} \sum_{j'=1}^{r_2} \sum_{k'=1}^{r_3} g_{i'j'k'} u_{ii'} v_{jj'} w_{kk'} + e_{ijk} \tag{1.1}$$

with the e_{ijk} *to be sufficiently small for prescribed* r_1, r_2, r_3. The matrices $U = [u_{ii'}]$,

*Received by the editors March 31, 2006; accepted for publication (in revised form) by N. Mastronardi October 11, 2006; published electronically September 25, 2008. This research was supported by the Russian Fund of Basic Research (grants 05-01-00721, 04-07-90336, and 06-01-08052) and a Priority Research Grant ONM-3 from the Department of Mathematical Sciences of the Russian Academy of Sciences.

http://www.siam.org/journals/simax/30-3/65589.html

†Institute of Numerical Mathematics, Russian Academy of Sciences, Gubkina Street, 8 IVM RAN, Moscow 119991, Russia (ivan@bach.inm.ras.ru, draug@bach.inm.ras.ru, tee@bach.inm.ras.ru).

$V = [v_{jj'}]$, $W = [w_{kk'}]$ will be referred to as *Tucker factors*, and the $r_1 \times r_2 \times r_3$ tensor $\mathcal{G} = [g_{i'j'k'}]$ as the *core tensor*.

A well-known method for the computation of the Tucker decomposition is based on the SVD algorithm. Consider three rectangular "unfolding" matrices of appropriate sizes $A^{(1)}, A^{(2)}, A^{(3)}$, which contain n-mode vectors (columns, rows, and fibers, respectively) of the tensor $\mathcal{A}$. The left ("short") singular vectors of the SVDs of these matrices

$$A^{(1)} = U\Sigma_1\Phi_1^\top, \qquad A^{(2)} = V\Sigma_2\Phi_2^\top, \qquad A^{(3)} = W\Sigma_3\Phi_3^\top \tag{1.2}$$

give the factors U, V, W of the Tucker decomposition, possibly after an appropriate truncation, and the core is computed as

$$g_{i'j'k'} = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ijk}u_{ii'}v_{jj'}w_{kk'}. \tag{1.3}$$

The tensor dimension can be large (for example, $n = 10^4 - 10^6$ for some tensors coming from three-dimensional integral equations). The array itself cannot even be stored in the operative memory as $\mathcal{O}(n^3)$ memory cells are needed. The computation of the SVDs in (1.2) by standard methods costs $\mathcal{O}(n^4)$ operations and is prohibitive for $n \geq 1000$.

However, we are chiefly interested in the case $r \ll n$, and the Tucker decomposition contains only $\mathcal{O}(rn + r^3)$ parameters. If a good approximation exists, we can ask if it can be computed using only a small amount of entries of the tensor $\mathcal{A}$. A similar question for matrices (two-dimensional tensors) was asked and positively answered in [14]. In the present paper we extend the positive answer to the case of three-dimensional tensors. More specifically, it will be shown that if the tensor admits a good Tucker approximation for some (small) rank r, then this approximation can be computed using only $\mathcal{O}(nr)$ entries with $\mathcal{O}(nr^3)$ complexity.

Prior to investigation of special low-parametric (data-sparse) representations obtained only from the knowledge of a small portion of the data entries, we use a general assumption that *some* low-parametric approximations exist. In other words, we consider the cases with sufficiently small approximate tensor rank estimates. Several estimates for many interesting for practial purposes cases are developed in [15, 18, 19]. We can mention also some practical algorithms using interpolation and other function approximation techniques or additional structural properties rather than the given arrays of data [15, 21]. [20] is closest to the paradigm of a completely data-based method (using no knowledge beyond the data themselves); however, [20] contains no proof for the existence of a sufficiently good low-rank representation and does not suggest a general adaptive procedure for selecting "most meaningful" entries. Recently, much attention has been paid to the approximation of a given matrix by a low-rank matrix using randomized algorithms, for example, see [23]. To our best knowledge, these algorithms are fast only asymptotically with very large constants in the estimates and cannot be applied in practice. Moreover, the authors do not report any numerical results in their articles, so we cannot compare their methods with our method. In this paper we present the existence results and the adaptive three-dimensional cross algorithms.

2. Notations and definitions. Let us recall some basic facts about tensors [10, 11].

DEFINITION 2.1. *The unfoldings of $n_1 \times n_2 \times n_3$ tensor $\mathcal{A}$ are rectangular matrices $A^{(1)}$ of size $n_1 \times n_2 n_3$, $A^{(2)}$ of size $n_2 \times n_1 n_3$, and $A^{(3)}$ of size $n_3 \times n_1 n_2$ with elements*

$$A^{(1)} = [a^1_{i(jk)}] = [a_{ijk}], \quad A^{(2)} = [a^2_{j(ki)}] = [a_{ijk}], \quad A^{(3)} = [a^3_{k(ij)}] = [a_{ijk}]. \tag{2.1}$$

The superscripts $1, 2, 3$ in the definitions of unfoldings define which index (first, second, or third) is used; two other indices are merged into one "long" index.

DEFINITION 2.2. *The norm of the $n_1 \times n_2 \times n_3$ tensor $\mathcal{A} = [a_{ijk}]$ is defined similarly to the Frobenius norm for matrices as*

$$||\mathcal{A}|| = ||\mathcal{A}||_F = \left(\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} a^2_{ijk} \right)^{1/2}.$$

Also, let

$$||\mathcal{A}||_C = \max_{i,j,k} |a_{ijk}|$$

be a Chebyshev norm *of a tensor $\mathcal{A}$.*

DEFINITION 2.3 (outer product). *If $\mathcal{A}$ is a p-index array $a_{i_1,i_2,\ldots,i_p}$ and $\mathcal{B}$ is a q-index array $b_{j_1,j_2,\ldots,j_q}$, then $\mathcal{C} = \mathcal{A} \otimes B$ is defined as a (p+q)-index array with elements*

$$c_{i_1,i_2,\ldots,i_p,j_1,j_2,\ldots,j_q} = a_{i_1,i_2,\ldots,i_p} b_{j_1,j_2,\ldots,j_q}.$$

Tensors can be multiplied by matrices along a specified index (mode) direction.

DEFINITION 2.4 (mode convolution or n-mode product). *If $\mathcal{A} = [a_{ijk}]$ is a $n_1 \times n_2 \times n_3$ array and U is a $m_1 \times n_1$ matrix, then their product $\mathcal{B} = [b_{ijk}]$ is a $m_1 \times n_2 \times n_3$ array defined as*

$$\mathcal{B} = \mathcal{A} \times_i U, \qquad b_{ijk} = \sum_{i'=1}^{n_1} u_{ii'} a_{i'jk}.$$

The operations $\mathcal{A} \times_j U$, $\mathcal{A} \times_k U$ are defined analogously, provided that U and $\mathcal{A}$ have appropriate sizes. In this notation, the Tucker decomposition (1.1) can be written as

$$\mathcal{A} = \mathcal{G} \times_i U \times_j V \times_k W.$$

We will say that a tensor has a *rank-(r_1, r_2, r_3) (Tucker) decomposition* if (1.1) holds [10, 11].

The important objects are *slices* of the three-dimensional arrays.

DEFINITION 2.5. *If $\mathcal{A} = [a_{ijk}]$ is a $n_1 \times n_2 \times n_3$ array, then its kth slice by the third index*[1] *is a $n_1 \times n_2$ matrix A_k with elements*

$$(A_k)_{ij} = a_{ijk}.$$

The "short" vectors along the modes i, j, and k will be referred to as *columns, rows,* and *fibers*, respectively.

[1] The slices by other two indices can also be defined, but it may lead to ambiguity. Since in this paper only kth slices are used, we omit these definitions here.

3. Existence theory. Suppose an $n_1 \times n_2 \times n_3$ tensor $\mathcal{A} = [a_{ijk}]$ is given and there exists a rank-(r_1, r_2, r_3) Tucker approximation to $\mathcal{A}$ with the accuracy ε:

$$\mathcal{A} = \mathcal{G} \times_i U \times_j V \times_k W + \mathcal{E}, \quad ||\mathcal{E}|| = \varepsilon. \tag{3.1}$$

If we are aware that such an approximation exists, then a generally different approximation of the same type with the accuracy bound $c\varepsilon$ (where $c > 1$ is a *deterioration coefficient*) can be constructed from the knowledge of roughly the same amount of entries as those explicitly involved in (3.1). We want to prove this together with a bound on the deterioration coefficient c depending only upon dimensions and ranks but not on the entries of the array.

THEOREM 3.1. *Suppose* (3.1) *holds for some* U, V, *and* W *and* $\mathcal{G}$. *Then there exist matrices* U', V', *and* W' *of sizes* $n_1 \times r_1$, $n_2 \times r_2$, *and* $n_3 \times r_3$ *and consisting of some* r_1 *columns,* r_2 *rows, and* r_3 *fibers, of* $\mathcal{A}$, *respectively, and there exists a tensor* $\mathcal{G}'$ *such that*

$$\mathcal{A} = \mathcal{G}' \times_i U' \times_j V' \times_k W' + \mathcal{E}',$$

where

$$||\mathcal{E}'||_C \le (r_1 r_2 r_3 + 2r_1 r_2 + 2r_1 + 1)\varepsilon.$$

Proof. Consider an unfolding matrix $A^{(1)}$ of the array $\mathcal{A}$. Since $\mathcal{A}$ has a rank-(r_1, r_2, r_3) approximation with accuracy ε, it is easy to see from (3.1) that $A^{(1)}$ has a rank-r_1 approximation with the same accuracy. A low-rank matrix can be approximated by its *skeleton decomposition*

$$A^{(1)} = C\hat{C}^{-1}B^{\top} + \mathcal{E}_1,$$

where C is a $n_1 \times r_1$ matrix containing some r_1 columns of $A^{(1)}$, B is a $n_2 n_3 \times r_1$ matrix containing some r_1 rows of $A^{(1)}$, and $\hat{C}$ is a submatrix on the intersection of these rows and columns. In [13] it was proved that if $\hat{C}$ is a *submatrix of maximal volume* (that is, the $r_1 \times r_1$ submatrix which has the largest absolute value of the determinant) in $\mathcal{A}^{(1)}$, then ε_1 is bounded as follows:

$$||\mathcal{E}_1||_C \le (r_1 + 1)\varepsilon,$$

where $||\cdot||_C$ denotes the largest magnitude element of a matrix (array). Also, if $\hat{C}$ is a maximal volume submatrix, it is easy to prove (cf. [13]) that the elements of $C\hat{C}^{-1}$ are not greater than 1 in modulus. Consequently,

$$\left| a_{ijk} - \sum_{s=1}^{r_1} \gamma_{is} z_{jks} \right| \le (r_1 + 1)\varepsilon,$$

where

$$|\gamma_{is}| \le 1, \qquad 1 \le i \le n_1,$$

and z_{jks} are in a one-to-one correspondence with the entries of $B^{\top}$ (for a fixed s these elements present a slice by the index i of the array $\mathcal{A}$). Also note that

$$\gamma_{is} = \sum_{l=1}^{r_1} u_{il}\phi_{ls},$$

where the matrix $[u_{il}]$ consists of some columns of $\mathcal{A}$.

Now let us look more closely at the matrix $B^\top$. In a reshaped form, it becomes the tensor with the elements z_{jks}. As previously, unfold this tensor along the index j. The ε-rank[2] of the unfolding matrix does not exceed r_2.

Again using the result of [13], we obtain the following inequalities:

$$\left| z_{jks} - \sum_{t=1}^{r_2} \sum_{\tau=1}^{r_2} \psi_{t\tau} v_{jt} w_{ks\tau} \right| \le (r_2 + 1)\varepsilon,$$

where the arrays $[v_{jt}]$ and $[w_{ks\tau}]$ consist of some rows and fibers of $\mathcal{A}$.

Unfolding the array $[w_{ks\tau}]$ by the index k, we observe that the ε-rank of the unfolding matrix cannot be larger than k_3. Hence, this matrix admits the skeleton approximation with the error bound

$$\left| w_{ks\tau} - \sum_{\alpha=1}^{k_3} \sum_{\beta=1}^{k_3} x_{k\alpha} \zeta_{\alpha\beta} y_{\alpha s\tau} \right| \le (r_3 + 1)\varepsilon.$$

Finally,

$$\begin{aligned}
\left| a_{ijk} - \sum_{l=1}^{r_1} \sum_{s=1}^{r_1} \sum_{t=1}^{r_2} \sum_{\tau=1}^{r_2} \sum_{\alpha=1}^{r_3} \sum_{\beta=1}^{r_3} (\phi_{ls} \psi_{t\tau} \zeta_{\alpha\beta} y_{\alpha t\tau}) u_{il} v_{jt} x_{k\alpha} \right| &\le \left| a_{ijk} - \sum_{s=1}^{r_1} \gamma_{is} z_{jks} \right| \\
+ \sum_{s=1}^{r_1} \left| z_{jks} - \sum_{t=1}^{r_2} \sum_{\tau=1}^{r_2} \psi_{t\tau} v_{jt} w_{ks\tau} \right| &+ \sum_{s=1}^{r_1} \sum_{\tau=1}^{r_2} \left| w_{ks\tau} - \sum_{\alpha=1}^{k_3} \sum_{\beta=1}^{k_3} x_{k\alpha} \zeta_{\alpha\beta} y_{\alpha s\tau} \right| \\
&\le (r_1 + 1)\varepsilon + r_1(r_2 + 1)\varepsilon + r_1 r_2 (r_3 + 1)\varepsilon,
\end{aligned}$$

which completes the proof. ∎

If $r_1 = r_2 = r_3 = r$, then the error bound becomes $(r+1)(r^2+r+1)\varepsilon \le (r+1)^3\varepsilon$. In the general case, we are not completely satisfied with the error bound of this theorem because it is not a symmetric function of r_1, r_2, r_3. Of course, the answer can be formally symmetrized, using different permutations of modes (for example, $n_3 \times n_2 \times n_1$) and taking the minimum of all these error bounds, but the obtained result seems to be rather artificial. So here we note that a "truly symmetric" version of this theorem is likely to need a different technique.

COROLLARY 3.2. *Under the premises of the theorem,*

$$||\mathcal{E}'||_F \le (r_1 r_2 r_3 + 2r_1 r_2 + 2r_1 + 1)\sqrt{n_1 n_2 n_3}\varepsilon.$$

4. The cross approximation method. For presentation purposes from now on we will assume that $n_1 = n_2 = n_3 = n$ and $r_1 = r_2 = r_3 = r$.

4.1. The two-dimensional-cross method. In the works [13, 14, 17] the problem of finding a rank-r approximation to a given matrix was connected with finding in matrix A a *submatrix of maximal volume* (that is, determinant in modulus) among all $r \times r$ submatrices. The latter problem is hard to solve. However, we may be satisfied with a "sufficiently good" submatrix and some heuristic algorithms. Since these algorithms are to fetch a cross of some columns and rows, we call them *cross*

[2]The matrix A is said to have ε-rank r if there exists a rank-r matrix B such that $\|A - B\| \le \varepsilon$.

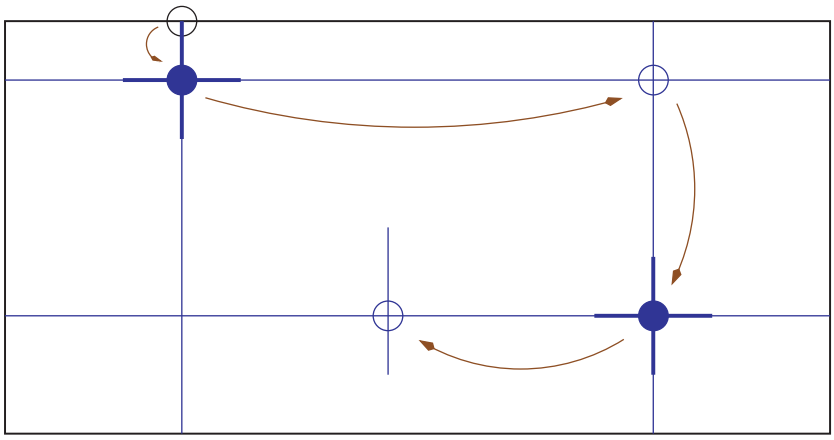


FIG. 1. *How a cross method works. Filled dots: elements used for the calculation of cross*(i_p, j_p). *Empty dots: row–pivot, step* (2).

algorithms. Probably the most simple and effective cross algorithm is the Gauss elimination method using some pivoting technique over dynamically selected sets of the entries of the "active matrix" (for a general description, see [9]). We will use here the column and row pivoting considered in [8]. This method is simple but may breakdown (quit when a good approximation is not obtained) if applied as it is. A cheap practical remedy proposed in [16] is a restarted version of this cross method. For the readers convenience, we give here a brief description of the algorithm.

ALGORITHM 1 (CROSS2D). Given a matrix A of approximate rank r, approximate it by a matrix $\tilde{A}_r$, which is a sum of r rank-1 matrices $u_p v_p^\top$ (so-called skeletons). The principle scheme is given in Figure 1.

(0) Numbering the steps by p, set $p = 1$. Choose some column in A, and assign its index to j_p.
(1) Calculate column j_p of the matrix A, and subtract from all elements the corresponding elements of already calculated skeletons. In the resulting vector find the largest magnitude element. Suppose it is located in the row i_p.
(2) Calculate the row i_p of the residue and the next pivot which is its largest magnitude element with a restriction that the element from the j_pth column cannot be chosen again (see Figure 1). Suppose this pivot is located in the j_{p+1}th column.
(3) Calculate the new cross centered at (i_p, j_p).
(4) If a stopping criterion is not satisfied, set $p := p + 1$, and go to step (1).

The approximation $\tilde{A}_p = \sum_{\alpha=1}^{p} u_\alpha v_\alpha^\top$ is considered good if

$$\|A - \tilde{A}_p\| \le \varepsilon \|A\|_F \approx \varepsilon \|\tilde{A}_p\|_F.$$

However, the exact computation of the error requires all matrix elements and n^2 operations, which is unacceptable. At the same time, the norm $\|\tilde{A}_p\|_F$ can be computed via the formula

$$\|\tilde{A}_p\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\sum_{\alpha=1}^{r} u_{i\alpha} v_{j\alpha}\right)^2 = \sum_{\alpha=1}^{r}\sum_{\alpha'=1}^{r}(u_\alpha, u_{\alpha'})(v_\alpha, v_{\alpha'})$$

using $\mathcal{O}(p^2 n)$ operations. And as a practical estimator of the error (stopping criterion),

we use the norm of a newly computed rank-1 correction. Specifically, we stop if

$$(n-p)\|u_p\|_2\|v_p\|_2 \le \varepsilon\|\tilde{A}_r\|_F.$$

The number $n-p$ is a heuristic constant. Note that after p steps of the cross algorithm exactly p rows and columns of the residue are zeroed, so if we assume that the error is "equally distributed" among the remaining $n-p$ rows, then we immediately arrive at the presented stopping criteria.

Such version of the cross method requires $2rn$ evaluations of matrix elements and $\mathcal{O}(r^2n)$ additional operations (the reason for counting the number of element computations is that the calculation of one element may be a very time-consuming operation). Even if the stopping criteria is satisfied, in some cases the obtained approximation is not good enough (but this does not happen very often). To make the method more robust, the *restart* step is performed: we create a sample from the elements of the residue matrix $A-\tilde{A}_r$. If the error estimated from that sample is large, we proceed with step (3) using the largest magnitude element in the sample as a pivot.

4.2. Towards the three-dimensional-cross method. Consider the unfoldings of the array $\mathcal{A}$ (rectangular matrices of sizes $n\times n^2$ defined by (2.1)), and apply to them the cross approximation algorithm. If the array $\mathcal{A}$ possesses a good Tucker rank-(r,r,r) approximation, then there exist rank-r approximations for the unfoldings $A^{(1)}$, $A^{(2)}$, and $A^{(3)}$ which are also good:

$$\tilde{A}^{(1)}_r = U\Psi^\top, \qquad \tilde{A}^{(2)}_r = V\Phi^\top, \qquad \tilde{A}^{(3)}_r = W\Upsilon^\top,$$

where U,V,W are $n\times r$ matrices with *orthonormal columns* and matrices Ψ,Φ,Υ are $n^2\times r$. The Tucker core is calculated by the convolution of the form (1.3) with a_{ijk} being replaced with their approximate values. For example, using the decomposition by the first direction, $\tilde{A}^{(1)}_r = U\Psi^\top$, we have

$$a_{ijk} \approx \tilde{a}_{ijk} = \sum_{\alpha=1}^{r} u_{i\alpha}\psi_{jk\alpha},$$

where the rows of the matrix $\Psi=[\psi_{(jk)\alpha}]$ are numbered by a pair of indices (jk). Substituting this into (1.3), we obtain

$$\begin{aligned} g_{i'j'k'} &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left(\sum_{\alpha=1}^{r} u_{i\alpha}\psi_{jk\alpha}\right)u_{ii'}v_{jj'}w_{kk'} \\ &= \sum_{\alpha=1}^{r}(u_\alpha,u_{i'})\left(\sum_{j=1}^{n}\sum_{k=1}^{n} v_{jj'}w_{kk'}\psi_{jk\alpha}\right) \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n} v_{jj'}w_{kk'}\psi_{jki'}. \end{aligned} \tag{4.1}$$

This computation needs $\mathcal{O}(n^2r)$ evaluations of the elements of $\mathcal{A}$ plus $\mathcal{O}(n^2r^2)$ operations.

Of course, $\mathcal{O}(n^2)$ is much smaller than the total number of elements in the array $\mathcal{A}$, but it is still too large when n is about 10^3.

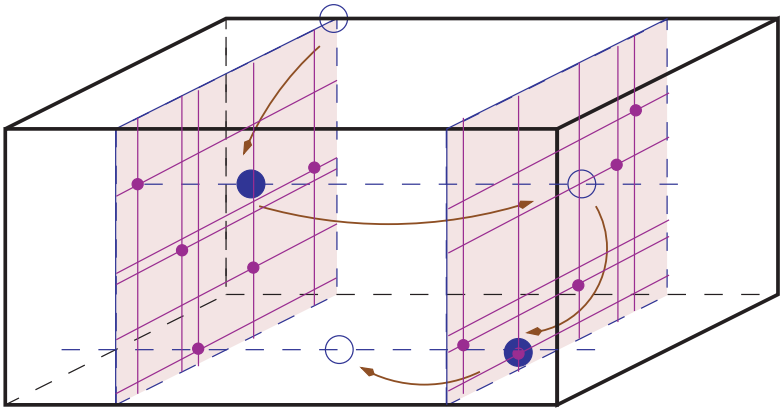


FIG. 2. *The work of the three-dimensional-cross method. The big filled and empty dots correspond to elements for the "outer" cross algorithm, and small dots show elements used for the "inner" cross algorithm, approximating a particular two-dimensional slice*

4.3. How to achieve linear complexity. We want to achieve linear complexity in n. To this end, we have to get rid of the computation of all elements in the slices of $\mathcal{A}$ used in the unfoldings (2.1) (then we avoid n^2-long vectors). We suggest to approximate the slices by the same cross algorithm developed for matrices. Since $\mathcal{A}$ has a good Tucker approximation with the accuracy ε, each slice $A_k = [(a_k)_{ij}]$ can be accurately approximated by a rank-r matrix. In what follows, we will never store a slice as a full $n \times n$ matrix and will never refer to all its elements. Instead, we deal only with some low-rank approximations for the slices.

ALGORITHM 2. Given an $n \times n \times n$ array $\mathcal{A}$, take one of the indices i, j, k as the "leading index," let it be k. Then consider the corresponding unfolding matrix of size $n \times n^2$, and approximate it applying the cross method. The columns of the unfolding matrix are calculated as usual, but each of the long rows is considered as a matrix of size $n \times n$ to be approximated by the same cross method. Therefore, at each step of the "outer" cross method we add one skeleton of form $A \otimes w$ with the long vector A represented by the sum of r skeletons in the "inner" cross method; therefore, we add a tensor of form $\sum_{q=1}^{r} u_q \otimes v_q \otimes w$. After r steps the approximant has the form

$$\tilde{A} = \sum_{p=1}^{r} \left(\sum_{q=1}^{r} u_q^{\{p\}} \otimes v_q^{\{p\}} \right) \otimes w^{\{p\}},$$

where $u_q^{\{p\}}$ and $v_q^{\{p\}}$, $q = 1, \ldots, r$, are vectors comprising the skeleton approximation to the long vector at the step with number p. Note that unlike vectors u and v, vectors w are numbered only by p, because k is the leading index. The principle scheme is given in Figure 2.

The expected number of arithmetic operations is *almost linear* in sizes of the array: the complexity is $\mathcal{O}(nr^d)$ operations for some small $d > 0$.

(0) Numbering the steps by p, set p to 1. Choose a slice $A_k = [(a_k)_{ij}]$ in $\mathcal{A}$, for example, and assume its index to k_1. Set $\tilde{\mathcal{A}} = 0$.

(1a) Find an approximation A_{k_p} to the k_pth slice of the residue $\mathcal{R} = \mathcal{A} - \tilde{\mathcal{A}}_p$ by the cross method:

$$A_{k_p} = \sum_{q=1}^{r} u_q v_q^{\top}.$$

(1b) Find the largest magnitude element in the matrix A_{k_p}; let it be located at (i_p, j_p).

(2) Compute the vector w corresponding to the fiber of $\mathcal{R}$ with index (i_p, j_p),

$$w_k = \mathcal{R}_{i_p, j_p, k},$$

perform the scaling

$$w := w / w_{k_p},$$

and find in w the largest magnitude element from those whose index is not equal to k_p.[3] Suppose it is located at the k_{p+1}th position of w.

(3) Compute a new approximation:

$$\tilde{\mathcal{A}} := \tilde{\mathcal{A}} + A_{k_p} \otimes w = \tilde{\mathcal{A}} + \left(\sum_{q=1}^{r} u_q v_q^\top \right) \otimes w = \tilde{\mathcal{A}} + \sum_{q=1}^{r} u_q \otimes v_q \otimes w.$$

(4) If the stopping criterion is not satisfied, set $p := p + 1$, and go to step (1).

In the end, the array $\mathcal{A}$ is approximated by $\tilde{\mathcal{A}} = [\tilde{a}_{ijk}]$ having a Tucker-like decomposition (also viewed as a trilinear, PARAFAC, or CANDECOMP decomposition [1, 5]) of the form (in elementwise notation)

$$\tilde{a}_{ijk} = \sum_{p=1}^{r} \left(\sum_{q=1}^{r} u_{iq}^{\{p\}} v_{jq}^{\{p\}} \right) w_k^{\{p\}} = \sum_{\alpha=1}^{r^2} u_{i\alpha} v_{j\alpha} w_{k\alpha}, \tag{4.2}$$

where in the second sum for simplicity all terms are numbered by a single index α instead of a complicated sum on the left.

During the implementation of this method, we encounter several problems that should be solved with a linear complexity in n:

- determine the largest magnitude element in a low-rank matrix, step (1)b;
- estimate the quantities in the relationships

$$\|\mathcal{A} - \tilde{\mathcal{A}}\|_F \le \varepsilon \|\mathcal{A}\|_F \approx \varepsilon \|\tilde{\mathcal{A}}\|_F$$

so as to have a sound stopping criterion, step (4).

The first problem is not trivial, and we do not know if there is an exact and fast way to find a maximal element in a low-rank matrix. However, we are able to design a heuristic algorithm, based on the submatrix of maximal volume. It manifests a very good practical performance (see Appendix A).

The stopping criterion in the Cross3D method is identical to the two-dimensional case, by the comparison of the approximant norm and the norm of a newly computed cross-correction. The norm $\|\tilde{\mathcal{A}}\|_F$ is computed by the formulas

$$\begin{aligned} \|\tilde{\mathcal{A}}\|_F^2 &= \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \left(\sum_{\alpha=1}^{r^2} u_{i\alpha} v_{j\alpha} w_{k\alpha} \right)^2 \\ &= \sum_{\alpha=1}^{r^2} \sum_{\alpha'=1}^{r^2} (u_\alpha, u_{\alpha'})(v_\alpha, v_{\alpha'})(w_\alpha, w_{\alpha'}). \end{aligned}$$

[3] It is worthy to note that we cannot also use elements with indices $k_1, \ldots, k_{p-1}$, but it can be verified that they are all zeroes, so they cannot have maximal modulus.

The cost of Algorithm 2 is $\mathcal{O}(nr^2)$ evaluations of the elements of $\mathcal{A}$ plus $\mathcal{O}(nr^4)$ arithmetic operations (at each outer step of the method we compute a new slice from which we should subtract the elements of the previously computed approximation, and that results in a relatively big constant r^4 at the size n). This is already a linear complexity. However, we are going to present a "clever" implementation with a significantly better performance.

4.4. The three-dimensional-cross algorithm. We can improve the efficiency of Algorithm 2 by using a more compact way to store and handle the slices A_{k_p} so that the required number of vectors to represent them is reduced from $\mathcal{O}(r^2)$ to $\mathcal{O}(r)$.

At each step we approximate the computed slices A_{k_p} in the format

$$A_{k_p} = UB_pV^\top, \tag{4.3}$$

where $n \times r$ matrices U, V are orthogonal and the core matrices B_p are $r \times r$. It is worth noting that (4.3) is also known as a Tucker2 decomposition, where only 2 of 3 modes are compressed. The storage for p slices is now $2nr+pr^2$ which is asymptotically equal to $\mathcal{O}(nr)$. The existence of matrices U, V, follows from the existence of a "good" Tucker approximation. In fact, we can try U and V as the Tucker factors. The computation of this simultaneous matrix decomposition is equivalent to the computation of the Tucker decomposition of a $n \times n \times p$ array:

$$\mathcal{A}' = [A_{k_1} \dots A_{k_p}].$$

Indeed, if

$$a_{ijk_p} = \sum_{i'=1}^{r} \sum_{j'=1}^{r} \sum_{p'=1}^{r} g_{i'j'p'} u_{ii'} v_{jj'} w_{pp'},$$

then, setting

$$\sum_{p'=1}^{r} g_{i'j'p'} w_{pp'} = b_{i'j'p} = (b_p)_{i'j'},$$

we immediately arrive at (4.3).

Another important modification concerns the computation of the slices. Suppose the p steps are done and we are going to compute the $p+1$th slice $A_{k_{p+1}}$. Instead of using the "full" cross method for this slice, we first find an approximation of the form

$$A_{k_{p+1}} \approx U\Phi V^\top,$$

where U, V come from (4.3) and a matrix Φ is $r \times r$. Such an approximation can be obtained quite cheaply by the following scheme:

- Find $r \times r$ *submatrices of maximal volume* in U and V. Suppose they have indices $i_1, \dots, i_r$ and $j_1, \dots, j_r$. Denote these submatrices by $\hat{U}$ and $\hat{V}$.
- Compute the $r \times r$ submatrix S in $A_{k_{p+1}}$ lying on the intersection of rows with indices $i_1, \dots, i_r$ and columns with indices $j_1, \dots j_r$.
- Compute

$$\Phi = \hat{U}^{-1} S \hat{V}^{-1}. \tag{4.4}$$

We can prove that this approximation approach is robust (see Appendix B). After Φ is computed, we check the approximation error by taking some random samples of a true matrix A_{k_p}. If the approximation is not good enough, then we perform some steps of the cross approximation algorithm, starting from a good approximation. However, as a rule, only a few steps (or even none) of the cross algorithm are required.

ALGORITHM 3 (CROSS3D). Suppose an $n \times n \times n$ three-way array $\mathcal{A}$ is given.

(0) Perform one step of Algorithm 2 (with $p = 1$). Upon completion, $p = 2$, and $\mathcal{A}$ is represented as

$$\tilde{A} = \left(\sum_{q=1}^{r} u_q^{\{1\}} (v_q^{\{1\}})^\top \right) \otimes w^{\{1\}} = \sum_{q=1}^{r} u_q^{\{1\}} \otimes v_q^{\{1\}} \otimes w^{\{1\}}.$$

Form matrices $U_1 = [u_q^{\{1\}}]$, $V_1 = [v_q^{\{1\}}]$, and compute orthonormal bases U, V of these subspaces using two QR-decompositions:

$$U_1 = UR_u, \qquad V_1 = VR_v.$$

Then $\mathcal{A}$ is represented as

$$\tilde{A} = (UR_u R_v^\top V^\top) \otimes w^{\{1\}} = (UB_1 V^\top) \otimes (w^{\{1\}}/\|w^{\{1\}}\|_2),$$

$$B_1 = R_u R_v^\top \|w^{\{1\}}\|_2.$$

Set $w^{\{1\}} := w^{\{1\}}/\|w^{\{1\}}\|$. (Note that in this algorithm we will keep bases U, V, and W orthonormal.) In the vector $w^{\{1\}}$ compute the largest magnitude element; suppose it has index k_2.

(1.1) Compute Φ from (4.4). If necessary, perform some additional steps of the cross method to obtain an approximation to the slice A_{k_p}:

$$A_{k_p} \approx \tilde{A}_{k_p} = U\Phi V^\top + \sum_{q=1}^{r_p} u_q^{\{p\}} (v_q^{\{p\}})^\top.$$

(Note that r_p is supposed to be small even compared to r.)

(1.2) Add new vectors $U_p = [u_q^{\{p\}}]$, $V_p = [v_q^{\{p\}}]$, $q = 1, \ldots, r$, to bases U, V, and orthogonalize the extended matrices $[UU_p]$, $[VV_p]$:

$$[UU_p] = [U\hat{U}_p] \begin{bmatrix} I & M_u \\ 0 & R_u \end{bmatrix}, \quad [VV_p] = [V\hat{V}_p] \begin{bmatrix} I & M_v \\ 0 & R_v \end{bmatrix},$$

$$U^\top \hat{U}_p = 0, \qquad V^\top \hat{V}_p = 0, \qquad \hat{U}_p^\top \hat{U}_p = I, \qquad \hat{V}_p^\top \hat{V}_p = I.$$

(1.3) Compute new $(r + r_p) \times (r + r_p)$ core B_p:

$$B_p = \begin{bmatrix} M_u \\ R_u \end{bmatrix} \begin{bmatrix} M_v^\top R_v^\top \end{bmatrix}.$$

Other slices in a new basis have the form

$$A_{k_q} = [U\hat{U}_p] \begin{bmatrix} B_q & 0 \\ 0 & 0 \end{bmatrix} [V\hat{V}_p]^\top, \qquad q = 1, \ldots, p-1.$$

Therefore, approximation $\tilde{\mathcal{A}}_{p-1}$ is represented as

$$\tilde{\mathcal{A}}_{p-1} = \sum_{q=1}^{p-1} \left(U' B'_q V'^{\top}\right) \otimes w_q,$$

$$U' = [U\hat{U}_p], \qquad V' = [V\hat{V}_p], \qquad B'_q = \begin{bmatrix} B_q & 0 \\ 0 & 0 \end{bmatrix}, \qquad q = 1, \ldots, p-1.$$

Set also $B'_p = B_p$.

(1.4) In the new slice A_{k_p} the largest magnitude element is found. Suppose it is located in (i_p, j_p).

(2.1) The fiber $w^{\{p\}}$ of the residue $\mathcal{A} - \tilde{\mathcal{A}}_{p-1}$ corresponding to (i_p, j_p) is computed.

(2.2) Vector $w^{\{p\}}$ is orthogonalized to vectors $W = [w^{\{1\}}, \ldots, w^{\{p-1\}}]$:

$$w^{\{p\}} = \sum_{q=1}^{p-1} c_q w^{\{q\}} + \hat{w}^{\{p\}}, \qquad (\omega^{\{q\}})^{\top} \hat{w}^{\{p\}} = 0, \qquad q = 1, \ldots, p-1.$$

Cores of old slices B'_q, $q = 1, \ldots, p-1$, are modified:

$$B''_q = B'_q + c_q B'_p, \qquad q = 1, \ldots, p-1.$$

Vector $\hat{w}^{\{p\}}$ is normalized:

$$w^{\{p\}} := \hat{w}^{\{p\}} / \|\hat{w}^{\{p\}}\|_2, \qquad B''_p = \|\hat{w}^{\{p\}}\|_2 B'_p.$$

(3) The approximation $\tilde{\mathcal{A}}_p$ is represented as

$$\tilde{\mathcal{A}}_p = \sum_{q=1}^{p} \left(U' B''_q V'^{\top}\right) \otimes w^{\{q\}}. \tag{4.5}$$

To reduce the sizes of the $(r+r_p) \times (r+r_p)$ matrices B''_q, we apply the Tucker reduction method.

(3.1) Create a three-way array $\mathcal{B}'' = [B''_1 \ldots B''_p]$ of size $(r+r_p) \times (r+r_p) \times p$, and compute its Tucker decomposition.

$$b''_{ijk} = \sum_{i'=1}^{r} \sum_{j'=1}^{r} \sum_{k'=1}^{r} g^{\clubsuit}_{i'j'k'} u^{\clubsuit}_{ii'} v^{\clubsuit}_{jj'} w^{\clubsuit}_{kk'}.$$

If we introduce matrices $B_{\clubsuit_k}$ with elements

$$(B_{\clubsuit_k})_{i'j'} = \sum_{k'=1}^{r} g^{\clubsuit}_{i'j'k'} w^{\clubsuit}_{kk'},$$

then we have

$$B''_k = U_{\clubsuit} B_{\clubsuit_k} V_{\clubsuit}^{\top}, \qquad k = 1, \ldots, p, \tag{4.6}$$

where matrices $U_{\clubsuit}$ and $V_{\clubsuit}$ are $(r+r_1) \times r$ and cores $B_{\clubsuit_k}$ are $r \times r$.

(3.2) Substituting (4.6) into (4.5), we obtain that

$$\tilde{\mathcal{A}}_p = \sum_{k=1}^{p} \Big((U' U_{\clubsuit}) B_{\clubsuit k} (V' V_{\clubsuit})^\top \Big) \otimes w^{\{k\}} = \sum_{k=1}^{p} \big(U B_k V^\top \big) \otimes w^{\{k\}}, \tag{4.7}$$

where $U = U' U_{\clubsuit}$, $V = V' V_{\clubsuit}$ are orthogonal and cores $B_k = B_{\clubsuit k}$ are $r \times r$. The format (4.3) is restored.

(4) Check stopping criteria; if it is not satisfied, go to (1.1).

At each step of the outer cross method we obtain a new Tucker approximant with the core of the dimension $(r + r_p) \times (r + r_p) \times r$; the purpose of steps (3.1)–(3.2) is to compress this tensor back.

The Algorithm 3 is the final version of Cross3D. The numerical complexity of the method is $\mathcal{O}(nr)$ evaluations of the array elements and $\mathcal{O}(nr^3)$ additional operations.

5. Numerical experiments. We illustrate the performance of our algorithm on some model tensors which allow good low-rank approximation.

Specifically, we consider the following two types of arrays:

$$\mathcal{A} = [a_{ijk}], \qquad a_{ijk} = \frac{1}{i + j + k}, \qquad 1 \le i, j, k \le n,$$

$$\mathcal{B} = [b_{ijk}], \qquad b_{ijk} = \frac{1}{\sqrt{i^2 + j^2 + k^2}}, \qquad 1 \le i, j, k \le n.$$

The rank estimates obtained in [15, 18, 19] have the form

$$r \le C(\log n \log^2 \varepsilon),$$

where ε is an error of the approximation, so the rank grows only logarithmically with n and ε.

These two examples arise from the numerical solution of integral equations. For example, the array $\mathcal{B}$ is obtained from the integral equation with kernel $\frac{1}{||x-y||}$ acting on a unit cube and being disretized by the Nyström method on a uniform grid.

Table 1 shows the ranks, accuracies, and size of the computed Tucker approximation for the array $\mathcal{A}$; Table 2 shows the same for $\mathcal{B}$. The accuracy of the approximation was computed by sampling the elements of the array, since it is not possible to check all the elements for large n. The size of the sample was determined by the following rule: if the sample size was doubled, the estimated error should change by no more than 10%. As it can be seen, the approximation method is robust and leads to astonishing memory savings: the arrays that would need in the full format an enormous storage of 2 petabytes ($2 \cdot 2^{50}$ PB) are compressed to the sizes of 100 MB.

Moreover, our algorithm works with arrays on this huge scale on a personal workstation. The timings made on a personal computer (`Pentium-4` with 3.4 Ghz clock) are shown in Figure 3. This figure confirms that the approximation time is almost linear in n. More precisely, we demonstrate the $\mathcal{O}(nr^3)$ complexity of the Cross3D algorithm with rank estimated as $r \sim \log n$ for fixed ε. Therefore, the complexity of the algorithm is estimated as

$$t \le cn \log^3 n.$$

In Figure 3 real timings, measured for different ε, are shown together with "theoretical" bounds $cn \log^3 n$, plotted for two different values of c. The somewhat irregular

TABLE 1
Numerical results

$$\mathcal{A} = [a_{ijk}], \qquad a_{ijk} = \frac{1}{i+j+k}, \qquad 1 \le i,j,k \le n.$$

Rank and accuracy of the decomposition.

ε	1.$_{10}-3$		1.$_{10}-5$		1.$_{10}-7$		1.$_{10}-9$	
n	r	err	r	err	r	err	r	err
64	5	$2.57_{10}-4$	8	$2.33_{10}-6$	10	$1.46_{10}-8$	12	$3.09_{10}-10$
128	6	$6.86_{10}-4$	8	$4.47_{10}-6$	11	$3.19_{10}-8$	13	$6.4_{10}-10$
256	6	$8.84_{10}-4$	9	$3.97_{10}-6$	12	$5.49_{10}-8$	15	$3.03_{10}-10$
512	7	$7.49_{10}-4$	10	$1.41_{10}-6$	13	$6.84_{10}-8$	16	$5.2_{10}-10$
1024	7	$5.71_{10}-4$	11	$4.83_{10}-6$	14	$4.09_{10}-8$	18	$2.74_{10}-10$
2048	7	$6.63_{10}-4$	12	$2.08_{10}-6$	16	$4.01_{10}-8$	19	$3.97_{10}-10$
4096	8	$3.23_{10}-4$	12	$6.32_{10}-6$	17	$3.44_{10}-8$	21	$3.47_{10}-10$
8192	8	$6.36_{10}-4$	13	$3.36_{10}-6$	18	$1.93_{10}-8$	22	$4.56_{10}-10$
16384	9	$7.95_{10}-4$	14	$3.52_{10}-6$	19	$7.21_{10}-8$	24	$5.64_{10}-10$
32768	9	$6.4_{10}-4$	14	$8.86_{10}-6$	20	$5.27_{10}-8$	25	$3.79_{10}-10$
65536	9	$4.07_{10}-4$	15	$6.31_{10}-6$	21	$2.51_{10}-8$	26	$5.25_{10}-10$

Rank and size (MB) of the Tucker decomposition.
The sizes smaller than 1MB are not shown.

ε		1.$_{10}-3$		1.$_{10}-5$		1.$_{10}-7$		1.$_{10}-9$	
n	full	r	mem	r	mem	r	mem	r	mem
64	2MB	5		8		10		12	
128	16MB	6		8		11		13	
256	128MB	6		9		12		15	
512	1GB	7		10		13		16	
1024	8GB	7		11		14		18	
2048	64GB	7		12		16		19	
4096	512GB	8	< 1	12	1.1	17	1.6	21	2
8192	4TB	8	1.5	13	2.5	18	3.5	22	4.2
16384	32TB	9	3.4	14	5.25	19	7.2	24	9
32768	256TB	9	6.75	14	10.5	20	15	25	19
65536	2PB	9	13.5	15	22	21	31	26	39

behavior on the timing plots is caused by the effects of caching (for small n) and by some rank overestimation by the stopping criteria for large n.

Two dense tensors considered come from a simple disretization of integral equations. Despite their "regularity" they are quite representative: in more complex cases our method behaves similarly. In other areas where tensor decomposition is used, the researchers often obtain more irregular and possibly sparse tensors. We want to note that sparseness is *ideal* for Cross3D because in that case the residue can be measured *exactly* and the pivots during the cross approximation stage can be also found exactly, leading to a *theoretically robust* method. The applications of the three-dimensional-cross approach to more complex tensors will be reported elsewhere.

Appendix A. How to find the maximal element in a slice. One of the important ingredients of the three-dimensional-cross method is the determination of the maximal element in a given low-rank matrix in linear time.

Suppose we have computed a skeleton approximation to a low-rank matrix

$$A = UV^{\top},$$

where U, V are $n \times r$, and we want to find the largest magnitude element in it. This

TABLE 2
Numerical results

$$\mathcal{B} = [b_{ijk}], \qquad b_{ijk} = \frac{1}{\sqrt{i^2 + j^2 + k^2}}, \qquad 1 \le i, j, k \le n.$$

Rank and accuracy of the Tucker decomposition.

ε	$1.10{-}3$		$1.10{-}5$		$1.10{-}7$		$1.10{-}9$	
n	r	err	r	err	r	err	r	err
64	7	$3.77_{10}{-}4$	11	$3.91_{10}{-}6$	14	$5.7_{10}{-}8$	18	$2.21_{10}{-}10$
128	8	$5.19_{10}{-}4$	12	$5.92_{10}{-}6$	17	$2._{10}{-}8$	20	$5.63_{10}{-}10$
256	9	$4.11_{10}{-}4$	14	$6.4_{10}{-}6$	19	$3.46_{10}{-}8$	23	$4.5_{10}{-}10$
512	10	$4.93_{10}{-}4$	15	$6.67_{10}{-}6$	21	$2.92_{10}{-}8$	26	$3.27_{10}{-}10$
1024	10	$5.47_{10}{-}4$	17	$3.21_{10}{-}6$	23	$3.95_{10}{-}8$	29	$4.73_{10}{-}10$
2048	11	$4.98_{10}{-}4$	18	$5.26_{10}{-}6$	25	$6.83_{10}{-}8$	31	$5.94_{10}{-}10$
4096	12	$8.4_{10}{-}4$	19	$4.25_{10}{-}6$	27	$3.56_{10}{-}8$	34	$3.38_{10}{-}10$
8192	12	$6.8_{10}{-}4$	20	$6._{10}{-}6$	28	$5.8_{10}{-}8$	36	$3.66_{10}{-}10$
16384	13	$2.69_{10}{-}4$	22	$4.78_{10}{-}6$	31	$5.65_{10}{-}8$	39	$2.67_{10}{-}10$
32768	13	$8.52_{10}{-}4$	23	$6.09_{10}{-}6$	32	$7.16_{10}{-}8$	41	$5.51_{10}{-}10$
65536	14	$6.27_{10}{-}4$	24	$6.52_{10}{-}6$	34	$7.89_{10}{-}8$	44	$1.41_{10}{-}9$

Rank and size (MB) of the Tucker decomposition.
Values less than 1MB are not shown

ε		$1.10{-}3$		$1.10{-}5$		$1.10{-}7$		$1.10{-}9$	
n	full	r	mem	r	mem	r	mem	r	mem
64	2MB	7		11		14		18	
128	16MB	8		12		17		20	
256	128MB	9		14		19		23	
512	1GB	10		15		21		26	
1024	8GB	10		17		23		29	
2048	64GB	11	< 1	18	< 1	25	1.18	31	1.46
4096	512GB	12	1.15	19	1.78	27	2.54	34	3.2
8192	4TB	12	2.25	20	3.75	28	5.3	36	6.8
16384	32TB	13	4.9	22	8.25	31	11.7	39	14.7
32768	256TB	13	9.75	23	17.25	32	24	41	31
65536	2PB	14	21	24	36	34	51	44	66

problem can be solved by comparing all the elements of the matrix, but it costs $\mathcal{O}(n^2)$ operations. The proposed algorithm is based on the following hypothesis.

Hypothesis. Consider $r \times r$ submatrices in a rank-r matrix A. Let B be a submatrix of maximal volume among all such submatrices. Then

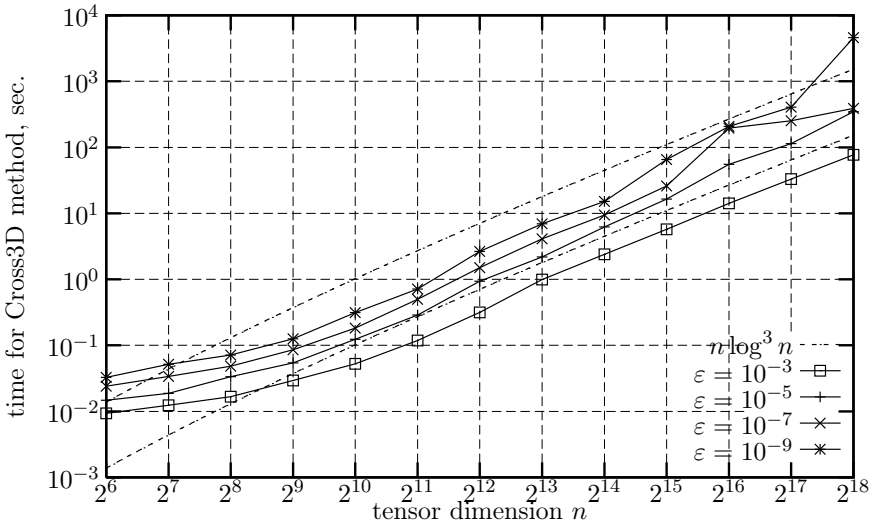
$$||B||_C \ge \frac{||A||_C}{r}.$$

So the maximal element in the submatrix of maximal volume cannot be very much different from the maximal element in the whole matrix A.

How does one determine a submatrix of maximal volume? This submatrix of A lies on the intersection of rows $i_1, \ldots, i_r$, coinciding with rows which contain the submatrix of maximal volume in U, and columns $j_1, \ldots, j_r$, which contain the submatrix of maximal volume in $V^\top$. To find the submatrix of maximal volume in a $n \times r$ matrix, we will use the algorithm, proposed in [12]. For the readers convenience we describe it below.

ALGORITHM 4. Suppose U is a $n \times r$ matrix and its $r \times r$ submatrix with maximal volume is needed.

$$\mathcal{A} = [a_{ijk}], \qquad a_{ijk} = \frac{1}{i+j+k}, \qquad 1 \le i,j,k \le n$$



$$\mathcal{B} = [b_{ijk}], \qquad b_{ijk} = \frac{1}{\sqrt{i^2+j^2+k^2}}, \qquad 1 \le i,j,k \le n$$

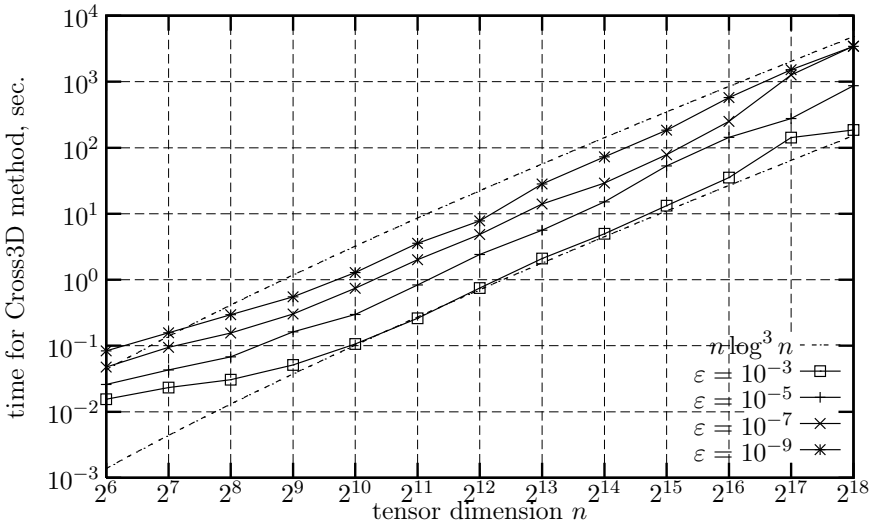


FIG. 3. *Approximation time, sec.*

(0) Let A_γ be a leading submatrix. In the beginning set A_γ to any nonsingular submatrix of A, and permute the rows so that A_γ is located in the first r rows.

(1) Compute

$$AA_\gamma^{-1} = \begin{bmatrix} I_{r\times r} \\ Z \end{bmatrix} = B.$$

(2) Find the largest magnitude element $|z_{ij}|$ in Z.
(3) **If** $\gamma = |z_{ij}| > 1$, **then**
Permute in B rows $r+i$ and j. The upper submatrix in B after the permutation has the form

$$\begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ * & * & \gamma & * & * & \\ & & & \ddots & & \\ & & & & 1 & \end{pmatrix},$$

and its determinant is equal to $\gamma \geq 1+\varepsilon$, that is, it increased. Denote by A_γ the new submatrix in the first r rows of A, and return to step (1).
Otherwise terminate the algorithm.

In practice, to avoid a huge number of transpositions a more "soft" stopping criteria is used in step (3). The algorithm stops if $|z_{ij}| \leq 1+\nu$, where ν is a some small parameter.

Appendix B. The $U\Phi V^\top$ decomposition. In this appendix we will prove that the usage of (4.4) for the construction of the low-rank approximation to a slice is "legal."

THEOREM B.1. *Suppose A is a $n_1 \times n_2$ matrix, U, V are $n_1 \times r_1$ and $n_2 \times r_2$ matrices with orthonormal columns, and there exists a matrix Φ such that*

$$A = U\Phi V^\top + E, \quad ||E||_F \leq \varepsilon.$$

If we compute Φ' by the formula (4.4)*, then*

$$||A - U\Phi' V^\top||_F \leq (\sqrt{n_1 r_1 n_2 r_2} + 1)\varepsilon.$$

Proof. If $\hat{U}$ and $\hat{V}$ are submatrices of maximal volume in U and V, respectively, and $\hat{A}$ is a submatrix in A lying on the intersection of the selected rows from U and columns from $V^\top$, then

$$\hat{A} = \hat{U}\Phi\hat{V}^\top + \hat{E},$$

where $\hat{E}$ is a submatrix of E occupying the same positions in E as $\hat{A}$ in A.

$$||\Phi - \Phi'|| \leq ||\hat{U}^{-1}||\ ||\hat{E}||\ ||\hat{V}^{-1}||.$$

The norms $\hat{U}^{-1}, \hat{V}^{-1}$ can be estimated as follows. We know that the elements of

$$U\hat{U}^{-1}$$

are not greater than 1 in modulus (because $\hat{U}$ is a submatrix of maximal volume). Therefore,

$$||\hat{U}^{-1}||_F \leq \sqrt{n_1 r_1}.$$

Using this estimate we immediately complete the proof as follows:

$$\|A - U\Phi' V^\top\|_F \leq \|U\Phi V^\top - U\Phi' V^\top\|_F + \|E\|_F = \|\Phi - \Phi'\| + \varepsilon \leq \sqrt{n_1 r_1}\sqrt{n_2 r_2}\varepsilon + \varepsilon. \qquad \square$$

Acknowledgment. We are very grateful to both of the referees of our paper. The remark of one of the referees helped us to discover a nasty bug in the program code.

REFERENCES

[1] R. A. Harshman, *Foundations of the Parafac procedure: Models and conditions for an explanatory multimodal factor analysis*, UCLA Working Papers in Phonetics, 16 (1970), pp. 1–84.
[2] Three-Mode Company, http://three-mode.leidenuniv.nl.
[3] P. Comon, *Tensor decomposition: State of the art and applications*, in IMA Conference on Mathematics in Signal Processing, Warwick, UK, http://www.i3s.unice.fr/~comon/FichiersPs/ima2000.ps
[4] J. D. Caroll and J. J. Chang, *Analysis of individual differences in multidimensional scaling via n-way generalization of Eckart–Young decomposition*, Psychometrika, 35 (1970), pp. 283–319.
[5] R. Bro, *PARAFAC: Tutorial and applications*, Chemom. Intelligent Lab. Systems, 38 (1997), pp. 149–171.
[6] G. Beylkin and M. M. Mohlenkamp, *Numerical operator calculus in higher dimensions*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 10246–10251.
[7] M. A. O. Vasilescu and D. Terzopoulos, *Multilinear image analysis for facial recognition*, Proceedings of the International Conference on Pattern Recognition, Quebec City, Canada, 2000, pp. 511–514.
[8] M. Bebendorf, *Approximation of boundary element matrices*, Numer. Math., 86 (2000), pp. 565–589.
[9] J. M. Ford and E. E. Tyrtyshnikov, *Combining Kronecker product approximation with discrete wavelet transforms to solve dense, function-related systems*, SIAM J. Sci. Comput., 25 (2003), pp. 961–981.
[10] L. De Lathauwer, B. De Moor, and J. Vandewalle, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1253–1278.
[11] L. De Lathauwer, B. De Moor, and J. Vandewalle, *On best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of high-order tensors*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1324–1342.
[12] S. A. Goreinov, *Pseudoskeleton approximations of block matrices generated by asymptotically smooth kernels*, Ph.D. Thesis, Institute on Numerical Mathematics, Russian Academy of Sciences, Moscow, Russia, 2001 (in Russian).
[13] S. A. Goreinov and E. E. Tyrtyshnikov, *The maximal-volume concept in approximation by low-rank matrices*, Contemp. Math., 208 (2001). pp. 47–51.
[14] S. A. Goreinov, E. E. Tyrtyshnikov, and N. L. Zamarashkin, *A theory of pseudo-skeleton approximations*, Linear Algebra Appl., 261 (1997), pp. 1–21.
[15] W. Hackbusch, B. N. Khoromskij, and E. E. Tyrtyshnikov, *Hierarchical Kronecker tensor-product approximations*, J. Numer. Math., 13 (2005), pp. 119–156.
[16] D. V. Savostianov, *Mosaic-skeleton approximations*, Master Thesis, Institute on Numerical Mathematics, Russian Academy of Sciences, Moscow, Russia, 2001 (in Russian).
[17] E. E. Tyrtyshnikov, *Incomplete cross approximation in the mosaic-skeleton method*, Computing, 4 (2000), pp. 367–380.
[18] E. E. Tyrtyshnikov, *Kronecker-product approximations for some function-related matrices*, Linear Algebra Appl., 379 (2004), pp. 423–437.
[19] E. E. Tyrtyshnikov, *Tensor approximations of matrices generated by asymptotically smooth functions*, Sb. Math., 194 (2003), pp. 941–954.
[20] I. Ibraghimov, *Application of the three-way decomposition for matrix compression*, Numer. Linear Algebra Appl., 9 (2002), pp. 551–565.
[21] V. Olshevsky, I. V. Oseledets, and E. E. Tyrtyshnikov, *Tensor properties of multilevel Toeplitz and related matrices*, Linear Algebra Appl., 412 (2006), pp. 1–21.
[22] L. R. Tucker, *Some mathematical notes on three-mode factor analysis*, Psychometrika, 31 (1966), pp. 279–311.
[23] P. Drineas, R. Kannan, and M. W. Mahoney, SIAM J. Comput., 36 (2006), pp. 158-183.