

Preservation of Linear Constraints in Approximation of Tensors

Eugene Tyrtyshnikov*

*Institute of Numerical Mathematics of Russian Academy of Sciences, Gubkin Street,
8, Moscow 119333, Russia.*

Received 13 March 2009; Accepted (in revised version) 8 April 2009

Abstract. For an arbitrary tensor (multi-index array) with linear constraints at each direction, it is proved that the factors of any minimal canonical tensor approximation to this tensor satisfy the same linear constraints for the corresponding directions.

AMS subject classifications: 15A12, 65F10, 65F15

Key words: Multi-index arrays, tensors, linear constraints, low rank approximation, canonical tensor decomposition, multilevel matrices.

1. Introduction

Linear constraints define many important classes of structured matrices (Toeplitz, Hankel, various sparse matrices of special patterns etc.). A combination of Toeplitz and tensor structures was first considered in [3]. The common case of linear constraints along with tensor approximations of two-level matrices was first studied in [6]. Some estimates of tensor ranks were suggested in [2, 10, 11]. The interest to tensor approximations in combination with linear constraints is well justified by their role as a base for construction of fast algorithms in difficult cases, a good example is a superfast algorithm for approximate inversion of two-level Toeplitz matrices recently proposed in [7].

A matrix A of order $n = p_1 p_2$ can be viewed as a matrix composed of blocks a_{ij} of size $p_2 \times p_2$, where the indices i, j run from 1 to $p = p_1$. In particular, A can be of the form

$$A = A_r = \sum_{t=1}^r U_t \otimes V_t, \quad (1.1)$$

where U_t and V_t are matrices of order p_1 and p_2 , respectively, and \otimes denotes the tensor (Kronecker) product of matrices:

$$U \otimes V = \begin{bmatrix} u_{11}V & \cdots & u_{1p}V \\ \cdots & \cdots & \cdots \\ u_{p1}V & \cdots & u_{pp}V \end{bmatrix}.$$

*Corresponding author. *Email address:* tee@inm.ras.ru (E. Tyrtyshnikov)

We denote by $\mathcal{T}_r = \mathcal{T}_r(p_1, p_2)$ the set of all matrices of the form (1.1) with real (for definiteness) entries for fixed values of r and p_1, p_2 , and we are especially interested in approximations

$$A \approx A_r \in \mathcal{T}_r$$

that minimize the Frobenius norm $\|A - A_r\|_F$ (the square root of the sum of squared entries in modulus). If $\|A - B\|_F > \|A - A_r\|_F$ for all $B \in \mathcal{T}_k$ with $k < r$ and $\|A - B\|_F \geq \|A - A_r\|_F$ for all $B \in \mathcal{T}_r$, then the minimizer matrix A_r will be called the *minimal* approximation of tensor rank r .

Let us assume that the blocks a_{ij} of a matrix A satisfy linear constraints

$$\sum_{i=1}^p \sum_{j=1}^p c_{ij} a_{ij} = 0 \tag{1.2}$$

with some fixed scalar coefficients c_{ij} . For this case in [6] it was discovered and proved that the entries of each of the matrices U_t of any minimal approximation A_r are subject to the same constraints (1.2). It follows, for instance, that if A is a block Toeplitz matrix (every block a_{ij} is a function of $i - j$) then each of the matrices U_t is a Toeplitz matrix. Similarly, if each of the blocks a_{ij} is a Toeplitz matrix then each of the matrices V_t ought to be Toeplitz.

In this paper we want to figure out to which extent the result of [6] can be generalized to the case of tensor approximations with an arbitrary fixed number of factors:

$$A_r = \sum_{t=1}^r U_t \otimes V_t \otimes \dots \otimes W_t. \tag{1.3}$$

Let the number of factors in every summand be equal to s and the orders of matrices U_t, V_t, \dots, W_t be $p = p_1, p_2, \dots, p_s$, respectively. Then the order of A_r is $n = p_1 p_2 \dots p_s$. Denote the set of all matrices of the form (1.3) by $\mathcal{T}_r^s = \mathcal{T}_r^s(p_1, p_2, \dots, p_s)$. For this case matrices A_r are used as approximations for a given matrix A of order $n = p_1 p_2 \dots p_s$.

The matrix A can be considered as a block matrix consisting of the blocks a_{ij} , $1 \leq i, j \leq p = p_1$. We will prove that from the viewpoint of preservation of linear constraints the case of arbitrary s is analogous to the case $s = 2$: if the equations (1.2) are valid then the minimality of approximation implies that the same relationships (1.2) hold true for each of the matrices U_t .

One essential difference is still there. Suppose that $A \in \mathcal{T}_{r_{\max}}$ and $A \notin \mathcal{T}_r$ whenever $r < r_{\max}$. Then for $s = 2$ the minimal approximation is constructed via the singular value decomposition (SVD) for any $1 \leq r \leq r_{\max}$, whereas in the case $s > 2$ there could be some values $1 < r < r_{\max}$ for which a minimal approximation of tensor rank r does not exist (cf. [5]). Moreover, there are no generalizations of the SVD to the case $s > 2$ that keep all the properties of this decomposition in the case $s = 2$ (some partial generalizations can be found in [1, 9]), and therefore, some other techniques are needed.

2. Minimal decompositions of tensors

It is convenient to reformulate our questions in the terms of approximation of tensors (multi-index arrays). Instead of a matrix A we consider an array of dimension s (tensor) $A = [A_{ij\dots k}]$, where the indices i, j, \dots, k take the values from finite sets I, J, \dots, K . Once the sets I, J, \dots, K are fixed, we can define the sum of tensors and the multiplication of a tensor by a scalar in a natural way enabling us to consider the set of tensors as a linear space.

The set of arrays A_r of the special form

$$(A_r)_{ij\dots k} = \sum_{t=1}^r U_{it} V_{jt} \cdots W_{kt}, \quad i \in I, \quad j \in J, \dots, k \in K,$$

will be denoted by \mathcal{T}_r . Since the array A_r is determined by matrices

$$U = [U_{it}], \quad V = [v_{jt}], \quad \dots, \quad W = [W_{kt}],$$

we shall write (due to Kruskal [4])

$$A_r = [U, V, \dots, W] = \sum_{t=1}^r U_t \otimes V_t \otimes \cdots \otimes W_t \tag{2.1}$$

and say that A_r is the sum of tensor products of one-dimensional arrays (tensors) U_t, V_t, \dots, W_t . We keep the same notation as in (1.3), but the ordering in which the entries of U_t, V_t, \dots, W_t are indexed is not prescribed and actually does not matter when this notation is used.

A decomposition of an array A of the form

$$A = \sum_{t=1}^r U_t \otimes V_t \otimes \cdots \otimes W_t \tag{2.2}$$

is called *minimal decomposition* of tensor rank r if $A \in \mathcal{T}_r$ but $A \notin \mathcal{T}_m$ for any $m < r$.

Lemma 2.1. *Any tensor possesses a minimal decomposition.*

Proof. Any tensor can be expressed in the form (2.2) with some number of summands r . Among all those decompositions we can obviously find one with minimal possible number of summands. □

Lemma 2.2. *If a decomposition (2.2) is minimal for a tensor A , then the tensors*

$$V_t \otimes \cdots \otimes W_t, \quad 1 \leq t \leq r,$$

are linearly independent.

Proof. From contrary, assume for definiteness that

$$V_{j_1} \cdots W_{k_1} = \sum_{t=2}^r \alpha_t V_{j_t} \cdots W_{k_t}.$$

Then

$$A_{ij \cdots k} = \sum_{t=2}^r (U_{it} + \alpha_t U_{i1}) V_{j_t} \cdots W_{k_t}.$$

Hence, $A \in \mathcal{T}_{r-1}$, which contradicts the minimality of decomposition (2.2). □

Theorem 2.1. *Assume that a tensor A satisfies linear constraints*

$$\sum_{i \in I} c_i A_{ij \cdots k} = 0, \quad j \in J, \dots, k \in K, \tag{2.3}$$

and possesses a minimal decomposition (2.2). Then

$$\sum_{i \in I} c_i U_{it} = 0, \quad 1 \leq t \leq r. \tag{2.4}$$

Proof. According to (2.3) and (2.2) we find

$$\sum_{t=1}^r \left(\sum_{i \in I} c_i U_{it} \right) V_{j_t} \cdots W_{k_t} = 0.$$

By Lemma 2.2 the tensors $V_t \otimes \cdots \otimes W_t$ are linearly independent, and the equations (2.4) follow from the fact that a zero linear combination of these tensors must have zero coefficients. □

Note that the indices i, j, \dots, k are equal “in rights”, and so Lemma 2.2 and Theorem 2.1 are valid when we consider summation in any of these indices. It stems, for example, that if an s -level Toeplitz (cf. [8, 12]) matrix $A = A_r$ is written as a sum of r tensor (Kronecker) products of the form (1.3) with the number of factors s and if (1.3) is its minimal decomposition then each of the matrices U_t, V_t, \dots, W_t is a Toeplitz matrix.

Remark that all assertions of this section are valid for tensors with entries from an arbitrary field.

3. Minimal approximations of tensors

By definition,

$$\|A\|_F = \left(\sum_{i \in I} \sum_{j \in J} \cdots \sum_{k \in K} |A_{ij \cdots k}|^2 \right)^{1/2}.$$

For a given array A , an array $A_r \in \mathcal{T}_r$ is called *minimal approximation* of tensor rank r if $\|A - A_m\|_F \geq \|A - A_r\|_F$ for all $m \leq r$ and $\|A - A_m\|_F > \|A - A_r\|_F$ for all $m < r$.

Lemma 3.1. *Given a matrix \mathcal{A} and a matrix G with r linearly independent columns, assume that a matrix F with r columns is such that for any matrix P with r columns*

$$\|\mathcal{A} - FG^\top\|_F \leq \|\mathcal{A} - PG^\top\|_F.$$

Then $F = \mathcal{A}Z$ for some matrix Z .

Proof. Let us write

$$F = \mathcal{A}Z + H,$$

where H is uniquely defined by the condition

$$\mathcal{A}^*H = 0.$$

Then

$$\|\mathcal{A} - FG^\top\|_F^2 = \|(\mathcal{A} - \mathcal{A}ZG^\top) - HG^\top\|_F^2 = \|\mathcal{A} - \mathcal{A}ZG^\top\|_F^2 + \|HG^\top\|_F^2.$$

It remains to observe that linear independent columns of G cause us to conclude that the equation $\|HG^\top\|_F = 0$ takes place if and only if $H = 0$. \square

Theorem 3.1. *Let a tensor A satisfy linear constraints (2.3) and a tensor A_r of the form (2.1) be its minimal approximation of tensor rank r . Then the linear constraints (2.4) are valid and the tensor A_r satisfies the same linear constraints (2.3).*

Proof. Build up a matrix \mathcal{A} from the elements $A_{ij\dots k}$ arranging them so that the index i points to rows while the columns are marked by the multi-index

$$v = (j, \dots, k).$$

Further, let $A_r = [U, V, \dots, W]$ be a minimal decomposition of tensor A_r . Let G be a matrix in which the column t contains the entries of the tensor $V_{jt} \dots W_{kt}$, the row position being defined by the same multi-index $v = (j, \dots, k)$. Then, as is readily seen,

$$\|A - A_r\|_F = \|\mathcal{A} - UG^\top\|_F.$$

By Lemma 2.2 the columns of G are linearly independent. Therefore, we may apply Lemma 3.1 for the matrices \mathcal{A} , $F = U$ and G . In the end we obtain

$$U = \mathcal{A}Z.$$

In chime with (2.3) we have

$$\sum_{i \in I} c_i \mathcal{A}_{iv} = 0.$$

It follows that

$$\sum_{i \in I} c_i U_{it} = \sum_v \left(\sum_{i \in I} c_i \mathcal{A}_{iv} \right) Z_{vt} = 0, \quad 1 \leq t \leq r,$$

which completes the proof. \square

Corollary 3.1. *Assume that a matrix A_r is a minimal tensor approximation of the form (1.3) for an s -level matrix A whose blocks satisfy linear constraints (1.2). Then the entries of the matrices U_t from (1.3) satisfy the same linear constraints (1.2).*

Remark that the results obtained are valid for linear constraints defined by summation in any of indices i, j, \dots, k . For example, if we consider summation in indices j and k then linear constraints for the original tensor are maintained for the tensors V_t and W_t for all t .

In conclusion, we would like to stress a nice practical value of our results. Assume that a d -level Toeplitz matrix A with the level sizes all equal to n is approximated by a matrix A_r of the form (1.3) where the matrices U_t, V_t, \dots, W_t are of order n . If the best possible r -term approximation exists and r is minimal possible value for the achieved approximation accuracy, then Theorem 3.1 states that the matrices U_t, V_t, \dots, W_t are Toeplitz. Hence, all these matrices are determined by $(2n-1)r$ parameters, instead of n^2r parameters, and the whole approximation problem for A reduces exactly to the same approximation problem for a d -dimensional array of size $(2n-1) \times \dots \times (2n-1)$.

Acknowledgments This work was supported by the Russian Fund for Basic Research (RFBR grant 08-01-00115, RFBR/DFG grant 09-01-91332, RFBR grant 09-01-12058) and Priority Research Programme of Department of Mathematical Sciences of Russian Academy of Sciences.

References

- [1] L. DE LATHAUWER, B. DE MOOR, AND J. VANDEWALLE, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., vol. 21, no. 4 (2000), pp. 1253–1278.
- [2] W. HACKBUSCH, B. N. KHOROMSKIJ, AND E. E. TYRTYSHNIKOV, *Hierarchical Kronecker tensor-product approximations*, J. Numer. Math., vol. 13 (2005), pp. 119–156.
- [3] J. KAMM AND J. G. NAGY, *Optimal Kronecker Product Approximations of Block Toeplitz Matrices*, SIAM J. Matrix Anal. Appl., vol. 22, no. 1 (2000), pp. 155–172.
- [4] J. B. KRUSKAL, *Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra Appl., 18 (1977), pp. 95–138.
- [5] L.-H. LIM AND V. DE SILVA, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM J. Matrix Anal. Appl., vol. 30, no. 3, (2008), pp. 1084–1127.
- [6] V. OLSHEVSKY, I. OSELEDETS, AND E. TYRTYSHNIKOV, *Tensor properties of multilevel Toeplitz and related matrices*, Linear Algebra Appl., 412 (2006), pp. 1–21.
- [7] V. OLSHEVSKY, I. OSELEDETS, AND E. TYRTYSHNIKOV, *Superfast inversion of two-level Toeplitz matrices using Newton iteration and tensor-displacement structure*, Operator Theory Advances and Applications, vol. 179 (2007), pp. 229–240.
- [8] E. E. TYRTYSHNIKOV, *Toeplitz matrices, some their analogs and applications*, OVM RAN, Moscow, 1989 (in Russian).
- [9] E. E. TYRTYSHNIKOV, *Matrix analysis and linear algebra*, Fizmatlit, Moscow, 2007 (in Russian).
- [10] E. E. TYRTYSHNIKOV, *Tensor approximations of matrices generated by asymptotically smooth functions*, Sbornik: Mathematics, vol. 194, no. 5-6 (2003), 941–954.
- [11] E. E. TYRTYSHNIKOV, *Kronecker-product approximations for some function-related matrices*, Linear Algebra Appl., 379 (2004), pp. 423–437.
- [12] V. V. VOEVODIN AND E. E. TYRTYSHNIKOV, *Computations processes with Toeplitz matrices*, Nauka, Moscow, 1987 (in Russian).