ON THE ADJOINT OF TENSORS AND SOME ASSOCIATED MUSINGS

MOODY T. CHU*

Abstract. The notion of adjoint plays a fundamental role in many applications and across different fields. This note studies the notion of adjoint when a multi-dimensional array is regarded as the representation of a linear transformation between tensor spaces. A variety of associated concepts such as symmetry, orthogonality, best approximation of outer product factorization, and rank reduction formula are discussed.

Key words. adjoint, linear transformation, orthogonality

AMS subject classifications. 15A15, 15A09, 15A23

1. Introduction. To convey the idea more easily, we shall concern ourselves in this paper with real-valued tensors only. By taking into account the conjugate symmetry, generalization to complex tensors can be done similarly. A tensor of order-k is typically represented by a k-way array

$$T = [\tau_{i_1,\ldots,i_k}] \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_k},$$

where elements $t_{i_1,\ldots,i_k}$ are accessed via k indices. When $I_1 = \ldots = I_k$, we say that T is a square tensor. An order-k square tensor T is said to be super-symmetric [2, 3, 5, 8] if

$$\tau_{i_1,\ldots,i_k} = \tau_{i_{\sigma(1)},\ldots,i_{\sigma(k)}} \tag{1.1}$$

with respect to all possible permutations σ over the integers $\{1,\ldots k\}$. Such a well-rounded super-symmetry, totally immune to permutations of indices, seems to be a natural generalization of the symmetric matrices. However, we shall argue that this straightforward structural emulation, interesting in itself and is equivalent to homogeneous polynomials [2, Section 3], does not carry a similar meaning of a self-adjoint operator and the associated properties.

In linear algebra, the transpose $A^\top$ of a (real-valued) matrix A is precisely the matrix representation of the adjoint of the linear transformation being represented by A. Because a matrix involves only two indices, its adjoint operator can easily be obtained by merely swapping the order of the indices. If we think of a multi-dimensional array as the representation of a linear transformation between tensor subspaces, how the transposition should be addressed? The purpose of this paper is to explore the notion of adjoint of a tensor and to understand how the many important concepts already studied in matrix theory, such as orthogonal transformation, Gram-Schmidt process, rank reduction and so on, can be and should be generalized to multi-dimensional arrays.

2. Tensor as a linear transformation. For convenience, the set of integers $\{1,\ldots n\}$ for a given n will be abbreviated to the symbol $[\![n]\!]$ henceforth. Suppose that the set $[\![k]\!]$ is partitioned as the union of two disjoint nonempty subsets $\boldsymbol{\alpha} = \{\alpha_1,\ldots,\alpha_s\}$ and $\boldsymbol{\beta} = \{\beta_1,\ldots,\beta_t\}$, where $s+t=k$. Elements in these subsets will be used as pointers of locations throughout this paper. The reason for employing pointers is because there is no easy way to visualize high-order tensors as matrices. Choosing various ways to partition $[\![k]\!]$ offers us a convenient tool to dissect a high-dimensional T and exam its cross-sections from different perspectives.

An element in the tensor T will be marked as $\tau^{(\boldsymbol{\alpha},\boldsymbol{\beta})}_{[\mathcal{I}|\mathcal{J}]}$ where $\mathcal{I} := (i_1,\ldots,i_s)$ and $\mathcal{J} := (j_1,\ldots,j_t)$ contain those indices at locations $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. Each index in the arrays $\mathcal{I}$ and $\mathcal{J}$ should be within the corresponding range of integers, e.g., $i_1 \in [\![I_{\alpha_1}]\!]$ and so on. Since $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are merely sets, any permutation of their elements and the corresponding shuffle of indices in $\mathcal{I}$ and $\mathcal{J}$ should not alter the tensor element they are representing. If the reference to a specific partitioning $(\boldsymbol{\alpha},\boldsymbol{\beta})$ is clear, then without causing ambiguity we abbreviate the element as $\tau_{[\mathcal{I}|\mathcal{J}]}$. For example, if $\boldsymbol{\alpha} = \{2,4\}$ and $\boldsymbol{\beta} = \{1,3,5,6\}$ are fixed and known, then

*Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205. (chu@math.ncsu.edu) This research was supported in part by the National Science Foundation under grant DMS-1316779.

$\tau_{2,3,1,5,6,4} = \tau_{[3,5|2,1,6,4]}$. There is no preference of $\boldsymbol{\alpha}$ over $\boldsymbol{\beta}$, so $\tau^{(\boldsymbol{\alpha},\boldsymbol{\beta})}_{[\mathcal{I}|\mathcal{J}]} = \tau^{(\boldsymbol{\beta},\boldsymbol{\alpha})}_{[\mathcal{J}|\mathcal{I}]}$ for any fixed partitioning $[\![k]\!] = \boldsymbol{\alpha} \cup \boldsymbol{\beta}$. Such a symmetry with respect to the partitioning does not indicate any symmetry with respect to the indices at all.

Given a fixed partitioning $[\![k]\!] = \boldsymbol{\alpha} \cup \boldsymbol{\beta}$, we shall regard an order-k tensor $T \in \mathbb{R}^{I_1 \times \ldots \times I_k}$ as a "matrix representation" of a linear operator mapping order-s tensors to order-t tensors [9]. Specifically, we identify T with the map

$$\mathscr{T}_{\boldsymbol{\beta}} : \mathbb{R}^{I_{\alpha_1} \times \ldots \times I_{\alpha_s}} \to \mathbb{R}^{I_{\beta_1} \times \ldots \times I_{\beta_t}}, \tag{2.1}$$

where, for any $A \in \mathbb{R}^{I_{\alpha_1} \times \ldots \times I_{\alpha_s}}$, we have

$$\mathscr{T}_{\boldsymbol{\beta}}(A) := T \circledast_{\boldsymbol{\beta}} A = \left[\langle \tau_{[:|\mathcal{J}]}, A\rangle\right] \in \mathbb{R}^{I_{\beta_1} \times \ldots \times I_{\beta_t}}. \tag{2.2}$$

In the above, $\tau_{[:|\mathcal{J}]} = \tau_{[:|j_1,\ldots,j_t]}$ denotes the $(j_1, \ldots, j_t)$-th “slice" in the $\boldsymbol{\beta}$ direction of the tensor T, that is, the index $j_\ell \in [\![I_{\beta_\ell}]\!]$ occurs at the β_ℓ-th location in the array $[\![k]\!]$ for $\ell = 1, \ldots t$, whereas the notation ":" represents a wild card at the $(\alpha_1, \ldots, \alpha_s)$ location and

$$\langle \tau_{[:|\mathcal{J}]}, A\rangle := \sum_{i_1=1}^{I_{\alpha_1}} \ldots \sum_{i_s=1}^{I_{\alpha_s}} \tau_{[i_1,\ldots,i_s|\mathcal{J}]} a_{i_1,\ldots,i_s} \tag{2.3}$$

is the Frobenius inner product generalized to multi-dimensional arrays. Indeed, (2.3) can be abbreviated as

$$(\mathscr{T}_{\boldsymbol{\beta}}(A))_{\mathcal{J}} = \sum_{\mathcal{I}} \tau_{[\mathcal{I}|\mathcal{J}]} a_{\mathcal{I}}, \tag{2.4}$$

where the summation of $\mathcal{I}$ runs through appropriate ranges of the indices $i_1, \ldots, i_{\alpha_s}$. In terms of this multi-index notation, the tensor-to-tensor operation $\circledast_{\boldsymbol{\beta}}$ defined in (2.2) generalizes the usual matrix-to-vector multiplication. At first glance, the expression (2.4) may seem awkward because in the matrix-to-vector multiplication $A\mathbf{x}$ we usually write the jth entry as $(A\mathbf{x})_j = \sum_i a_{ji} x_i$. Such a distinction between rows and columns is now replaced by the locators $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. For instance, we may unambiguously rewrite the summation in (2.4) as

$$(\mathscr{T}_{\boldsymbol{\beta}}(A))_{\mathcal{J}} = \sum_{\mathcal{I}} \tau^{(\boldsymbol{\beta},\boldsymbol{\alpha})}_{[\mathcal{J}|\mathcal{I}]} a_{\mathcal{I}}, \tag{2.5}$$

which is more analogous to the classical matrix-to-vector multiplication.

When dealing with multi-dimensional arrays, it is no longer realistic to visualize them in the shapes we prefer to make out. It might be more practical to have a unified and systematic way to store the data as a one-dimensional array. Indeed, that is precisely how a computer stores an array in the memory. Using the Matlab as an example, the rule of composing this 1-D array is to enumerate data column by column, each appended to the last. More specifically, the entry $\tau_{i_1,\ldots,i_k}$ of an order-k tensor $T \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_k}$ is saved at the location

$$(i_k - 1)I_{k-1}I_{k-2} \ldots I_1 + (i_{k-1} - 1)I_{k-2} \ldots I_1 + ... + (i_2 - 1)I_1 + i_1 \tag{2.6}$$

of the linear array. In the subsequent discussion, we shall comply with this rule whenever we want to fold or reshape a tensor. For example, to emulate the usual matrix representation, it might be convenient to rearrange T in the way that elements are read in the order $[\beta_1, \ldots, \beta_t, \alpha_1, \ldots, \alpha_s]$. We say that the operator $\mathscr{T}_{\boldsymbol{\beta}}$ is represent by the tensor $T^{\boldsymbol{\alpha}}_{\boldsymbol{\beta}}$ generated by the Matlab command

```
T_beta_alpha = permute(T,[beta,alpha])
```

where `[beta,alpha]` indicates the order of the subscripts to be accessed when identifying a particular element. If elements of $T_{\boldsymbol{\beta}}^{\boldsymbol{\alpha}}$ are renamed as $\mathfrak{t}_{j_1,\ldots,j_t,i_1,\ldots,i_s}$ for $(j_1,\ldots,j_t) \in [\![I_{\beta_1}]\!] \times \ldots \times [\![I_{\beta_t}]\!]$ and $(i_1,\ldots,i_s) \in [\![I_{\alpha_1}]\!] \times \ldots \times [\![I_{\alpha_s}]\!]$, then

$$(\mathscr{T}_{\boldsymbol{\beta}}(A))_{j_1,\ldots,j_t} = \sum_{i_1=1}^{I_{\alpha_1}} \cdots \sum_{i_s=1}^{I_{\alpha_s}} \mathfrak{t}_{j_1,\ldots,j_t,i_1,\ldots,i_s} a_{i_1,\ldots,i_s}, \tag{2.7}$$

which is equivalent to (2.5).

3. Adjoint operator. The adjoint of $\mathscr{T}_{\boldsymbol{\beta}}$ should be a linear transformation

$$\mathscr{T}_{\boldsymbol{\beta}}^* : \mathbb{R}^{I_{\beta_1} \times \ldots \times I_{\beta_t}} \to \mathbb{R}^{I_{\alpha_1} \times \ldots \times I_{\alpha_s}} \tag{3.1}$$

such that the Lagrange's identity [7]

$$\langle \mathscr{T}_{\boldsymbol{\beta}}(A), B \rangle = \left\langle A, \mathscr{T}_{\boldsymbol{\beta}}^*(B) \right\rangle \tag{3.2}$$

is satisfied for all $A \in \mathbb{R}^{I_{\alpha_1} \times \ldots \times I_{\alpha_s}}$ and $B \in \mathbb{R}^{I_{\beta_1} \times \ldots \times I_{\beta_t}}$.

LEMMA 3.1. *Given an order-k tensor T and a fixed partitioning $[\![k]\!] = \boldsymbol{\alpha} \cup \boldsymbol{\beta}$, then*

$$\mathscr{T}_{\boldsymbol{\beta}}^* = \mathscr{T}_{\boldsymbol{\alpha}}. \tag{3.3}$$

The representation of $\mathscr{T}_{\boldsymbol{\beta}}^*$ therefore is the tensor $T_{\boldsymbol{\alpha}}^{\boldsymbol{\beta}}$.

Proof. Observe that

$$\langle \mathscr{T}_{\boldsymbol{\beta}}(A), B \rangle = \sum_{\mathcal{J}} \left(\sum_{\mathcal{I}} \tau_{[\mathcal{I}|\mathcal{J}]} a_{\mathcal{I}} \right) b_{\mathcal{J}} = \sum_{\mathcal{I}} \left(\sum_{\mathcal{J}} \tau_{[\mathcal{I}|\mathcal{J}]} b_{\mathcal{J}} \right) a_{\mathcal{I}}.$$

It follows that

$$(\mathscr{T}_{\boldsymbol{\beta}}^*(B))_{\mathcal{I}} = \sum_{\mathcal{J}} \tau_{[\mathcal{I}|\mathcal{J}]} b_{\mathcal{J}}. \tag{3.4}$$

Recall that $\tau_{[\mathcal{I}|\mathcal{J}]} = \tau_{[\mathcal{I}|\mathcal{J}]}^{(\boldsymbol{\alpha},\boldsymbol{\beta})} = \tau_{[\mathcal{J}|\mathcal{I}]}^{(\boldsymbol{\beta},\boldsymbol{\alpha})}$. By the definition of (2.2), the action in (3.4) is the same as that of $\mathscr{T}_{\boldsymbol{\alpha}}$. $\square$

In the case $k = 2$ and $T \in \mathbb{R}^{m \times n}$, then $\mathscr{T}_1 : \mathbb{R}^n \to \mathbb{R}^m$ and its matrix representation is precisely T, while $\mathscr{T}_2 : \mathbb{R}^m \to \mathbb{R}^n$ and is represented by $T^\top$. It is clear that $\mathscr{T}_2 = \mathscr{T}_1^* = T^\top$. Note that the 2-D arrays T and $T^\top$ contain the same set of data. Their structural difference is only a matter of being perceived from different perspectives as rows and columns. Lemma 3.1 conveniently generalizes the notion of adjoint to tensors — The "transpose" of a tensor T, viewed from the $\boldsymbol{\beta}$ point of view, is the very same T viewed from the $\boldsymbol{\alpha}$ point of view, and vice versa. For matrices, there are only two ways to orient the data, i.e., by rows or by columns. For high-order tensors, the notion of an adjoint is $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ dependent.

Example 1. It might be illustrative to display these structures by an order-3 tensor, although such an attempt to visualize higher-order tensors is infeasible (and useless). Consider the partitioning with $\boldsymbol{\alpha} = \{1, 2\}$ and $\boldsymbol{\beta} = \{3\}$. The order-3 tensor T may be visualized as an $I_3 \times 1$ block matrix of which each block is an $I_1 \times I_2$ matrix, whereas the action $\circledast_3$ is a 2-dimensional contraction defined by

$$\mathscr{T}_3(X) \equiv \underbrace{\begin{bmatrix} \boxed{\tau_{:,1}} \\ \boxed{\tau_{:,2}} \\ \vdots \\ \boxed{\tau_{:,I_3}} \end{bmatrix}}_{T} \circledast_3 X = [\langle \tau_{:,j}, X \rangle] \in \mathbb{R}^{I_3},$$

for $X \in \mathbb{R}^{I_1 \times I_2}$. The adjoint $\mathscr{T}_3^* = \mathscr{T}_{1,2}$ can be visualized as an $I_1 \times I_2$ block matrix of which each block is an $I_3 \times 1$ column vector with action defined by

$$\mathscr{T}_{1,2}(X) \equiv \underbrace{\begin{bmatrix} \boxed{\tau_{1,1,:}} & \boxed{\tau_{1,2,:}} & \dots & \boxed{\tau_{1,I_2,:}} \\ \boxed{\tau_{2,1,:}} & & \dots & \boxed{\tau_{2,I_2,:}} \\ \vdots & & & \vdots \\ \boxed{\tau_{I_1,1,:}} & & & \boxed{\tau_{I_1,I_2}} \end{bmatrix}}_{T^*} \circledast_{12} \mathbf{x} = [\langle \tau_{i,j,:}, \mathbf{x} \rangle] \in \mathbb{R}^{I_1 \times I_2},$$

for $\mathbf{x} \in \mathbb{R}^{I_3}$. Both T and T^* contain the same set of entries but are organized differently, depending on $(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

Under the context of operators, a self-adjoint tensor therefore should be such that

$$\mathscr{T}_{\boldsymbol{\alpha}} = \mathscr{T}_{\boldsymbol{\alpha}}^*. \tag{3.5}$$

In turn, by Lemma 3.1, we need $\mathscr{T}_{\boldsymbol{\alpha}} = \mathscr{T}_{\boldsymbol{\beta}}$. It thus becomes necessary that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ have the same cardinality, say, s, and $I_{\alpha_\ell} = I_{\beta_\ell}$, for $\ell = 1, \dots, s$. It follows immediately that an odd order tensor can never be self-adjoint regardless how the partitioning is taken. Two important remarks are due. First, the notion of super-symmetry does allow odd order tensors, so it does not agree with the conventional concept of adjoint. Second, the super-symmetry can happen only in square tensors, but there is no such a restriction on self-adjoint tensors. For self-adjoint tensors, the condition (3.5) requires that

$$\langle \tau^{(\boldsymbol{\alpha},\boldsymbol{\beta})}_{[:|\mathcal{J}]}, A \rangle = \langle \tau^{(\boldsymbol{\beta},\boldsymbol{\alpha})}_{[:|\mathcal{J}]}, A \rangle \tag{3.6}$$

for all order-s tensors $A \in \mathbb{R}^{I_{\alpha_1} \times \dots \times I_{\alpha_s}}$ and all multi-indices $\mathcal{J} \in [\![I_{\alpha_1}]\!] \times \dots \times [\![I_{\alpha_s}]\!]$. Therefore, a real-valued order-$2s$ tensor T is "symmetric" with respect to the partitioning $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ if and only if

$$\tau^{(\boldsymbol{\alpha},\boldsymbol{\beta})}_{[\mathcal{I}|\mathcal{J}]} = \tau^{(\boldsymbol{\beta},\boldsymbol{\alpha})}_{[\mathcal{I}|\mathcal{J}]} = \tau^{(\boldsymbol{\alpha},\boldsymbol{\beta})}_{[\mathcal{J}|\mathcal{I}]} \tag{3.7}$$

for all multi-indices $\mathcal{I}, \mathcal{J} \in [\![I_{\alpha_1}]\!] \times \dots \times [\![I_{\alpha_s}]\!]$. Once a tensor T is permuted into $T^{\boldsymbol{\alpha}}_{\boldsymbol{\beta}}$ and $T^{\boldsymbol{\beta}}_{\boldsymbol{\alpha}}$, which is easy to do computationally, we can check the symmetry by examining whether

$$T^{\boldsymbol{\alpha}}_{\boldsymbol{\beta}} = T^{\boldsymbol{\beta}}_{\boldsymbol{\alpha}}. \tag{3.8}$$

Example 2. Consider the order-4 tensor $T \in \mathbb{R}^{2\times 3\times 2\times 3}$ which we lay out as the "matrix" of 2×3 blocks of 2×3 matrices depicted below,

$$T = \begin{bmatrix} \boxed{\begin{matrix} \tau_{1111} & \tau_{1211} & \tau_{1311} \\ \tau_{2111} & \tau_{2211} & \tau_{2311} \end{matrix}} & \boxed{\begin{matrix} \tau_{1112} & \tau_{1212} & \tau_{1312} \\ \tau_{2112} & \tau_{2212} & \tau_{2312} \end{matrix}} & \boxed{\begin{matrix} \tau_{1113} & \tau_{1213} & \tau_{1313} \\ \tau_{2113} & \tau_{2213} & \tau_{2313} \end{matrix}} \\ \boxed{\begin{matrix} \tau_{1121} & \tau_{1221} & \tau_{1321} \\ \tau_{2121} & \tau_{2221} & \tau_{2321} \end{matrix}} & \boxed{\begin{matrix} \tau_{1122} & \tau_{1222} & \tau_{1322} \\ \tau_{2122} & \tau_{2222} & \tau_{2322} \end{matrix}} & \boxed{\begin{matrix} \tau_{1123} & \tau_{1223} & \tau_{1323} \\ \tau_{2123} & \tau_{2223} & \tau_{2323} \end{matrix}} \end{bmatrix},$$

where $\tau_{i,j,k,\ell}$ is located at the (i,j)-th entry of the (k,ℓ)-th block. Then, by (3.7), T is symmetric with respect to the partitioning $\boldsymbol{\alpha} = \{1, 2\}$ and $\boldsymbol{\beta} = \{3, 4\}$ if $\tau_{i,j,k,\ell} = \tau_{k,\ell,i,j}$. We pair these relationships in colors, except that each of the six entries in black stands alone. We may rearrange the very same entries of T in the order

```
S= permute(T,[2,4,1,3])
```

as a tensor $S \in \mathbb{R}^{3\times3\times2\times2}$. Lay out S as a 2×2 block matrix of 3×3 blocks

$$S = [\sigma_{i,j,k,\ell}] = \left[\begin{array}{c|c} \begin{matrix} \tau_{1111} & \tau_{1112} & \tau_{1113} \\ \tau_{1211} & \tau_{1212} & \tau_{1213} \\ \tau_{1311} & \tau_{1312} & \tau_{1313} \end{matrix} & \begin{matrix} \tau_{1121} & \tau_{1122} & \tau_{1123} \\ \tau_{1221} & \tau_{1222} & \tau_{1223} \\ \tau_{1321} & \tau_{1322} & \tau_{1323} \end{matrix} \\ \hline \begin{matrix} \tau_{2111} & \tau_{2112} & \tau_{2113} \\ \tau_{2211} & \tau_{2212} & \tau_{2213} \\ \tau_{2311} & \tau_{2312} & \tau_{2313} \end{matrix} & \begin{matrix} \tau_{2121} & \tau_{2122} & \tau_{2123} \\ \tau_{2221} & \tau_{2222} & \tau_{2223} \\ \tau_{2321} & \tau_{2322} & \tau_{2323} \end{matrix} \end{array}\right].$$

It is interesting to observe that the symmetry of the tensor T implies that the flattened S shown above is a symmetric matrix. To further check whether S is symmetric with respect to the partitions of either $(\{1,3\},\{2,4\})$ or $(\{1,4\},\{2,3\})$, we really should check if $\sigma_{i,j,k,\ell} = \sigma_{j,i,\ell,k}$. For example, $\sigma_{1312} = \tau_{1123} = \tau_{2311} = \sigma_{3121}$. It turns out that in this case, S is still symmetric. The visual display here is deceiving and is actually useless for high-order tensors. Our point is that the way the data are organized can affect their interpretation, so much so that even the symmetry might have different meaning.

4. Orthogonal tensor. An orthogonal transformation is a linear transformation on a real inner product space that preserves the inner product. With the availability of its transpose, a square matrix $Q \in \mathbb{R}^{n\times n}$ is said to be orthogonal if $Q^\top Q$ is equal to the identity matrix $\mathscr{I}_n$. Equivalently, the "columns" of Q are mutually orthonormal. Many important properties follow from orthogonal transformations. A natural question to ask is what is meant by an orthogonal tensor. Which "columns" of a tensor should be mutually orthonormal?

Without any presumption except that a partitioning $(\boldsymbol{\alpha},\boldsymbol{\beta})$ is given, we first consider the composition

$$\mathbb{R}^{I_{\alpha_1}\times\ldots\times I_{\alpha_s}} \xrightarrow{\mathscr{T}_{\boldsymbol{\beta}}} \mathbb{R}^{I_{\beta_1}\times\ldots\times I_{\beta_t}} \xrightarrow{\mathscr{T}_{\boldsymbol{\alpha}}} \mathbb{R}^{I_{\alpha_1}\times\ldots\times I_{\alpha_s}}. \tag{4.1}$$

It is desired that $\mathscr{T}_{\boldsymbol{\alpha}}(\mathscr{T}_{\boldsymbol{\beta}}(A)) = T \circledast_{\boldsymbol{\alpha}} (T \circledast_{\boldsymbol{\beta}} A) = A$ for every $A \in \mathbb{R}^{I_{\alpha_1}\times\ldots\times I_{\alpha_s}}$. With respect to any given multi-index $\mathcal{I} \in [\![I_{\alpha_1}]\!] \times \ldots \times [\![I_{\alpha_s}]\!]$, observe that

$$(\mathscr{T}_{\boldsymbol{\alpha}}(\mathscr{T}_{\boldsymbol{\beta}}(A)))_{\mathcal{I}} = \sum_{\mathcal{J}} \tau_{[\mathcal{I}|\mathcal{J}]} \sum_{\mathcal{K}} \tau_{[\mathcal{K}|\mathcal{J}]} a_{\mathcal{K}} = \sum_{\mathcal{K}} \left(\sum_{\mathcal{J}} \tau_{[\mathcal{I}|\mathcal{J}]} \tau_{[\mathcal{K}|\mathcal{J}]} \right) a_{\mathcal{K}}. \tag{4.2}$$

We therefore want

$$\sum_{\mathcal{J}} \tau_{[\mathcal{I}|\mathcal{J}]} \tau_{[\mathcal{K}|\mathcal{J}]} = \delta_{\mathcal{I}\mathcal{K}}, \tag{4.3}$$

where $\delta_{\mathcal{I}\mathcal{K}}$ is the Kronecker delta notation applied to the multi-indices $\mathcal{I}$ and $\mathcal{K}$ at the locations indicated by $\boldsymbol{\alpha}$. The relationship (4.3) is analogous to $QQ^\top = \mathscr{I}_n$, except that the matrix-to-matrix multiplication should be interpreted as the composition $\mathscr{T}_{\boldsymbol{\alpha}} \circ \mathscr{T}_{\boldsymbol{\alpha}}^* = \mathscr{I}_{\boldsymbol{\alpha}}$ where $\mathscr{I}_{\boldsymbol{\alpha}}$ denotes the identity map over $\mathbb{R}^{I_{\alpha_1}\times\ldots\times I_{\alpha_s}}$. Also, if we interpret the order-t subtensor $\tau_{[\mathcal{I}|:]} \in \mathbb{R}^{I_{\beta_1}\times\ldots\times I_{\beta_t}}$ as the $\mathcal{I}$-th row of $\mathscr{T}_{\boldsymbol{\alpha}}$[1], then (4.3) is saying that the rows of $\mathscr{T}_{\boldsymbol{\alpha}}$ are mutually orthonormal. Likewise, we want to see that $\mathscr{T}_{\boldsymbol{\beta}}(\mathscr{T}_{\boldsymbol{\alpha}}(B)) = T \circledast_{\boldsymbol{\beta}} (T \circledast_{\boldsymbol{\alpha}} B) = B$ for every $B \in \mathbb{R}^{I_{\beta_1}\times\ldots\times I_{\beta_t}}$. This relationship

$$(\mathscr{T}_{\boldsymbol{\beta}}(\mathscr{T}_{\boldsymbol{\alpha}}(B)))_{\mathcal{J}} = \sum_{\mathcal{I}} \tau_{[\mathcal{I}|\mathcal{J}]} \sum_{\mathcal{K}} \tau_{[\mathcal{I}|\mathcal{K}]} b_{\mathcal{K}} = \sum_{\mathcal{K}} \left(\sum_{\mathcal{I}} \tau_{[\mathcal{I}|\mathcal{J}]} \tau_{[\mathcal{I}|\mathcal{K}]} \right) b_{\mathcal{K}} \tag{4.4}$$

implies that

$$\sum_{\mathcal{I}} \tau_{[\mathcal{I}|\mathcal{J}]} \tau_{[\mathcal{I}|\mathcal{K}]} = \delta_{\mathcal{J}\mathcal{K}}. \tag{4.5}$$

[1] In view of (2.5), it is natural to interpret $\tau_{[\mathcal{I}|:]}$ as the $\mathcal{I}$-th column of $\mathscr{T}_{\boldsymbol{\beta}}$.

The summation involved in (4.5) is equivalent to $\mathscr{T}_{\boldsymbol{\alpha}}^{*} \circ \mathscr{T}_{\boldsymbol{\alpha}} = \mathscr{I}_{\boldsymbol{\beta}}$ and is the analogue of $Q^{\top}Q = \mathscr{I}_n$. The order-$s$ subtensor $\tau_{[:|\mathcal{J}]} \in \mathbb{R}^{I_{\alpha_1} \times \ldots \times I_{\alpha_s}}$ is interpreted as the $\mathcal{J}$-th column of $\mathscr{T}_{\boldsymbol{\alpha}}$.

A tensor T is said to be orthogonal with respect to the partitioning $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ if its representation $\mathscr{T}_{\boldsymbol{\alpha}}$ satisfies both (4.3) and (4.5). What is interesting is that $\mathscr{T}_{\boldsymbol{\alpha}}$ need not be a square tensor to be orthogonal as we shall demonstrate in the following example.

Example 3. Consider the setting similar to Example 1 with the partitioning $\boldsymbol{\alpha} = \{2, 3\}$ and $\boldsymbol{\beta} = \{1\}$. Then (4.3) becomes

$$\sum_{j=1}^{I_1} \tau_{j,i_2,i_3} \tau_{j,k_2,k_3} = \delta_{(i_2,i_3),(k_2,k_3)} \tag{4.6}$$

which means that "columns" $\tau_{:,i_2,i_3} \in \mathbb{R}^{I_1}$ should be mutually orthonormal. There are a total of $I_2 I_3$ many such columns. Obviously, such an orthogonality is impossible if $I_1 < I_2 I_3$. Similarly, (4.5) becomes

$$\sum_{i_2=1}^{I_2} \sum_{i_3=1}^{I_3} \tau_{j,i_2,i_3} \tau_{k,i_2,i_3} = \delta_{jk} \tag{4.7}$$

which means that "blocks" $\tau_{i,:} \in \mathbb{R}^{I_2 \times I_3}$ are mutually orthonormal. If $I_1 > I_2 I_3$, then the number of blocks will be greater than the dimension of the ambient space. In short, a necessary condition for both (4.6) and (4.7) to hold simultaneously is that $I_1 = I_2 I_3$. The special case when $I_1 = I_2 = n$ and $I_3 = 1$ corresponds precisely to the orthogonal matrices which form an $\frac{n(n-1)}{2}$-dimensional manifold. For the general case when $I_1 = I_2 I_3 = n$, the condition (4.6) corresponds to the phenomenon that the flatten $n \times n$ matrix

$$\Omega := \left[\tau_{:,1,1}, \tau_{:,1,2}, \ldots \tau_{:,1,I_3}, \tau_{:,2,1}, \ldots \tau_{:,I_2,I_3}\right],$$

where each $\tau_{:,i_2,i_3}$ is an n-dimensional column vector, is orthogonal. Additionally, by folding each row of Ω into an $I_2 \times I_3$ matrix, then (4.7) is automatically satisfied. It is interesting to note that, in contrast to the fact that an orthogonal matrix is always a square matrix, the order-3 tensor satisfying both (4.6) and (4.7) is not a square tensor. The condition $I_1 = I_2 I_3$ is a generalization of an orthogonal matrix being square. The following lemma asserts such an observation is true in general.

THEOREM 4.1. *Given a partitioning $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_s\}$ and $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_t\}$, a necessary condition for both (4.6) and (4.7) to hold simultaneously is that*

$$\prod_{i=1}^{s} I_{\alpha_i} = \prod_{j=1}^{t} I_{\beta_j}. \tag{4.8}$$

In this case, (4.6) holds if and only if (4.7) holds.

Proof. The reason for (4.8) to hold is parallel to what we have already argued in Example 3. We shall prove only one direction. For each $\mathcal{I} \in [\![I_{\alpha_1}]\!] \times \ldots \times [\![I_{\alpha_s}]\!]$, $\tau_{[\mathcal{I}|:]}$ is a tensor in $\mathbb{R}^{I_{\beta_1} \times \ldots \times I_{\beta_t}}$. For these tensors $\tau_{[\mathcal{I}|:]}$ to be mutually orthonormal with respect to the Frobenius inner product, as is dictated by (4.3), it is necessary that $\prod_{i=1}^{s} I_{\alpha_i} \leq \prod_{j=1}^{t} I_{\beta_j}$. A similar argument works for the other direction.

Suppose now that (4.3) holds. Vectorize each tensor $\tau_{[I|:]}$ into a $\prod_{j=1}^{t} I_{\beta_j}$-dimensional column vector according to (2.6). There are a total of $\prod_{i=1}^{s} I_{\alpha_i}$ many such vectors. The matrix obtained by assembling these vectors column-wise is a square matrix by (4.8) and is orthogonal by (4.3). As such, its rows which fold into $\tau_{[:|J]}$ are also mutually orthonormal. $\square$

5. QR decomposition. There is no need to reiterate the theory and the applications of the QR decomposition of any given matrix. Can an order-k tensor be factorized in a similar way? What is meant by an "upper triangular" tensor?

6. Approximation of outer product factorization. There are many ways to define multiplication between tensors [6]. We have seen the operator type of multiplication $\circledast_{\boldsymbol{\beta}}$ earlier in (2.2). The so called outer product is perhaps the most fundamental operation. Give an order-m tensor $F \in \mathbb{R}^{I_1 \times \ldots \times I_m}$ and an order-n tensor $G \in \mathbb{R}^{J_1 \times \ldots \times J_n}$, the tensor product $P = F \otimes G$ is an order-$(m+n)$ tensor defined by

$$p_{i_1,\ldots,i_m,j_1,\ldots,j_n} := f_{i_1,\ldots,i_m} g_{j_1,\ldots,j_n}, \tag{6.1}$$

or simply put,

$$p_{\mathcal{I}\mathcal{J}} = f_{\mathcal{I}} g_{\mathcal{J}} \tag{6.2}$$

where $\mathcal{I} \in [\![I_1]\!] \times \ldots \times [\![I_m]\!]$ and $\mathcal{J} \in [\![J_1]\!] \times \ldots \times [\![J_n]\!]$. The outer product of two column vectors $\mathbf{a} \in \mathbb{R}^m$ and $\mathbf{b} \in \mathbb{R}^n$ is the rank one matrix $\mathbf{a}\mathbf{b}^\top \in \mathbb{R}^{m \times n}$. The outer product for tensors can be calculated in a similar way if F and G are "vectorized" properly. With the conversion described in (2.6), the outer product can be calculated effectively in the following way.

```
% Matlab command numel(A) returns the number of elements in array A

f = reshape(F,numel(F),1);
g = reshape(G,1,numel(G));
p = f*g;
P = reshape(t,[size(F),size(G)]);
```

Regarding (6.2) as a factorization of P, we are curious about the converse of the outer product by. Given an order-$(m+n)$ tensor $P \in \mathbb{R}^{I_1 \times \ldots \times I_m \times J_1 \times \ldots \times J_n}$, when can it be factorized as the tensor product of two lower order tensors F and G of fixed m and n? There might not be an easy answer. If the factorization is not possible, then what is its best approximation as the tensor product of two lower order tensors F and G of fixed m and n? The latter is the minimization problem

$$\min_{F \in \mathbb{R}^{I_1 \times \ldots \times I_m}, G \in \mathbb{R}^{J_1 \times \ldots \times J_n}} \|P - F \otimes G\|_F. \tag{6.3}$$

For matrices, this is the classical problem of the best rank-1 approximation. The answer is known precisely. The absolute minimizer is given by left and right singular vectors associated with the largest singular value. For tensors, this is an interesting nonlinear approximation with fixed facets. We can solve it in a similar way which we outline below.

Since m and n are specified, define $\mathrm{c}(F) := \prod_{i=1}^m I_i$ and $\mathrm{c}(G) := \prod_{j=1}^n J_j$. These are the cardinalities of elements expected in F and G, respectively. Flatten the given P as a $\mathrm{c}(F) \times \mathrm{c}(G)$ matrix, denoted by $flat\_P$. There should be no confusion in this flattening process if we follow the rule specified in (2.6), that is, P is first recorded as a column vector and then reshaped into a matrix. Let $\mathbf{f} \in \mathbb{R}^{\mathrm{c}(F)}$ and $\mathbf{g} \in \mathbb{R}^{\mathrm{c}(G)}$ be the left and right singular vectors of $flat\_P$ corresponding to the largest singular value s. Define F and G by reshaping the columns $s\mathbf{f}$ and $\mathbf{g}$ into an order-m tensor in $\mathbb{R}^{I_1 \times \ldots \times I_m}$ and an order-n tensor in $\mathbb{R}^{J_1 \times \ldots \times J_n}$, respectively. In this way, the above procedure can be coded as follows.

```
% Given a tensor P and desirable m, n

sizeP = size(P);
sizeF = sizeP(1:m);
sizeG = sizeP(m+1:end);

totalP = prod(sizeP);
totalF = prod(sizeF);
```

```
totalG = prod(sizeG);

flat_P = reshape(P,totalF,totalG);

[f,s,g] = svds(flat_P,1);
F = s*reshape(f,sizeF);
G = reshape(g,sizeG);
```

LEMMA 6.1. *With F and G constructed in the way described above, $F \otimes G$ is the best outer production approximation to P.*

Proof. Let $\mathcal{I} = (i_1, \ldots, i_m) \in [\![I_1]\!] \times \ldots \times [\![I_m]\!]$ and $\mathcal{J} = (j_1, \ldots, j_n) \in [\![J_1]\!] \times \ldots \times [\![J_n]\!]$ denote the general multi-indices of size m and n, respectively. Then the corresponding linear indices are

$$i = i_1 + \sum_{s=1}^{m-1} (i_{s+1} - 1) \prod_{t=1}^{s} I_s,$$
$$j = j_1 + \sum_{\mu=1}^{n-1} (j_{\mu+1} - 1) \prod_{\nu=1}^{\mu} J_\mu$$

respectively. The (i,j)-entry of $flat\_P$ therefore should have the linear index

$$\begin{aligned}
\ell &= (j-1)\mathrm{c}(F) + i \\
&= \left(\left(j_1 + \sum_{\mu=1}^{n-1} (j_{\mu+1} - 1) \prod_{\nu=1}^{\mu} J_\mu \right) - 1 \right) \prod_{i=1}^{m} I_i + \left(i_1 + \sum_{s=1}^{m-1} (i_{s+1} - 1) \prod_{t=1}^{s} I_s \right) \\
&= \sum_{\mu=1}^{n-1} (j_{\mu+1} - 1) \prod_{\nu=1}^{\mu} J_\mu \prod_{t=1}^{s} I_s + (j_1 - 1) \prod_{t=1}^{s} I_s + \sum_{s=1}^{m-1} (i_{s+1} - 1) \prod_{t=1}^{s} I_s + i_1
\end{aligned}$$

which corresponds to the multi-index $(i_1, \ldots, i_m, j_1, \ldots, j_n)$. We therefore can write

$$\begin{aligned}
\|P - F \otimes G\|_F^2 &= \sum_{\mathcal{I}=(1,\ldots,1)}^{(I_1,\ldots,I_m)} \sum_{\mathcal{J}=(1,\ldots,1)}^{(J_1,\ldots,J_n)} |P_{\mathcal{I}\mathcal{J}} - F_{\mathcal{I}} G_{\mathcal{J}}|^2 \\
&= \sum_{i=1}^{\mathrm{c}(F)} \sum_{j=1}^{\mathrm{c}(G)} |(flat_P)_{ij} - s f_i g_j|^2 = \|flat_P - s\mathbf{f}\mathbf{g}^\top\|_F^2,
\end{aligned}$$

whereas the last quantity is optimal by the Eckart-Young Theorem. □

Question: Can this procedure be generalized to three or more factors? Indeed, if we can do it for three factors, then we can do it consecutively and obtain the best rank-1 tensor approximation.

7. Wedderburn rank-1 reduction formula. Given an arbitrary matrix $A \in \mathbb{R}^{m \times n}$, suppose that $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ are arbitrary vectors such that $\mathbf{y}^\top A\mathbf{x} \neq 0$. The Wedderburn rank-1 reduction formula asserting that the matrix

$$B := A - \frac{A\mathbf{x}\mathbf{y}^\top A}{\mathbf{y}^\top A\mathbf{x}} \tag{7.1}$$

has rank exactly one less than the rank of A plays a significant role in matrix factorization. It is demonstrated in [1] that perhaps all known matrix factorizations can be derived from such a formula with appropriately chosen $\mathbf{x}$ and $\mathbf{y}$. When considering a given tensor T as a linear transformation as we have discussed earlier, what should be the analogue of the rank reduction formula [4]?

To answer this question, we must first clarify the meaning of a rank. There is an array of various definitions for tensor ranks[2], so much so that the nomenclature of ranks associated to different fields might have entirely different meaning. Here we consider T as a linear transformation, so by the rank we refer to the dimension of the range space and the Fredholm alternative theorem holds. In this context, it is certain that this number is $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ dependent and is not the same as the so called tensor rank[3]. We first describe a specific combination of two tensor products.

LEMMA 7.1. *Give arbitrary order-m tensors $F \in \mathbb{R}^{I_1 \times \ldots \times I_m}$ and order-n tensor $G, H \in \mathbb{R}^{J_1 \times \ldots \times J_n}$, then with respect to $\boldsymbol{\beta} = \{1, \ldots m\}$ the identity*

$$(F \otimes G) \circledast_{\boldsymbol{\beta}} H = \langle G, H \rangle F \tag{7.2}$$

holds.

Proof. Recall that $P = F \otimes G$ is an order-$(m+n)$ tensor in $\mathbb{R}^{I_1 \times \ldots \times I_m \times J_1, \times \ldots, \times J_n}$. Considering P as the matrix representation of the linear transformation

$$\mathscr{P}_{\boldsymbol{\beta}} : \mathbb{R}^{J_1, \times \ldots, \times J_n} \to \mathbb{R}^{I_1 \times \ldots \times I_m},$$

we should have

$$P \circledast_{\boldsymbol{\beta}} H = \left[\sum_{j_1=1}^{J_1} \ldots \sum_{j_n=1}^{J_n} f_{i_1, \ldots, i_m} g_{j_1, \ldots, j_n} h_{j_1, \ldots, j_{sn}} \right] = \left[f_{\mathcal{I}} \sum_{\mathcal{J}} g_{\mathcal{J}} h_{\mathcal{J}} \right]. \tag{7.3}$$

The summation over $\mathcal{J}$ is precisely the Frobenious inner product $\langle G, H \rangle$. $\square$

We claim that the Wedderburn rank one reduction formula for tensors should be written in the following way.

LEMMA 7.2. *Given an order-k tensor $T \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_k}$, let it represent the linear transformation $\mathscr{T}_{\boldsymbol{\beta}} : \mathbb{R}^{I_{\alpha_1} \times \ldots \times I_{\alpha_s}} \to \mathbb{R}^{I_{\beta_1} \times \ldots \times I_{\beta_t}}$ with respect to a prescribed partitioning $[\![k]\!] = \boldsymbol{\alpha} \cup \boldsymbol{\beta}$. Let $T^\top$ denote the representation of the adjoint operator $\mathscr{T}_{\boldsymbol{\alpha}} : \mathbb{R}^{I_{\beta_1} \times \ldots \times I_{\beta_t}} \to \mathbb{R}^{I_{\alpha_1} \times \ldots \times I_{\alpha_s}}$. Suppose that $X \in \mathbb{R}^{I_{\alpha_1} \times \ldots \times I_{\alpha_s}}$ and $Y \in \mathbb{R}^{I_{\beta_1} \times \ldots \times I_{\beta_t}}$ are arbitrary tensors such that $\langle T \circledast_{\boldsymbol{\beta}} X, Y \rangle \neq 0$. Then, with respect to the same partitioning $[\![k]\!] = \boldsymbol{\alpha} \cup \boldsymbol{\beta}$, the linear transformation $\mathcal{S}_{\boldsymbol{\beta}}$ represented by the tensor*

$$S := T_{\boldsymbol{\beta}}^{\boldsymbol{\alpha}} - \frac{(T \circledast_{\boldsymbol{\beta}} X) \otimes (T^\top \circledast_{\boldsymbol{\alpha}} Y)}{\langle T \circledast_{\boldsymbol{\beta}} X, Y \rangle}. \tag{7.4}$$

has rank exactly one less than the rank of $\mathscr{T}_{\boldsymbol{\beta}}$.

Proof. We know that $F = T \circledast_{\boldsymbol{\beta}} X$ is in $\mathbb{R}^{I_{\beta_1} \times \ldots \times I_{\beta_t}}$ and $G = T^\top \circledast_{\boldsymbol{\alpha}} Y$ is in $\mathbb{R}^{I_{\alpha_1} \times \ldots \times I_{\alpha_s}}$. Since $X \in \mathbb{R}^{I_{\alpha_1} \times \ldots \times I_{\alpha_s}}$, it follows from Lemma 7.1 that $\mathscr{S}_{\boldsymbol{\beta}}(X) = S \circledast_{\boldsymbol{\beta}} X = 0$, showing that the operator $\mathscr{S}_{\boldsymbol{\beta}}$ has one extra "vector" in its null space than $\mathscr{T}_{\boldsymbol{\beta}}$. $\square$

What is interesting about (7.4) is not only about its semblance to (7.1), but also about the complexity of multiplications when high-dimensional data are involved. For example, when starting that the order-k tensor $(T \circledast_{\boldsymbol{\beta}} X) \otimes (T^\top \circledast_{\boldsymbol{\alpha}} Y)$ has a null space that is orthogonal to the order-s tensor $T^\top \circledast_{\boldsymbol{\alpha}} Y$, we must also state the null space is with respect to the operator $\circledast_{\boldsymbol{\beta}}$ from $\mathbb{R}^{I_{\alpha_1} \times \ldots \times I_{\alpha_s}}$ to $\mathbb{R}^{I_{\beta_1} \times \ldots \times I_{\beta_t}}$ and that orthogonality is with respect to the Frobenious inner product over $\mathbb{R}^{I_{\alpha_1} \times \ldots \times I_{\alpha_s}}$.

REFERENCES

[2] Including, for example, multilinear rank, outer product rank (also known as tensor rank, see footnote 3), border rank, Kruskal rank, symmetric tensor rank, and so on.

[3] An order-k tensor that is the outer product of k vectors is said to be simple. A tensor T is said to be of (tensor) rank r if r the smallest integer such that A is the sum of r simple tensors.

[1] M. T. Chu, R. E. Funderlic, and G. H. Golub, *A rank-one reduction formula and its applications to matrix factorizations*, SIAM Rev., 37 (1995), pp. 512–530.
[2] P. Comon, G. Golub, L.-H. Lim, and B. Mourrain, *Symmetric tensors and symmetric tensor rank*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1254–1279.
[3] S. Friedland, *Best rank one approximation of real symmetric tensors can be chosen symmetric*, Front. Math. China, 8 (2013), pp. 19–40.
[4] S. A. Goreinov, I. V. Oseledets, and D. V. Savostyanov, *Wedderburn rank reduction and Krylov subspace method for tensor approximation. Part 1: Tucker case*, SIAM J. Sci. Comput., 34 (2012), pp. A1–A27.
[5] T. G. Kolda, *Numerical optimization for symmetric tensor decomposition*, Math. Program., 151 (2015), pp. 225–248.
[6] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500.
[7] G. I. Marchuk, *Construction of adjoint operators in non-linear problems of mathematical physics*, Sbornik: Mathematics, 189 (1998), p. 1505.
[8] L. Qi, *Eigenvalues of a real supersymmetric tensor*, J. Symb. Comput., 40 (2005), pp. 1302–1324.
[9] L. Wang and M. T. Chu, *On the global convergence of the alternating least squares method for rank-one approximation to generic tensors*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 1058–1072.