BLOCK TENSORS AND SYMMETRIC EMBEDDINGS

STEFAN RAGNARSSON* AND CHARLES F. VAN LOAN†

Abstract. Well known connections exist between the singular value decomposition of a matrix A and the Schur decomposition of its symmetric embedding $\mathbf{sym}(A) = ([\,0\,A\,;\,A^T\,0\,])$. In particular, if σ is a singular value of A then $+\sigma$ and $-\sigma$ are eigenvalues of the symmetric embedding. The top and bottom halves of $\mathbf{sym}(A)$'s eigenvectors are singular vectors for A. Power methods applied to A can be related to power methods applied to $\mathbf{sym}(A)$. The rank of $\mathbf{sym}(A)$ is twice the rank of A. In this paper we develop similar connections for tensors by building on L-H. Lim's variational approach to tensor singular values and vectors. We show how to embed a general order-d tensor $\mathcal{A}$ into an order-d symmetric tensor $\mathbf{sym}(\mathcal{A})$. Through the embedding we relate power methods for $\mathcal{A}$'s singular values to power methods for $\mathbf{sym}(\mathcal{A})$'s eigenvalues. Finally, we connect the multilinear and outer product rank of $\mathcal{A}$ to the multilinear and outer product rank of $\mathbf{sym}(\mathcal{A})$.

Key words. tensor, block tensor, symmetric tensor, tensor rank

AMS subject classifications. 15A18, 15A69, 65F15

1. Introduction. If $A \in \mathbb{R}^{n_1 \times n_2}$, then there are well-known connections between its singular value decomposition (SVD) and the eigenvalue and eigenvector properties of the symmetric matrix

$$\mathbf{sym}(A) \;=\; \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}. \tag{1.1}$$

If $A = U\Sigma V^T$ is the SVD of A, then for $k = 1{:}\mathrm{rank}(A)$

$$\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} u_k \\ \pm v_k \end{bmatrix} \;=\; \pm\sigma_k \begin{bmatrix} u_k \\ \pm v_k \end{bmatrix} \tag{1.2}$$

where $u_k = U(:,k)$, $v_k = V(:,k)$, and $\sigma_k = \Sigma(k,k)$. Another way to connect A and $\mathbf{sym}(A)$ is through the Rayleigh quotients

$$\phi_A(u,v) \;=\; \frac{u^T A v}{\|\,u\,\|_2 \|\,v\,\|_2} \;=\; \left(\sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} A(i_1,i_2) u(i_1) v(i_2) \right) \Big/ \left(\|\,u\,\|_2 \|\,v\,\|_2 \right) \tag{1.3}$$

and

$$\phi_A^{(sym)}(x) \;=\; \frac{1}{2}\frac{x^T C x}{x^T x} \;=\; \frac{1}{2} \left(\sum_{i_1=1}^{N} \sum_{i_2=1}^{N} C(i_1,i_2) x(i_1) x(i_2) \right) \Big/ \|\,x\,\|_2^2 \tag{1.4}$$

where $u \in \mathbb{R}^{n_1}$, $v \in \mathbb{R}^{n_2}$, $N = n_1 + n_2$, $x \in \mathbb{R}^N$, and $C = \mathbf{sym}(A)$. If x is a stationary vector for $\phi_A^{(sym)}$, then $u = x(1{:}n_1)$ and $v = x(n_1+1{:}n_1+n_2)$ render a stationary value for ϕ_A. See [8, p.448].

In this paper we discuss these notions as they apply to tensors. An order-d tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is a real d-dimensional array $\mathcal{A}(1{:}n_1, \ldots, 1{:}n_d)$ where the index range in the k-th mode is from 1 to n_k. The idea of embedding a general tensor into a larger symmetric tensor having the same order is developed in §2. This requires

*Center for Applied Mathematics, Cornell University Ithaca, NY 14853, `str23@cornell.edu`

†Department of Computer Science, Cornell University, Ithaca, NY 14853, `cv@cs.cornell.edu`.
Both authors are supported in part by NSF contract DMS-1016284.

having a facility with block tensors. Fundamental orderings, unfoldings, and multilinear summations are discussed in §3 and used in §4 where we characterize various multilinear Rayleigh quotients and their stationary values and vectors. This builds on the variational approach to tensor singular values developed in [15]. In §5 we provide a symmetric embedding analysis of several higher-order power methods for tensors that have recently been proposed [10, 11, 5, 6, 13]. Results that relate the multilinear and outer product ranks of a tensor to the corresponding ranks of its symmetric embedding are presented in §6. A brief conclusion section follows.

Before we proceed with the rest of the paper, we use the case of third-order tensors to preview some of the main ideas and to establish notation. (The busy reader already familiar with basic tensor computations and notation may safely skip to §2.) The starting point is to define the trilinear Rayleigh quotient

$$\phi_{\mathcal{A}}(u,v,w) \;=\; \left(\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\sum_{i_3=1}^{n_3}\mathcal{A}(i_1,i_2,i_3)u(i_1)v(i_2)w(i_3)\right)\Big/\left(\parallel u\parallel_2\parallel v\parallel_2\parallel w\parallel_2\right) \tag{1.5}$$

where $\mathcal{A}\in\mathbb{R}^{n_1\times n_2\times n_3}$,$u\in\mathbb{R}^{n_1}$, $v\in\mathbb{R}^{n_2}$, and $w\in\mathbb{R}^{n_3}$. Calligraphic characters are used for tensors: $\mathcal{A}(i_1,i_2,i_3)$ is entry (i_1,i_2,i_3) of $\mathcal{A}$.

The singular values and vectors of $\mathcal{A}$ are the critical values and vectors of $\phi_{\mathcal{A}}$ as formulated in [15]. A simple expression for the gradient $\nabla\phi_{\mathcal{A}}$ is made possible by unfolding $\mathcal{A}=(a_{ijk})$ in each of its three modes and aggregating the u, v, and w vectors with the Kronecker product. To illustrate, suppose $n_1=4$, $n_2=3$, and $n_3=2$ and define the *modal unfoldings* $\mathcal{A}_{(1)}$, $\mathcal{A}_{(2)}$, and $\mathcal{A}_{(3)}$ by

$$\begin{aligned}
\mathcal{A}_{(1)} &= \begin{bmatrix} a_{111} & a_{121} & a_{131} & a_{112} & a_{122} & a_{132}\\ a_{211} & a_{221} & a_{231} & a_{212} & a_{222} & a_{232}\\ a_{311} & a_{321} & a_{331} & a_{312} & a_{322} & a_{332}\\ a_{411} & a_{421} & a_{431} & a_{412} & a_{422} & a_{432}\end{bmatrix}\\
\mathcal{A}_{(2)} &= \begin{bmatrix} a_{111} & a_{211} & a_{311} & a_{411} & a_{112} & a_{212} & a_{312} & a_{412}\\ a_{121} & a_{221} & a_{321} & a_{421} & a_{122} & a_{222} & a_{322} & a_{422}\\ a_{131} & a_{231} & a_{331} & a_{431} & a_{132} & a_{232} & a_{332} & a_{432}\end{bmatrix}\\
\mathcal{A}_{(3)} &= \begin{bmatrix} a_{111} & a_{211} & a_{311} & a_{411} & a_{121} & a_{221} & a_{321} & a_{421} & a_{131} & a_{231} & a_{331} & a_{431}\\ a_{112} & a_{212} & a_{312} & a_{412} & a_{122} & a_{222} & a_{322} & a_{422} & a_{132} & a_{232} & a_{332} & a_{432}\end{bmatrix}.
\end{aligned} \tag{1.6}$$

The columns of these matrices are *fibers*. A fiber of a tensor is obtained by fixing all but one of the indices. For example, the third column of the unfolding

$$\mathcal{A}_{(1)} \;=\; \left[\;\mathcal{A}(:,1,1)\quad\mathcal{A}(:,2,1)\quad\mathcal{A}(:,3,1)\quad\mathcal{A}(:,1,2)\quad\mathcal{A}(:,2,2)\quad\mathcal{A}(:,3,2)\;\right]$$

is the fiber

$$\mathcal{A}(:,3,1) \;=\; \begin{bmatrix}\mathcal{A}(1,3,1)\\ \mathcal{A}(2,3,1)\\ \mathcal{A}(3,3,1)\\ \mathcal{A}(4,3,1)\end{bmatrix}$$

obtained by fixing the 2-mode index at 3 and the 3-mode index at 1. It is necessary to specify the order in which the fibers appear in a modal unfolding. The choice exhibited in (1.6) has the property that

$$\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\sum_{i_3=1}^{n_3}\mathcal{A}(i_1,i_2,i_3)u(i_1)v(i_2)w(i_3) \;=\; \begin{cases} u^T\,\mathcal{A}_{(1)}\; w\otimes v \\ v^T\,\mathcal{A}_{(2)}\; w\otimes u \\ w^T\mathcal{A}_{(3)}\; v\otimes u \end{cases} \tag{1.7}$$

which makes it easy to specify the stationary vectors of $\phi_{\mathcal{A}}$. If u, v, and w are unit vectors, then the gradient of $\phi_{\mathcal{A}}$ is given by

$$\nabla\phi_{\mathcal{A}}(u,v,w) \;=\; \begin{bmatrix} \mathcal{A}_{(1)}\; w\otimes v \\ \mathcal{A}_{(2)}\; w\otimes u \\ \mathcal{A}_{(3)}\; v\otimes u \end{bmatrix} \;-\; \phi_{\mathcal{A}}(u,v,w)\begin{bmatrix} u \\ v \\ w \end{bmatrix}. \tag{1.8}$$

We remark that if $\mathcal{A}$ is an order-2 tensor, then (1.8) collapses to the familiar matrix-SVD equations $Av = \sigma u$ and $A^T u = \sigma v$.

A central contribution of this paper revolves around the tensor version of the **sym** matrix (1.1) and the associated Rayleigh quotient $\phi_{\mathcal{A}}^{(sym)}$ that is defined in (1.4). Just as **sym**-of-a-matrix sets up a symmetric block matrix whose entries are either zero or matrix transpositions, **sym**-of-a-tensor sets up a symmetric block tensor whose entries are either zero or a tensor transposition.

If $\mathcal{A}\in\mathbb{R}^{n_1\times n_2\times n_3}$, then there are $6 = 3!$ possible transpositions identified by the notation $\mathcal{A}^{<\,[i\;j\;k]\,>}$ where $[i\;j\;k]$ is a permutation of $[1\;2\;3]$:

$$\mathcal{B} \;=\; \left\{\begin{array}{c} \mathcal{A}^{<\,[1\;2\;3]\,>} \\ \mathcal{A}^{<\,[1\;3\;2]\,>} \\ \mathcal{A}^{<\,[2\;1\;3]\,>} \\ \mathcal{A}^{<\,[2\;3\;1]\,>} \\ \mathcal{A}^{<\,[3\;1\;2]\,>} \\ \mathcal{A}^{<\,[3\;2\;1]\,>} \end{array}\right\} \quad\Longrightarrow\quad \left\{\begin{array}{c} \mathcal{B}(i,j,k) \\ \mathcal{B}(i,k,j) \\ \mathcal{B}(j,i,k) \\ \mathcal{B}(j,k,i) \\ \mathcal{B}(k,i,j) \\ \mathcal{B}(k,j,i) \end{array}\right\} \;=\; \mathcal{A}(i,j,k) \tag{1.9}$$

for $i = 1{:}n_1,\; j = 1{:}n_2,\; k = 1{:}n_3$.

The symmetric embedding of a 3rd-order tensor results in a 3-by-3-by-3 block tensor, a kind of Rubik's cube built from 27 (possibly non-cubical) boxes. If $\mathcal{A}\in\mathbb{R}^{n_1\times n_2\times n_3}$ and $N = n_1+n_2+n_3$, then $\textbf{sym}(\mathcal{A}) \;=\; C\in\mathbb{R}^{N\times N\times N}$ is the 3-by-3-by-3 block tensor whose ijk block is specified by

$$\mathcal{C}_{[i\;j\;k]} \;=\; \begin{cases} \mathcal{A}^{<\,[i\;j\;k]\,>} & \text{if } [i\;j\;k] \text{ is a permutation of } [1\;2\;3] \\ 0\in\mathbb{R}^{n_i\times n_j\times n_k} & \text{otherwise.} \end{cases} \tag{1.10}$$

See Fig 1.1. The blocks in a block tensor such as $\mathcal{C}$ can be specified using the colon notation. For example, if $n_1 = 4$, $n_2 = 3$ and $n_3 = 2$, then

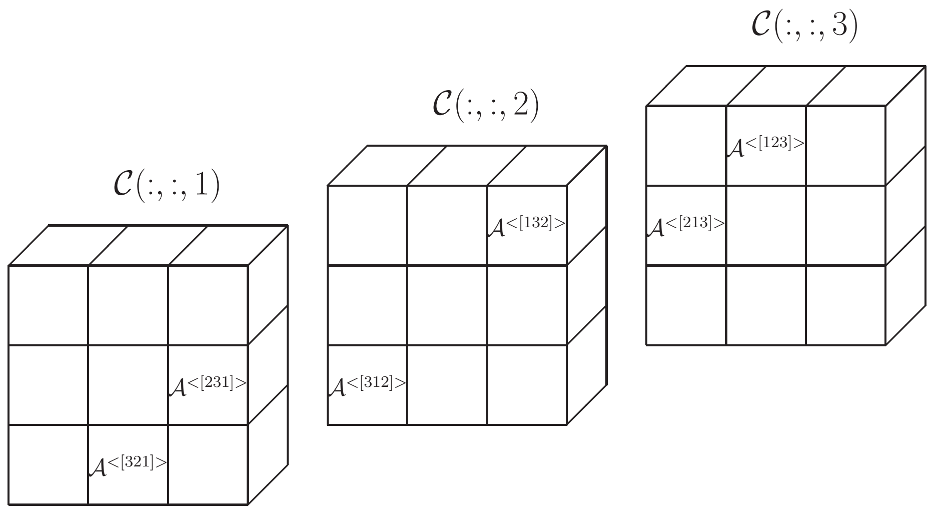

FIG. 1.1. *The Symmetric Embedding of an Order-3 Tensor*

$$\begin{array}{lllll} \mathcal{C}_{[1\,2\,3]} & = & \mathcal{C}(1{:}4, 5{:}7, 8{:}9) & = & \mathcal{A}^{<[1\,2\,3]>} \in \mathbb{R}^{n_1\times n_2\times n_3} \\ \mathcal{C}_{[1\,3\,2]} & = & \mathcal{C}(1{:}4, 8{:}9, 5{:}7) & = & \mathcal{A}^{<[1\,3\,2]>} \in \mathbb{R}^{n_1\times n_3\times n_2} \\ \mathcal{C}_{[2\,1\,3]} & = & \mathcal{C}(5{:}7, 1{:}4, 8{:}9) & = & \mathcal{A}^{<[2\,1\,3]>} \in \mathbb{R}^{n_2\times n_1\times n_3} \\ \mathcal{C}_{[2\,3\,1]} & = & \mathcal{C}(5{:}7, 8{:}9, 1{:}4) & = & \mathcal{A}^{<[2\,3\,1]>} \in \mathbb{R}^{n_2\times n_3\times n_1} \\ \mathcal{C}_{[3\,1\,2]} & = & \mathcal{C}(8{:}9, 1{:}4, 5{:}7) & = & \mathcal{A}^{<[3\,1\,2]>} \in \mathbb{R}^{n_3\times n_1\times n_2} \\ \mathcal{C}_{[3\,2\,1]} & = & \mathcal{C}(8{:}9, 5{:}7, 1{:}4) & = & \mathcal{A}^{<[3\,2\,1]>} \in \mathbb{R}^{n_3\times n_2\times n_1} \end{array} . \tag{1.11}$$

We will prove in section 2.3 that the tensor $\mathcal{C}$ is in fact symmetric.

The last topic to cover in our order-3 preview is the generalization of the Rayleigh quotient $\phi_{\mathcal{A}}^{(sym)}$ defined in (1.4). If $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$, $\mathcal{C} = \mathbf{sym}(\mathcal{A})$, $N = n_1 + n_2 + n_3$, and $x \in \mathbb{R}^N$, then $\phi_{\mathcal{A}}^{(sym)}$ is defined by

$$\phi_{\mathcal{A}}^{(sym)}(x) \;=\; \frac{1}{3!}\left(\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{i_3=1}^{N}\mathcal{C}(i_1,i_2,i_3)x(i_1)x(i_2)x(i_3)\right)\Big/ \parallel x \parallel_2^3 \tag{1.12}$$

It will be shown in section 4.3 that if

$$x \;=\; \left[\begin{array}{c} u \\ v \\ w \end{array}\right] \begin{array}{l} \}n_1 \\ \}n_2 \\ \}n_3 \end{array}$$

satisfies $\nabla\phi_{\mathcal{A}}^{(sym)}(x) = 0$, then

$$\nabla_u\,\phi_{\mathcal{A}}(u,v,w) \;=\; 0 \qquad \nabla_v\,\phi_{\mathcal{A}}(u,v,w) \;=\; 0 \qquad \nabla_w\,\phi_{\mathcal{A}}(u,v,w) \;=\; 0$$

where ∇_z refers to the gradient with respect to the components in vector z. Moreover, it will be shown that

$$x_{+-} = \begin{bmatrix} u \\ v \\ -w \end{bmatrix} \qquad x_{-+} = \begin{bmatrix} u \\ -v \\ w \end{bmatrix} \qquad x_{--} = \begin{bmatrix} u \\ -v \\ -w \end{bmatrix}$$

are also stationary vectors for $\phi_{\mathcal{A}}^{(sym)}$ and

$$\phi_{\mathcal{A}}(u,v,w) \;=\; \phi_{\mathcal{A}}^{(sym)}(x) \;=\; \phi_{\mathcal{A}}^{(sym)}(x_{--}) \;=\; -\phi_{\mathcal{A}}^{(sym)}(x_{-+}) \;=\; -\phi_{\mathcal{A}}^{(sym)}(x_{+-}).$$

2. The Symmetric Embedding. Block matrix manipulation is such a fixture in numerical linear algebra that we take for granted the correctness of facts like

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^T = \begin{bmatrix} A_{11}^T & A_{21}^T \\ A_{12}^T & A_{22}^T \end{bmatrix}. \tag{2.1}$$

Formal verification requires showing that the (i,j) entries on both sides of the equation are equal for all valid ij pairs.

The symmetric embedding of a tensor involves generalizations of both transposition and blocking so this section begins by discussing these notions and establishing the tensor analog of (2.1). Since vectors of subscripts are prominent in the presentation, we elevate their notational status with boldface font, e.g., $\mathbf{p} = [\,4\,1\,2\,3\,]$. We let $\mathbf{1}$ denote the vector of ones and assume that dimension is clear from context. More generally, if N is an integer, then $\mathbf{N}$ is the vector of all N's. Finally, if $\mathbf{i}$ and $\mathbf{j}$ have equal length, then $\mathbf{i} \le \mathbf{j}$ means that $i_k \le j_k$ for all k.

2.1. Blocking. If s and t are integers with $s \le t$, then (as in MATLAB) let $s{:}t$ denote the row vector $[s, s+1, \cdots, t]$. We refer to a vector with this form as an *index range vector*. The act of blocking an m_1-by-m_2 matrix C is the act of partitioning the index range vectors $1{:}m_1$ and $1{:}m_2$:

$$\mathbf{r}^{(1)} \;=\; 1{:}m_1 \;=\; \left[\; \mathbf{r}_1^{(1)} \;\middle|\; \cdots \;\middle|\; \mathbf{r}_{b_1}^{(1)} \;\right] \qquad \mathbf{r}^{(2)} \;=\; 1{:}m_2 \;=\; \left[\; \mathbf{r}_1^{(2)} \;\middle|\; \cdots \;\middle|\; \mathbf{r}_{b_2}^{(2)} \;\right] \tag{2.2}$$

Given (2.2), we are able to regard C as a $b_1 \times b_2$ block matrix (C_{i_1,i_2}) where block C_{i_1,i_2} has $\texttt{length}(\mathbf{r}_{i_1}^{(1)})$ rows and $\texttt{length}(\mathbf{r}_{i_2}^{(2)})$ columns. It is easy (although messy) to "locate" a particular entry of a particular block. Indeed,

$$C_{i_1,i_2}(j_1,j_2) \;=\; C(\,\rho_{i_1}^{(1)} + j_1\,,\; \rho_{i_2}^{(2)} + j_2\,)$$

where

$$\rho_{i_k}^{(k)} \;=\; \texttt{length}(\mathbf{r}_1^{(k)}) + \texttt{length}(\mathbf{r}_2^{(k)}) + \cdots + \texttt{length}(\mathbf{r}_{i_k-1}^{(k)}) \tag{2.3}$$

for $k = 1{:}2$.

To block an order-d tensor $\mathcal{C} \in \mathbb{R}^{m_1 \times \cdots \times m_d}$ we proceed analogously. The index-range vectors $1{:}m_1, \ldots, 1{:}m_d$ are partitioned

$$\mathbf{r}^{(k)} \;=\; 1{:}m_k \;=\; \left[\; \mathbf{r}_1^{(k)} \;\middle|\; \cdots \;\middle|\; \mathbf{r}_{b_k}^{(k)} \;\right] \qquad k = 1{:}d \tag{2.4}$$

and this permits us to regard $\mathcal{C}$ as a $b_1 \times \cdots \times b_d$ block tensor. If $\mathbf{i} = [i_1, \ldots, i_d]$, then the $\mathbf{i}$-th block is the subtensor

$$\mathcal{C}_{\mathbf{i}} \;=\; \mathcal{C}_{i_1,\ldots,i_d} \;=\; \mathcal{C}(\mathbf{r}_{i_1}^{(1)}, \ldots, \mathbf{r}_{i_d}^{(d)}).$$

If $\mathbf{j} = [j_1, \ldots, j_d]$, then the $\mathbf{j}$-th entry of this subtensor is given by

$$\mathcal{C}_{\mathbf{i}}(\mathbf{j}) \;=\; \mathcal{C}(\,\rho^{(1)}_{i_1} + j_1\,, \ldots,\, \rho^{(d)}_{i_d} + j_d) \in \mathbb{R} \tag{2.5}$$

where $\rho^{(k)}_{i_k}$ is specified by (2.3) for $k = 1{:}d$.

To illustrate equations (2.3)-(2.5), if $\mathcal{C} \in \mathbb{R}^{9\times 7\times 5\times 6}$ and

$$\begin{array}{lcll} 1{:}9 &=& \left[\begin{array}{cccc|cc|ccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9\end{array}\right] & (b_1 = 3)\\ 1{:}7 &=& \left[\begin{array}{ccccc|cc} 1 & 2 & 3 & 4 & 5 & 6 & 7\end{array}\right] & (b_2 = 2)\\ 1{:}5 &=& \left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 5\end{array}\right] & (b_3 = 2)\\ 1{:}6 &=& \left[\begin{array}{cc|cc|cc} 1 & 2 & 3 & 4 & 5 & 6\end{array}\right] & (b_4 = 3)\end{array},$$

then we are choosing to regard $\mathcal{C}$ as a $3 \times 2 \times 2 \times 3$ block tensor. Thus, if $\mathbf{i} = [3\,1\,2\,1]$ then $\mathcal{C}_{\mathbf{i}} = \mathcal{C}(7{:}9, 1{:}5, 5{:}5, 1{:}2)$ and

$$\mathcal{C}_{\mathbf{i}}(\mathbf{j}) \;=\; \mathcal{C}(6 + j_1\,,\, j_2\,,\, 4 + j_3\,,\, j_4)$$

where $\mathbf{1} \le \mathbf{j} \le [3\,5\,1\,2]$.

2.2. Tensor Transposition. If $\mathcal{A} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ and $\mathbf{p} = [p_1, \ldots, p_d]$ is a permutation of $1{:}d$, then $\mathcal{A}^{<\mathbf{p}>} \in \mathbb{R}^{n_{p_1}\times\cdots\times n_{p_d}}$ denotes the $\mathbf{p}$-transpose of $\mathcal{A}$ defined by

$$\mathcal{A}^{<\mathbf{p}>}(j_{p_1}, \ldots, j_{p_d}) \;=\; \mathcal{A}(j_1, \ldots, j_d)$$

where $1 \le j_k \le n_k$ for $k = 1{:}d$. A more succinct way of saying the same thing is

$$\mathcal{A}^{<\mathbf{p}>}(\mathbf{j}(\mathbf{p})) = \mathcal{A}(\mathbf{j}) \qquad \mathbf{1} \le \mathbf{j} \le \mathbf{n}.$$

If $\mathcal{A}$ is an order-2 tensor, then $\mathcal{A}^{<\,[2\;1]\,>}(j_2, j_1) = \mathcal{A}(j_1, j_2)$. It is also easy to verify that if $\mathbf{f}$ and $\mathbf{g}$ are both permutations of $1{:}d$, then

$$(\mathcal{A}^{<\mathbf{f}>})^{<\mathbf{g}>} \;=\; \mathcal{A}^{<\mathbf{f}(\mathbf{g})>}. \tag{2.6}$$

A transposition of a block tensor renders another block tensor. The following lemma makes this precise and generalizes (2.1).

Lemma 2.1. *Suppose $\mathcal{C} \in \mathbb{R}^{m_1\times\cdots\times m_d}$ is a $b_1 \times \cdots \times b_d$ block tensor with block dimensions defined by the partitioning (2.4). Let $\mathcal{C}_{\mathbf{i}}$ denote its $\mathbf{i}$-th block where $\mathbf{i} = [i_1, \ldots, i_d]$. If $\mathbf{p} = [p_1, \ldots, p_d]$ is a permutation of $1{:}d$ and $\mathcal{B} = \mathcal{C}^{<\mathbf{p}>}$, then the tensor $\mathcal{B} \in \mathbb{R}^{m_{p_1}\times\cdots\times m_{p_d}}$ is a $b_{p_1} \times \cdots \times b_{p_d}$ block tensor where each block $\mathcal{B}_{\mathbf{i}(\mathbf{p})}$ is defined by $\mathcal{B}_{\mathbf{i}(\mathbf{p})} = \mathcal{C}_{\mathbf{i}}^{<\mathbf{p}>}$.*

Proof. If $1 \le j_k \le m_k$ for $k = 1{:}d$, then from (2.4) and (2.5) we have

$$\mathcal{C}_{\mathbf{i}}^{<\mathbf{p}>}(j_{p_1}, \ldots, j_{p_d}) \;=\; \mathcal{C}_{\mathbf{i}}(j_1, \ldots, j_p) \;=\; \mathcal{C}(\rho^{(1)}_{i_1} + j_1, \ldots, \rho^{(d)}_{i_d} + j_d)$$

On the other hand, $\mathcal{B} = \mathcal{C}^{<\mathbf{p}>}$ and so

$$\mathcal{C}(\rho^{(1)}_{i_1} + j_1, \ldots, \rho^{(d)}_{i_d} + j_d) \;=\; \mathcal{B}(\rho^{(p_1)}_{i_{p_1}} + j_{p_1}, \ldots, \rho^{(p_d)}_{i_{p_d}} + j_{p_d}) \;=\; \mathcal{B}_{\mathbf{i}(\mathbf{p})}(j_{p_1}, \ldots, j_{p_d}).$$

Thus, $\mathcal{B}_{\mathbf{i}(\mathbf{p})}(\mathbf{j}(\mathbf{p})) \;=\; \mathcal{C}_{\mathbf{i}}^{<\mathbf{p}>}(\mathbf{j}(\mathbf{p}))$ for all $\mathbf{j}$, i.e., $\mathcal{B}_{\mathbf{i}(\mathbf{p})} = \mathcal{C}_{\mathbf{i}}^{<\mathbf{p}>}$. ☐

2.3. The sym(·) Operation. An order-d tensor $\mathcal{C} \in \mathbb{R}^{N\times\cdots\times N}$ is *symmetric* if $\mathcal{C} = \mathcal{C}^{<\mathbf{p}>}$ for any permutation $\mathbf{p}$ of $1{:}d$. The tensor analog of (1.1) involves constructing an order-d symmetric tensor $\mathbf{sym}(\mathcal{A})$ whose blocks are either zero or carefully chosen transposes of $\mathcal{A}$. In particular, if $\mathcal{A} \in \mathbb{R}^{n_1\times\cdots\times n_d}$, then

$$\mathbf{sym}(\mathcal{A}) \in \mathbb{R}^{N\times\cdots\times N} \qquad N = n_1 + \cdots n_d$$

is a block tensor defined by the partitioning $1{:}N = [\, \mathbf{r}_1 \mid \cdots \mid \mathbf{r}_d \,]$ where

$$\mathbf{r}_k = (1 + n_1 + \cdots + n_{k-1}){:}(n_1 + \cdots + n_k) \qquad k = 1{:}d. \tag{2.7}$$

The $\mathbf{i}$-th block of $\mathcal{C} = \mathbf{sym}(\mathcal{A})$ is given by

$$\mathcal{C}_{\mathbf{i}} = \begin{cases} A^{<\mathbf{i}>} & \text{if } \mathbf{i} \text{ is a permutation of } 1{:}d \\ \\ 0 & \text{otherwise} \end{cases}$$

for all $\mathbf{i}$ that satisfy $\mathbf{1} \le \mathbf{i} \le \mathbf{d}$. Note that $\mathcal{C}_{\mathbf{i}}$ is $n_{i_1} \times n_{i_2} \times \cdots \times n_{i_d}$. We confirm that $\mathbf{sym}(\mathcal{A})$ is symmetric.

LEMMA 2.2. *If $\mathcal{A} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ and $\mathcal{C} = \mathbf{sym}(\mathcal{A})$, then $\mathcal{C}$ is symmetric.*

Proof. Let $\mathbf{p}$ be an arbitrary permutation of $1{:}d$. We must show that if $\mathcal{B} = C^{<\mathbf{p}>}$ then $\mathcal{B} = \mathcal{C}$. Since $\mathcal{C}$ as a block tensor is $d\times d\times\cdots\times d$, it follows from Lemma 2.1 that $\mathcal{B}$ has the same block structure and

$$\mathcal{B}_{\mathbf{i}(\mathbf{p})} = \mathcal{C}_{\mathbf{i}}^{<\mathbf{p}>}$$

for all $\mathbf{i}$ that satisfy $\mathbf{1} \le \mathbf{i} \le \mathbf{d}$. If $\mathbf{i}$ is a permutation of $1{:}d$, then $\mathcal{C}_{\mathbf{i}} = A^{<\mathbf{i}>}$ and by using (2.6) we conclude that

$$\mathcal{B}_{\mathbf{i}(\mathbf{p})} = (\mathcal{A}^{<\mathbf{i}>})^{<\mathbf{p}>} = \mathcal{A}^{<\mathbf{i}(\mathbf{p})>} = \mathcal{C}_{\mathbf{i}(\mathbf{p})}$$

If $\mathbf{i}$ is *not* a permutation of $1{:}d$, then both $\mathcal{C}_{\mathbf{i}}$ and $\mathcal{C}_{\mathbf{i}(\mathbf{p})}$ are zero and so

$$\mathcal{B}_{\mathbf{i}(\mathbf{p})} = \mathcal{C}_{\mathbf{i}}^{<\mathbf{p}>} = 0 = \mathcal{C}_{\mathbf{i}(\mathbf{p})}.$$

Since $\mathcal{B}$ and $\mathcal{C}$ agree block-by-block, they are the same. ☐

3. Orderings, Unfoldings, and Summations. In numerical multilinear algebra it is frequently necessary to reshape a given tensor into a vector or a matrix and vice versa. In this section we collect results that make these maneuvers precise.

3.1. The col Ordering. If $\mathbf{i}$ and $\mathbf{s}$ are length-e index vectors and $\mathbf{1} \le \mathbf{i} \le \mathbf{s}$, then we define the integer-valued function *ivec* by

$$ivec(\mathbf{i}, \mathbf{s}) = i_1 + (i_2 - 1)s_1 + \cdots + (i_e - 1)(s_1 \cdots s_{e-1})$$

If $\mathcal{F} \in \mathbb{R}^{s_1\times\cdots\times s_e}$, then $v = \mathbf{vec}(\mathcal{F}) \in \mathbb{R}^{s_1\cdots s_e}$ is the column vector defined by

$$v(ivec(\mathbf{i}, \mathbf{s})) = \mathcal{F}(\mathbf{i}) \qquad \mathbf{1} \le \mathbf{i} \le \mathbf{s}.$$

Note that if $e = 2$, then $\mathcal{F}$ is a matrix and $\mathbf{vec}(\mathcal{F})$ stacks its columns. We also observe that if $w_k \in \mathbb{R}^{s_k}$ for $k = 1{:}e$, then

$$w = w_e \otimes \cdots \otimes w_1 \qquad \Leftrightarrow \qquad w(ivec(\mathbf{i}, \mathbf{s})) = w_1(i_1) \cdots w_e(i_e). \tag{3.1}$$

3.2. Modal Unfoldings. In the gradient calculations that follow, it is particularly convenient to "flatten" the given tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ into a matrix. If

$$\tilde{\mathbf{n}} = [\ \mathbf{n}(1{:}k-1) \quad \mathbf{n}(k+1{:}d)\ ], \tag{3.2}$$

$$\tilde{\mathbf{i}} = [\ \mathbf{i}(1{:}k-1) \quad \mathbf{i}(k+1{:}d)\ ], \tag{3.3}$$

then the *mode-k unfolding* $\mathcal{A}_{(k)}$ is defined by

$$\mathcal{A}_{(k)}(i_k, ivec(\tilde{\mathbf{i}}, \tilde{\mathbf{n}})) \;=\; \mathcal{A}(\mathbf{i}) \qquad \mathbf{1} \le \mathbf{i} \le \mathbf{n}. \tag{3.4}$$

This matrix has n_k rows and $n_1 \cdots n_{k-1} n_{k+1} \cdots n_d$ columns. A third-order instance of this important concept is displayed in equation (1.6). We mention that there are other ways to order the columns in $\mathcal{A}_{(k)}$. See [14].

While the columns of $\mathcal{A}_{(k)}$ are mode-k fibers, its rows are reshapings of its *mode-k subtensors*. In particular, if $1 \le r \le n_k$, then

$$\mathcal{A}_{(k)}(r,:) \;=\; \mathbf{vec}(\mathcal{B}^{(r)})^T$$

where the mode-k subtensor $\mathcal{B}^{(r)}$ has order $d-1$ and is defined by

$$\mathcal{B}^{(r)}(i_1, \ldots, i_{k-1}, i_{k+1}, \ldots, i_d) \;=\; \mathcal{A}(i_1, \ldots, i_{k-1}, r, i_{k+1}, \ldots, i_d).$$

The partitioning of an order-d tensor into order-$(d-1)$ tensors is just a generalization of partitioning a matrix into its columns.

3.3. Summations. It is handy to have a multi-index summation notation in order to describe general versions of the summations that appear in (1.5) and (1.12). If $\mathbf{n}$ is a length-d index vector, then

$$\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \;\equiv\; \sum_{i_1=1}^{n_1} \cdots \sum_{i_d=1}^{n_d}.$$

The summation that defines the multilinear Rayleigh quotient (1.5) can be written in matrix-vector terms.

LEMMA 3.1. *If $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ and $u_k \in \mathbb{R}^{n_k}$ for $k = 1{:}d$, then*

$$\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \mathcal{A}(\mathbf{i}) u_1(i_1) \cdots u_d(i_d) \;=\; \mathbf{vec}(\mathcal{A})^T u_d \otimes \cdots \otimes u_1. \tag{3.5}$$

Moreover, for $k = 1{:}d$ we have

$$\sum_{\mathbf{i}=1}^{\mathbf{n}} \mathcal{A}(\mathbf{i}) u_1(i_1) \cdots u_d(i_d) \;=\; u_k^T \, \mathcal{A}_{(k)} \, \tilde{u}_k \tag{3.6}$$

where

$$\tilde{u}_k \;=\; (u_d \otimes \cdots \otimes u_{k+1} \otimes u_{k-1} \otimes \cdots \otimes u_1). \tag{3.7}$$

Proof. If $a = \mathbf{vec}(\mathcal{A})$ and $b = u_d \otimes \cdots \otimes u_1$, then using the definition of **vec** and equations (3.1)-(3.4), we have

$$\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \mathcal{A}(\mathbf{i}) u_1(i_1) \cdots u_d(i_d) \;=\; \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} a(ivec(\mathbf{i},\mathbf{n})) \cdot b(ivec(\mathbf{i},\mathbf{n})) \;=\; \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} a(\mathbf{i}) b(\mathbf{i}) \;=\; a^T b.$$

This proves (3.5). Using the modal subtensor interpretation of $\mathcal{A}_{(k)}$ that we discussed in §3.2 and definitions (3.2) and (3.3), we have

$$\begin{aligned}\sum_{\mathbf{i}=1}^{\mathbf{n}} \mathcal{A}(\mathbf{i}) u_1(i_1) \cdots u_d(i_d) &= \sum_{i_k=1}^{n_k} u_k(i_k) \left(\sum_{\tilde{\mathbf{i}}=\mathbf{1}}^{\tilde{\mathbf{n}}} \mathcal{B}^{(i_k)}(\tilde{\mathbf{i}}) \tilde{u}(\tilde{\mathbf{i}}) \right) \\ &= \sum_{i_k=1}^{n_k} u_k(i_k) \left(\mathcal{A}_{(k)}(i_k, :) \tilde{u}_k \right) \;=\; u_k^T \mathcal{A}_{(k)} \tilde{u}_k\end{aligned}$$

which establishes (3.6). ☐

Summations that involve symmetric tensors are important in later sections. The following notation for the multiple Kronecker product of a single vector is handy:

$$x^{\otimes d} = \underbrace{x \otimes \cdots \otimes x}_{d \text{ times}}.$$

Note that if $x \in \mathbb{R}^N$, then $x^{\otimes d} \in \mathbb{R}^{N^d}$.

Lemma 3.2. *If $\mathcal{C} \in \mathbb{R}^{N \times \cdots \times N}$ is a symmetric order-d tensor and $x \in \mathbb{R}^N$, then*

$$\sum_{\mathbf{i}=1}^{\mathbf{N}} \mathcal{C}(\mathbf{i}) x(i_1) \cdots x(i_d) \;=\; x^T \, \mathcal{C}_{(1)} \, x^{\otimes(d-1)} \tag{3.8}$$

Proof. This follows from Lemma 3.1 by setting $n_k = N$ and $u_k = x$ for $k = 1{:}d$. Note that because $\mathcal{C}$ is symmetric, $\mathcal{C}_{(1)} = \cdots = \mathcal{C}_{(d)}$. ☐

The summation (3.8) has a special characterization if $\mathcal{C} = \mathbf{sym}(\mathcal{A})$. To pursue this we will have to navigate $\mathcal{C}$'s block structure and to that end we define the index vectors **L** and **R** as follows:

$$\mathbf{L} \;=\; \begin{bmatrix} 1 \\ n_1 + 1 \\ \vdots \\ n_1 + \cdots + n_{d-1} + 1 \end{bmatrix} \qquad \mathbf{R} \;=\; \begin{bmatrix} n_1 \\ n_1 + n_2 \\ \vdots \\ n_1 + \cdots + n_d \end{bmatrix}. \tag{3.9}$$

Note that if $\mathbf{1} \le \mathbf{p} \le \mathbf{d}$, then

$$\mathcal{C}_{\mathbf{p}} \;=\; \mathcal{C}(\mathbf{L}(p_1){:}\mathbf{R}(p_1), \ldots, \mathbf{L}(p_d){:}\mathbf{R}(p_d))$$

is $\mathcal{C}$'s **p**-th block.

Lemma 3.3. *Suppose $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, $\mathcal{C} \;=\; \mathbf{sym}(\mathcal{A})$, and $N = n_1 + \cdots + n_d$. If $x \in \mathbb{R}^N$ is partitioned as follows*

$$x \;=\; \begin{bmatrix} u_1 \\ \vdots \\ u_d \end{bmatrix} \qquad u_k \in \mathbb{R}^{n_k},$$

and $\tilde{u}_1, \ldots, \tilde{u}_d$ *are defined by (3.7), then*

$$\mathcal{C}_{(1)} x^{\otimes(d-1)} \;=\; (d-1)! \begin{bmatrix} \mathcal{A}_{(1)}\tilde{u}_1 \\ \vdots \\ \mathcal{A}_{(d)}\tilde{u}_d \end{bmatrix} \tag{3.10}$$

and

$$\sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{N}} \mathcal{C}(\mathbf{j}) x(j_1)\cdots x(j_d) \;=\; d! \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \mathcal{A}(\mathbf{i}) u_1(i_1)\cdots u_d(i_d). \tag{3.11}$$

Proof. If $v = \mathcal{C}_{(1)} x^{\otimes(d-1)}$ and

$$e_j = I_N(:,j) \;=\; \begin{bmatrix} w_1 \\ \vdots \\ w_d \end{bmatrix} \tag{3.12}$$

is partitioned conformally with x, then for $j = 1{:}N$ we have

$$\begin{aligned} v(j) &= \sum_{\mathbf{i}(2:d)=\mathbf{1}}^{\mathbf{N}} \mathcal{C}(j, i_2, \ldots, i_d) x(i_2)\cdots x(i_d) \\ &= \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{N}} \mathcal{C}(\mathbf{i}) e_j(i_1) x(i_2)\cdots x(i_d) \\ &= \sum_{\mathbf{p}=\mathbf{1}}^{\mathbf{d}} \sum_{\mathbf{i}=\mathbf{L}(\mathbf{p})}^{\mathbf{R}(\mathbf{p})} \mathcal{C}(\mathbf{i}) e_j(i_1) x(i_2)\cdots x(i_d) \\ &= \sum_{\mathbf{p}=\mathbf{1}}^{\mathbf{d}} \left(\sum_{\mathbf{k}=\mathbf{1}}^{\mathbf{n}(\mathbf{p})} \mathcal{C}_{\mathbf{p}}(\mathbf{k}) w_{p_1}(k_1) u_{p_2}(k_2)\cdots u_{p_d}(k_d) \right) \end{aligned}$$

Now suppose that $\mathbf{L}(q) \le j \le \mathbf{R}(q)$, $j = \mathbf{L}(q) + r - 1$. From (3.10) we must show that v_j is the rth component of $\mathcal{A}_{(q)}\tilde{u}_q$.

To that end observe that $\mathcal{C}_{\mathbf{p}}(\mathbf{k}) w_{p_1}(k_1)$ is necessarily zero unless $p_1 = q$, $k_1 = r$, and $\mathbf{p}$ is a permutation of $1{:}d$. Assuming this to be the case and defining the vectors $v_1, \ldots, v_d$ by

$$v_i \;=\; \begin{cases} u_i & \text{if } i \neq q \\ w_q & \text{otherwise} \end{cases},$$

we see using (3.6) that

$$\begin{aligned}\sum_{\mathbf{k}=\mathbf{1}}^{\mathbf{n}(\mathbf{p})} \mathcal{C}_{\mathbf{p}}(\mathbf{k}) w_{p_1}(k_1) u_{p_2}(k_2) \cdots u_{p_d}(k_d) &= \sum_{\mathbf{k}=\mathbf{1}}^{\mathbf{n}(\mathbf{p})} \mathcal{A}^{<p>}(\mathbf{k}) v_{p_1}(k_1) v_{p_2}(k_2) \cdots v_{p_d}(k_d) \\ &= \sum_{\mathbf{k}=\mathbf{1}}^{\mathbf{n}} \mathcal{A}(\mathbf{k}) v_1(k_1) v_2(k_2) \cdots v_d(k_d) \\ &= v_q^T \mathcal{A}_{(q)} v_d \otimes \cdots \otimes v_{q+1} \otimes v_{q-1} \otimes \cdots \otimes v_1 \\ &= w_q^T \mathcal{A}_{(q)} u_d \otimes \cdots \otimes u_{q+1} \otimes u_{q-1} \otimes \cdots \otimes u_1 \\ &= w_q^T \mathcal{A}_{(q)} \tilde{u}_q .\end{aligned}$$

Observe that the number of $\mathbf{p}$ that satisfy $\mathbf{1} \le \mathbf{p} \le \mathbf{d}$ subject to the constraint $p_1 = q$ is $(d-1)!$ and conclude from (3.12) that $w_q = I_{n_q}(:, r)$. It follows that

$$v(j) \;=\; \sum_{\mathbf{p}\,=\,\mathbf{1}}^{\mathbf{d}} w_q^T \mathcal{A}_{(q)} \tilde{u}_q \;=\; (d-1)! \left[\mathcal{A}_{(q)} \tilde{u}_q \right]_r .$$

This establishes (3.10). Equation (3.11) follows from

$$x^T \mathcal{C}_{(1)} x^{\otimes (d-1)} \;=\; \sum_{k=1}^{d} (d-1)! u_k^T \mathcal{A}_{(k)} \tilde{u}_k$$

and Lemmas 3.1 and 3.2. ▯

4. Rayleigh Quotients and Stationary Values.

Suppose $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ and $u_k \in \mathbb{R}^{n_k}$ for $k = 1{:}d$. Analogous to (1.3) we define the multilinear Rayleigh Quotient

$$\phi_{\mathcal{A}}(u_1, \ldots, u_d) \;=\; \left(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \mathcal{A}(\mathbf{i}) u_1(i_1) \cdots u_d(i_d) \right) \Bigg/ \left(\| u_1 \|_2 \cdots \| u_d \|_2 \right) \tag{4.1}$$

If $\mathcal{C} = \mathbf{sym}(\mathcal{A})$, $N = n_1 + \cdots n_d$, and $x \in \mathbb{R}^N$, then corresponding to (1.4) we have

$$\phi_{\mathcal{A}}^{(sym)}(x) \;=\; \frac{1}{d!} \left(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{N}} \mathcal{C}(\mathbf{i}) x(i_1) \cdots x(i_d) \right) \Bigg/ \left(\| x \|_2 \right)^d \tag{4.2}$$

In this section we examine these multilinear Rayleigh quotients, specify their gradients, and relate the singular values of $\mathcal{A}$ to the eigenvalues of $\mathbf{sym}(\mathcal{A})$.

4.1. The Singular Values of a General Tensor.

The gradient of $\phi_{\mathcal{A}}(u_1, \ldots, u_d)$ relates to a collection of matrix-vector products that involve the modal unfoldings of $\mathcal{A}$ and Kronecker products of the u-vectors.

Theorem 4.1. *If $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ and for $k = 1{:}d$ the vectors $u_k \in \mathbb{R}^{n_k}$ each have unit 2-norm, then*

$$\nabla \phi_{\mathcal{A}}(u_1, \ldots, u_d) \;=\; \begin{bmatrix} \mathcal{A}_{(1)} \tilde{u}_1 \\ \vdots \\ \mathcal{A}_{(d)} \tilde{u}_d \end{bmatrix} \;-\; \phi_{\mathcal{A}}(u_1, \ldots, u_d) \cdot \begin{bmatrix} u_1 \\ \vdots \\ u_d \end{bmatrix}$$

where $\tilde{u}_k = (u_d \otimes \cdots \otimes u_{k+1} \otimes u_{k-1} \otimes \cdots \otimes u_1)$.

Proof. From Lemma 3.1 we have

$$\phi_{\mathcal{A}}(u_1,\ldots,u_d) = \left(u_k^T\, \mathcal{A}_{(k)}\, \tilde{u}_k\right) / \left(\| u_1 \|_2 \cdots \| u_d \|_2\right).$$

For $k = 1{:}d$ we have $\nabla_{u_k}\left(u_k^T\, \mathcal{A}_{(k)}\, \tilde{u}_k\right) = \mathcal{A}_{(k)}\, \tilde{u}_k$ and $\nabla_{u_k}(\| u_1 \|_2 \cdots \| u_d \|_2) = u_k$. and so

$$\nabla_{u_k}\phi_{\mathcal{A}} = \frac{(\| u_1 \|_2 \cdots \| u_d \|_2)\mathcal{A}_{(k)}\tilde{u}_k \;-\; (u_k^T\mathcal{A}_{(k)}\tilde{u}_k)u_k}{(\| u_1 \|_2 \cdots \| u_d \|_2)^2}$$

$$= \mathcal{A}_{(k)}\tilde{u}_k \;-\; \phi_{\mathcal{A}}(u_1,\ldots,u_d)u_k.$$

The theorem follows by simply "stacking" these subvectors of the gradient. □

The variational approach to tensor singular values and vectors set forth in [15] is based on equating the gradient of $\phi_{\mathcal{A}}$ to zero.

Definition 4.2. *The scalar* $\sigma \in \mathbb{R}$ *is a* singular value *of a general tensor* $\mathcal{A} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ *if there are unit vectors* $u_k \in \mathbb{R}^{n_k}$ *such that*

$$\mathcal{A}_{(k)}\tilde{u}_k = \sigma u_k, \tag{4.3}$$

for $k = 1{:}d$*. The vector* u_k *is the* mode-k singular vector *associated with* σ*.*

The normalization condition $u_k^T u_k = 1$ is necessary, since if $v_k = a u_k$ for $k = 1{:}d$ then $\mathcal{A}_{(k)}\tilde{v}_k = a^{d-1}\mathcal{A}_{(k)}\tilde{u}_k = a^{k-1}\sigma u_k = (a^{k-2}\sigma)v_k$ for any $a \in \mathbb{R}$. It can be shown that at least one singular value and associated singular vectors exist for any tensor (cf. [15]).

4.2. The Eigenvalues of a Symmetric Tensor. For a symmetric tensor $\mathcal{C}$, the stationary values of $\phi_{\mathcal{C}}(x,\ldots,x)$ define the notion of a tensor eigenvalue.

Theorem 4.3. *If* $\mathcal{C} \in \mathbb{R}^{N\times\cdots\times N}$ *is symmetric and* $x \in \mathbb{R}^N$ *has unit norm, then*

$$\nabla_x\phi_{\mathcal{C}}(x,\ldots,x) = d\left(\mathcal{C}_{(1)}x^{\otimes(d-1)} \;-\; \phi_{\mathcal{C}}(x,\ldots,x)x\right).$$

Proof. From Lemma 3.2 we have

$$\phi_{\mathcal{C}}(x,\ldots,x) = x^T\mathcal{C}_{(1)}x^{\otimes(d-1)} / \| x \|^d.$$

Since

$$\nabla_x x^T\mathcal{C}_{(1)}x^{\otimes(d-1)} = d\mathcal{C}_{(1)}x^{\otimes(d-1)}$$

and

$$\nabla_x (x^Tx)^{d/2} = d(x^Tx)^{d/2-1}x$$

it follows that

$$\begin{aligned}\nabla_x\phi_{\mathcal{C}}(x,\ldots,x) &= d\,\frac{(x^Tx)^{d/2}\mathcal{C}_{(1)}x^{\otimes(d-1)} \;-\; \left(x^T\mathcal{C}_{(1)}x^{\otimes(d-1)}\right)(x^Tx)^{d/2-1}x}{(x^Tx)^d}\\ &= d\left(\mathcal{C}_{(1)}x^{\otimes(d-1)} \;-\; \left(x^T\mathcal{C}_{(k)}x^{\otimes(d-1)}\right)x\right)\\ &= d\left(\mathcal{C}_{(1)}x^{\otimes(d-1)} \;-\; \phi_{\mathcal{C}}(x,\ldots,x)x\right)\end{aligned}$$

completing the proof of the theorem. □

By setting the gradient of $\phi_{\mathcal{C}}(x,\ldots,x)$ to zero we arrive at the notion of a tensor eigenvalue [18].

Definition 4.4. *If $\mathcal{C} \in \mathbb{R}^{N\times\cdots\times N}$ is symmetric and $x \in \mathbb{R}^N$ is a unit vector such that*

$$\mathcal{C}_{(1)} x^{\otimes(d-1)} \;=\; \lambda x \tag{4.4}$$

then $\lambda = \phi_{\mathcal{C}}(x,\ldots,x)$ is an eigenvalue of $\mathcal{C}$ and x the associated eigenvector.

Note that if $\mathcal{C}_{(1)} x^{\otimes(d-1)} \;=\; \lambda x$ and $\alpha \in \mathbb{R}$, then $\mathcal{C}_{(1)}(\alpha x)^{\otimes(d-1)} \;=\; (\alpha^{d-2}\lambda)(\alpha x)$. Thus, we resolve a uniqueness issue by requiring tensor eigenvectors to have unit length, something that is not necessary in the matrix $(d = 2)$ case.

In [18, 19] it is shown that eigenvalues and associated eigenvectors always exist for symmetric tensors. Recently it has been shown that a symmetric tensor has at most $((d-1)^N - 1)/(d-2)$ eigenvalues, counted with multiplicity [1].

4.3. The Eigenvalues of sym($\mathcal{A}$). Since $\mathcal{C} = \textbf{sym}(\mathcal{A})$ is so structured, we anticipate that the eigenvalue-defining equation $\nabla\phi_{\mathcal{A}}^{(sym)}(x) = 0$ will have some special features. From the definitions (4.1) and (4.2) and Theorem 4.3, we have

$$\nabla\phi_{\mathcal{A}}^{(sym)}(x) \;=\; \frac{1}{d!}\nabla\phi_{\mathcal{C}}(x,\ldots,x) \;=\; \frac{1}{(d-1)!}\Big(\mathcal{C}_{(1)}x^{\otimes(d-1)} \;-\; \phi_{\mathcal{C}}(x,\ldots,x)x\Big) \tag{4.5}$$

We first characterize the gradient of $\phi_{\mathcal{A}}^{(sym)}$ in terms of matrix-vector products that involve $\mathcal{A}$'s modal unfoldings.

Theorem 4.5. *If $\mathcal{A} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ and x has unit 2-norm, then*

$$\nabla_x\phi_{\mathcal{A}}^{(sym)}(x) \;=\; \begin{bmatrix} \mathcal{A}_{(1)}\tilde{u}_1 \\ \vdots \\ \mathcal{A}_{(d)}\tilde{u}_d \end{bmatrix} \;-\; d\begin{bmatrix} (u_1^T\mathcal{A}_{(1)}\tilde{u}_1)u_1 \\ \vdots \\ (u_d^T\mathcal{A}_{(d)}\tilde{u}_d)u_d \end{bmatrix}. \tag{4.6}$$

Proof. From Lemmas 3.1 and 3.3 and the definitions (4.1) and 4.2) we have

$$\begin{aligned}\phi_{\mathcal{A}}^{(sym)}(x) &= \frac{1}{d!}\left(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{N}}\mathcal{C}(\mathbf{i})x(i_1)\cdots x(i_d)\right)\Big/\left(\|\, x\,\|_2\right)^d \\ &= \left(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}}\mathcal{A}(\mathbf{i})u_1(i_1)\cdots u_(i_d)\right)\Big/\left(\|\, x\,\|_2\right)^d \\ &= \frac{u_k^T\mathcal{A}_{(k)}\tilde{u}_k}{(u_1^Tu_1 + \cdots + u_d^Tu_d)^{d/2}}\end{aligned}$$

for $k = 1{:}d$. Since

$$\nabla_{u_k}(u_1^Tu_1 \;+\; \cdots +\; u_d^Tu_d)^{d/2} \;=\; d\cdot(u_1^Tu_1 \;+\; \cdots +\; u_d^Tu_d)^{d/2-1}u_k \;=\; d\cdot u_k.$$

Since $x^Tx = u_1^Tu_1 + \cdots u_d^Tu_d$, we can conclude that

$$\begin{aligned}\nabla_{u_k}\phi_{\mathcal{A}}^{(sym)}(x) &= \frac{(x^Tx)^{d/2}\mathcal{A}_{(k)}\tilde{u}_k - d\cdot(u_k^T\mathcal{A}_{(k)}\tilde{u}_k)(x^Tx)^{d/2-1}u_k}{(x^Tx)^d} \\ &= \mathcal{A}_{(k)}\tilde{u}_k \;-\; d\cdot(u_k^T\mathcal{A}_{(k)}\tilde{u}_k)u_k\end{aligned}$$

completing the proof of the theorem □

It turns out that if the gradient of $\phi_{\mathcal{A}}^{(sym)}(x)$ is zero, then the vector x generally has the property that each subvector u_k has the same norm.

COROLLARY 4.6. *If* $\nabla\phi_{\mathcal{A}}^{(sym)}(x) = 0$ *and* $x^Tx = 1$, *then either* $\mathcal{A}_{(k)}\tilde{u}_k = 0$ *for* $k = 1{:}d$ *or*

$$\begin{bmatrix} \mathcal{A}_{(1)}\tilde{u}_1 \\ \vdots \\ \mathcal{A}_{(d)}\tilde{u}_d \end{bmatrix} = d \cdot \phi_{\mathcal{A}}^{(sym)}(x) \begin{bmatrix} u_1 \\ \vdots \\ u_d \end{bmatrix}$$

and $\| u_1 \|_2 = \| u_2 \|_2 = \cdots = \| u_d \|_2 = 1/\sqrt{d}$.

Proof. Since $\nabla\phi_{\mathcal{A}}^{(sym)}(x) = 0$, we know from Theorem 4.5 that

$$\mathcal{A}_{(k)}\tilde{u}_k = d \cdot (u_k^T\mathcal{A}_{(k)}\tilde{u}_k)u_k$$

for $k = 1{:}d$. Thus,

$$u_k^T\mathcal{A}_{(k)}\tilde{u}_k = d \cdot (u_k^T\mathcal{A}_{(k)}\tilde{u}_k)(u_k^Tu_k).$$

From Lemma 3.1, if $u_k^T\mathcal{A}_{(k)}\tilde{u}_k = 0$ for some k, then it is zero for all k. In this case we conclude from (4.6) that $\mathcal{A}_{(k)}\tilde{u}_k = 0$ for $k = 1{:}d$. Otherwise, $1 = du_k^Tu_k$, $k = 1{:}d$. It follows that $\| u_1 \|_2 = \cdots = \| u_d \|_2 = 1/\sqrt{d}$. □

We are now ready for the main result that relates the eigenvalues and vectors of $\mathbf{sym}(\mathcal{A})$ to the singular values and vectors of $\mathcal{A}$.

THEOREM 4.7. *If* σ *is a nonzero singular value of* $\mathcal{A} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ *with unit modal singular vectors* $u_1, \ldots, u_d$, *then*

$$x_\alpha = \frac{1}{\sqrt{d}} \begin{bmatrix} \alpha_1u_1 \\ \alpha_2u_2 \\ \vdots \\ \alpha_du_d \end{bmatrix} \qquad \alpha = [1, \pm1, \ldots, \pm1]$$

is an eigenvector for $\mathbf{sym}(\mathcal{A})$ *corresponding to eigenvalue*

$$\lambda_\alpha = \alpha_1\alpha_2\cdots\alpha_d \frac{d!}{\sqrt{d^d}}\sigma.$$

Note that α_1 *is set to +1 to resolve a uniqueness issue. See discussion after definition 4.4 and also equation (1.2) for the matrix case.*

Proof. We must show that $g = \nabla\phi_{\mathcal{A}}^{(sym)}(x_\alpha) = 0$. If $\tau = \alpha_1\cdots\alpha_d$ then for $k = 1{:}d$ we have from (4.6) that

$$g_k = \frac{\tau}{\alpha_kd^{(d-1)/2}}\mathcal{A}_k\tilde{u}_k - d\frac{\alpha_k\tau}{d^{(d+1)/2}}\left(u_k^T\mathcal{A}_{(k)}\tilde{u}_k\right)u_k$$

But since $\sigma = u_k^T\mathcal{A}_k\tilde{u}_k$ and $\mathcal{A}_k\tilde{u}_k = \sigma u_k$, we have

$$g_k = \alpha_k\tau\left(\frac{1}{d^{(d-1)/2}} - \frac{1}{d^{(d-1)/2}}\right)u_k = 0.$$

Since $\lambda_\alpha \;=\; x_\alpha^T \mathcal{C}_{(1)} x_\alpha^{\otimes(d-1)}$ we have from Lemma 3.3 that

$$\lambda_\alpha = \left(\frac{1}{\sqrt{d}} \begin{bmatrix} \alpha_1 u_1 \\ \vdots \\ \alpha_d u_d \end{bmatrix} \right)^T \left(\frac{(d-1)!}{d^{(d-1)/2}} \begin{bmatrix} (\tau/\alpha_1)\mathcal{A}_{(1)}\tilde{u}_1 \\ \vdots \\ (\tau/\alpha_d)\mathcal{A}_{(d)}\tilde{u}_d \end{bmatrix} \right)$$

$$= \frac{1}{\sqrt{d}} \cdot \frac{(d-1)!}{d^{(d-1)/2}} \cdot \tau \cdot \sum_{k=1}^{d} u_k^T \mathcal{A}_{(k)} \tilde{u}_k \;=\; \frac{(d-1)!}{\sqrt{d^d}} \cdot \tau \cdot \sum_{k=1}^{d} \sigma \;=\; \frac{d!}{\sqrt{d^d}} \cdot \tau \cdot \sigma$$

completing the proof of the theorem. □

Thus, for each singular value and vector for $\mathcal{A}$ we have 2^{d-1} eigenvalue/eigenvector pairs for $\mathbf{sym}(\mathcal{A})$.

4.4. Connections to the Multilinear Transform. Suppose $\mathcal{F} \in \mathbb{R}^{s_1 \times \cdots \times s_d}$ and $B_k \in \mathbb{R}^{s_k \times t_k}$ for $k = 1{:}d$. The tensor $\mathcal{T} \in \mathbb{R}^{t_1 \times \cdots \times t_d}$ defined by

$$\mathcal{T}(\mathbf{i}) = \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{s}} \mathcal{F}(\mathbf{j}) B_1(j_1, i_1) B_2(j_2, i_2) \cdots B_k(j_k, i_k). \tag{4.7}$$

is the *multilinear transform* [7] of tensor $\mathcal{F}$ by the matrices $B_1, \ldots, B_d$ and is denoted by

$$\mathcal{T} \;=\; \mathcal{F} \cdot (B_1, B_2, \ldots, B_d). \tag{4.8}$$

We also define

$$(B_1, B_2, \ldots, B_d) \cdot \mathcal{F} \;\;\equiv\;\; \mathcal{F} \cdot (B_1^T, B_2^T, \ldots, B_d^T). \tag{4.9}$$

Some of the key summations and vectors above can be expressed neatly through this transformation. For example, if $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ and $u_k \in \mathbb{R}^{n_k}$ for $k = 1{:}d$, then

$$\mathcal{A} \cdot (u_1, \ldots, u_d) \;=\; \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \mathcal{A}(\mathbf{i}) u_1(i_1) \cdots u_d(i_d) = u_1^T \mathcal{A}_{(1)} \tilde{u}_1$$

and

$$\mathcal{A} \cdot (u_1, \ldots, u_{k-1}, I_{n_k}, u_{k+1}, \ldots, u_d) \;=\; \mathcal{A}_{(k)} \tilde{u}_k.$$

5. Higher Order Power Methods. We now briefly review various tensor power methods and consider them in light of the singular- and eigenvalue connection between $\mathcal{A}$ and $\mathbf{sym}(\mathcal{A})$.

5.1. The HOPM. The matrix power method method can be generalized to tensors by replacing the matrix-vector multipication with multilinear transforms. The *Higher-Order Power Method* of [5, 6] for finding a singular value and associated singular vectors of general order-d tensors proceeds in an alternating fashion to update each of the mode-j singular vectors u_j.

Different initial values for the u_j vectors will in general result in convergence to different singular values. See Section 5.4 for a discussion on popular choices for higher-order power method initial values.

Algorithm 1 The higher-order power method (HOPM) [5, 6]

Given an order-d tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$.

Require: $u_j^{(0)} \in \mathbb{R}^{n_j}$ with $\|u_j^{(0)}\|_2 = 1$. Let $\sigma^{(0)} = (u_1^{(0)})^T \mathcal{A}_{(1)} \tilde{u}_1^{(0)}$.

for $k = 0, 1, \ldots$ **do**
  for $j = 0, 1, \ldots$ **do**
    $\hat{u}_j^{(k+1)} \leftarrow \mathcal{A}_{(j)} u_d^{(k+1)} \otimes \cdots \otimes u_{j+1}^{(k+1)} \otimes u_{j-1}^{(k)} \otimes \cdots \otimes u_1^{(k)}$
    $u_j^{(k+1)} \leftarrow \hat{u}_j^{(k+1)} / \|\hat{u}_j^{(k+1)}\|_2$
  end for
  $\sigma^{(k+1)} \leftarrow (u_1^{(k+1)})^T \mathcal{A}_{(1)} \tilde{u}_1^{(k+1)}$
end for

The HOPM can also be viewed as a way of finding the best rank-1 tensor approximation $\hat{\mathcal{A}}$ to $\mathcal{A}$ [5]. Specifically, a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is said to be *rank-1* if for $k = 1{:}d$ there exist vectors $t_i \in \mathbb{R}^{n_i}$ such that for all $\mathbf{i} = \mathbf{1}, \ldots, \mathbf{n}$

$$\mathcal{T}(\mathbf{i}) = t_1(i_1) t_2(i_2) \cdots t_d(i_d) \tag{5.1}$$

and we then say that $\mathcal{T}$ is the *tensor outer product* of the vectors $t_1, \ldots, t_d$, denoted by

$$\mathcal{T} = t_1 \circ t_2 \circ \cdots \circ t_d. \tag{5.2}$$

It can be shown that the HOPM converges to a local minimum of the functional $f(\hat{\mathcal{A}}) \equiv \| \mathcal{A} - \hat{\mathcal{A}} \|_F^2$, where $\hat{\mathcal{A}} = \sigma\, u_1 \circ \cdots \circ u_d$ is a rank-1 approximation to $\mathcal{A}$ and the Frobenius norm of a tensor $\mathcal{T}$ is defined as $\| \mathcal{T} \|_F \equiv \sqrt{\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \mathcal{T}(\mathbf{i})^2}$. See [11].

The HOPM can be applied to an order-d symmetric $N \times \cdots \times N$ tensor, starting with a symmetric initial guess $u_1^{(0)} = u_2^{(0)} = \cdots = u_d^{(0)} \in \mathbb{R}^N$. The solution found by the algorithm will be symmetric but intermediate results may break symmetry. Indeed, after one iteration the u_j vectors will in general all be distinct, but $u_j^{(k)} \to u$ as $k \to \infty$ for some $u \in \mathbb{R}^N$ [5].

5.2. The S-HOPM. Recently, [10] investigated a modified version of the HOPM for symmetric tensors which was originally dismissed by [5] as unreliable since in general it is not guaranteed to converge. This algorithm is called the *Symmetric Higher Order Power Method* (S-HOPM) and converges for certain classes of symmetric tensors. For example, suppose $\mathcal{C}$ is a symmetric tensor of even order and that M is a square unfolding of $\mathcal{C}$. If M is semidefinite then the S-HOPM converges [10].

Algorithm 2 Symmetric higher-order power method (S-HOPM) [5, 10]

Given an order-d symmetric tensor $\mathcal{C} \in \mathbb{R}^{N \times \cdots \times N}$.

Require: $x^{(0)} \in \mathbb{R}^N$ with $\|x^{(0)}\|_2 = 1$. Let $\lambda^{(0)} = (x^{(0)})^T \mathcal{C}_{(1)} (x^{(0)})^{\otimes(d-1)}$.

for $k = 0, 1, \ldots$ **do**
  $\hat{x}^{(k+1)} \leftarrow \mathcal{C}_{(1)} (x^{(k)})^{\otimes(d-1)}$
  $x^{(k+1)} \leftarrow \hat{x}^{(k+1)} / \|\hat{x}^{(k+1)}\|_2$
  $\lambda^{(k+1)} \leftarrow (x^{(k+1)})^T \mathcal{C}_{(1)} (x^{(k+1)})^{\otimes(d-1)}$
end for

This approach avoids the awkward situation, mentioned previously, of encountering non-symmetric intermediate values when using the HOPM on a symmetric tensor.

Since $\mathbf{sym}(\mathcal{A})$ is symmetric for any tensor $\mathcal{A}$, the S-HOPM can be applied to $\mathcal{A}$ through its embedding. By using facts previously established, we can reduce all operations on $\mathbf{sym}(\mathcal{A})$ to equivalent ones on $\mathcal{A}$.

Algorithm 3 Symmetric higher-order power method on $\mathbf{sym}(\mathcal{A})$

Given an order-d tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$.

Require: $u_j^{(0)} \in \mathbb{R}^{n_j}$ with $\|u_j^{(0)}\|_2 = 1$. Let $\sigma^{(0)} = (u_1^{(0)})^T \mathcal{A}_{(1)} \tilde{u}_1^{(0)}$.

for $k = 0, 1, \ldots$ **do**
    for $j = 0, 1, \ldots$ **do**
        $\hat{u}_j^{(k+1)} \leftarrow \mathcal{A}_{(j)} \tilde{u}_j^{(k)}$
        $u_j^{(k+1)} \leftarrow \hat{u}_j^{(k+1)} / \|\hat{u}_j^{(k+1)}\|_2$
    end for
    $\sigma^{(k+1)} \leftarrow (u_1^{(k+1)})^T \mathcal{A}_{(1)} \tilde{u}_1^{(k+1)}$
end for

This algorithm computes a singular value σ for $\mathcal{A}$ and the mode-j singular vectors u_j. The normalization used in Algorithm 3 is slightly different than a direct application of the S-HOPM on $\mathbf{sym}(\mathcal{A})$ would imply; the S-HOPM would set $u_j^{(k+1)} = \hat{u}_j^{(k+1)} / \sqrt{\|\hat{u}_1^{(k+1)}\|_2^2 + \cdots + \|\hat{u}_d^{(k+1)}\|_2^2}$. However, numerical experiments suggest that using $u_j^{(k+1)} = \hat{u}_j^{(k+1)} / \|\hat{u}_j^{(k+1)}\|$ improves convergence. If $\mathcal{A}$ is itself symmetric, then Algorithm 3 reduces to the S-HOPM as all the u_j will be equal, assuming $u_1^{(0)} = \cdots = u_d^{(0)}$.

Note that Algorithm 3 is very similar to the regular HOPM except the most recently available information on $u_1, \ldots, u_{j-1}$ is not used when computing $u_j^{(k+1)}$ for $j > 1$. The difference between the HOPM and Algorithm 3 is thus somewhat like the difference between the Jacobi and Gauss-Seidel iterative linear system solvers [8].

Unlike the HOPM, Algorithm 3 does not always converge and since it can be shown that a square unfolding of $\mathbf{sym}(\mathcal{A})$ is indefinite unless all the entries in $\mathcal{A}$ are zero, the convergence criteria in [10] do not apply.

5.3. The SS-HOPM and $\mathbf{sym}(\cdot)$. Recently, Kolda and Mayo [13] developed a shifted version of the S-HOPM and proved that for a suitable choice of shift their algorithm will converge to an eigenpair (λ, x) for any symmetric tensor $\mathcal{C}$.

Algorithm 4 Shifted symmetric higher-order power method (SS-HOPM) [13]

Given an order-d symmetric tensor $\mathcal{C} \in \mathbb{R}^{N \times \cdots \times N}$.

Require: $x^{(0)} \in \mathbb{R}^N$ with $\|x^{(0)}\|_2 = 1$. Let $\lambda^{(0)} = (x^{(0)})^T \mathcal{C}_{(1)} (x^{(0)})^{\otimes(d-1)}$.

for $k = 0, 1, \ldots$ **do**
    $\hat{x}^{(k+1)} \leftarrow \mathcal{C}_{(1)} (x^{(k)})^{\otimes(d-1)} + \alpha_{\mathcal{C}} x^{(k)}$
    $x^{(k+1)} \leftarrow \hat{x}^{(k+1)} / \|\hat{x}^{(k+1)}\|_2$
    $\lambda^{(k+1)} \leftarrow (x^{(k+1)})^T \mathcal{C}_{(1)} (x^{(k+1)})^{\otimes(d-1)}$
end for

If the shift $\alpha_{\mathcal{C}}$ satisfies $|\alpha_{\mathcal{C}}| > (d-1) \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{N}} |\mathcal{C}(\mathbf{i})|$ then the SS-HOPM will converge to an eigenpair [13].

When $\mathcal{C} = \mathbf{sym}(\mathcal{A})$ the algorithm can be simplified and expressed in terms of operations on $\mathcal{A}$.

Algorithm 5 Shifted symmetric higher-order power method on $\mathbf{sym}(\mathcal{A})$

Given an order-d tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$.

Require: $u_j^{(0)} \in \mathbb{R}^{n_j}$ with $\|u_j^{(0)}\|_2 = 1/\sqrt{d}$. Let $\sigma^{(0)} = (u_1^{(0)})^T \mathcal{A}_{(1)} \tilde{u}_1^{(0)}$.

for $k = 0, 1, \ldots$ **do**
  for $j = 0, 1, \ldots$ **do**
    $\hat{u}_j^{(k+1)} \leftarrow \mathcal{A}_{(j)} \tilde{u}_j^{(k)} + d\, \alpha_{\mathcal{A}} u_j^{(k)}$
  end for
  for $j = 0, 1, \ldots$ **do**
    $u_j^{(k+1)} \leftarrow \hat{u}_j^{(k+1)} / \sqrt{\|\hat{u}_1^{(k+1)}\|_2^2 + \cdots + \|\hat{u}_d^{(k+1)}\|_2^2}$
  end for
  $\sigma^{(k+1)} \leftarrow (u_1^{(k+1)})^T \mathcal{A}_{(1)} \tilde{u}_1^{(k+1)}$
end for

Using Theorem 4.7, a simple normalization of the values returned by this algorithm gives a singular value and associated unit norm singular vectors of $\mathcal{A}$.

The shift $\alpha_{\mathcal{A}}$ must satisfy $|\alpha_{\mathcal{A}}| > (d-1)\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} |\mathcal{A}(\mathbf{i})|$ to guarantee convergence, although a smaller shift might be sufficient for any particular tensor $\mathcal{A}$.

Example. Let $\mathcal{A}$ be the $2 \times 2 \times 2 \times 2$ tensor given by the unfolding

$$\mathcal{A}_{(1)} = \begin{bmatrix} 1.1650 & 0.2641 & -0.6965 & 1.2460 & 0.0751 & -1.4462 & 0.0591 & 0.5774 \\ 0.6268 & 0.8717 & 1.6961 & -0.6390 & 0.3516 & -0.7012 & 1.7971 & -0.3600 \end{bmatrix}.$$

We ran 100 trials of the HOPM, Algorithm 3 and Algorithm 5 using different random starting points $u_i^{(0)}$ chosen from a uniform distribution on $[-1, 1]^{n_i}$ and suitably normalized for each algorithm. The algorithms are considered to have converged when $|\sigma^{(k+1)} - \sigma^{(k)}| < 10^{-16}$. For this example, all three algorithms converged for every starting point.

The HOPM found the singular values 2.7248 and 1.7960. Algorithm 3 converged to $\sigma = \pm 2.7248$. Algorithm 5 with a positive shift $\alpha_{\mathcal{A}}$ found 2.7248 and 1.7960 and using a negative shift produced the values -2.7248 and -1.7960.

For this tensor $\mathcal{A}$ the theory suggests a shift $\alpha_{\mathcal{A}}$ greater than 37.72 in absolute value to guarantee convergence. However, using $\alpha_{\mathcal{A}}$ as small as 1 will still lead to convergence and does so in many fewer iterations, sometimes by as much as a factor of 30 when compared to the suggested shift. Setting $\alpha_{\mathcal{A}}$ to zero caused the algorithm to fail to converge for all chosen starting points.

5.4. Initialization. A standard way to initialize higher-order power methods is to use a truncated form of the Higher-Order Singular Value Decomposition (HOSVD) of [4],

$$\mathcal{A} = (U_1, U_2, \ldots, U_d) \cdot \mathcal{S}. \tag{5.3}$$

where $\mathcal{S} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is the *core tensor*, the $U_i \in \mathbb{R}^{n_i \times n_i}$ are orthogonal and related to the modal unfoldings of $\mathcal{A}$ through the matrix SVD equations $\mathcal{A}_{(i)} = U_i \Sigma_i V_i^T$.

To initialize the HOPM, for example, the values $u_j^{(0)} = U_j(:, 1)$ have been shown [5] to often lie close to the best rank-1 approximation to $\mathcal{A}$.

If desired, it is possible to create the HOSVD of $\mathcal{C} = \mathbf{sym}(\mathcal{A})$ from the HOSVD of $\mathcal{A}$. For example, if $\mathcal{C} = (U_{\mathcal{C}}, \ldots, U_{\mathcal{C}}) \cdot \mathcal{S}_{\mathcal{C}}$ is the HOSVD of $\mathcal{C}$ then it can be shown that $U_{\mathcal{C}}$ is a column permutation of the block-diagonal matrix $\mathbf{diag}(U_1, U_2, \ldots, U_d)$.

There are many other ways to initialize tensor power methods. In [11] Regalia and Kofidis derive a procedure for symmetric tensors that can outperform the HOSVD-based approach.

Another possibility is to compute a tensor generalization of the QR decomposition with partial pivoting, of the form $\mathcal{A} = (Q_1, \ldots, Q_d) \cdot \mathcal{R}$ where $\mathcal{A}_{(k)} = Q_k R_k \Pi_k$ are the pivoted QR decompositions of the unfolding $\mathcal{A}_{(k)}$. It can be shown that this "HOQRD" decomposition retains some of the approximation properties of the truncated matrix pivoted QR decomposition and can thus give a reasonable initial guess for a tensor power method. As for the HOSVD, the HOQRD of $\mathbf{sym}(\mathcal{A})$ can be constructed from the HOQRD of $\mathcal{A}$.

6. Tensor Rank and the sym Operation. There are several definitions of tensor rank, each of which represents some reasonable generalization of matrix rank. For an excellent review see [7]. In this brief section we relate the multilinear rank and the outer product rank of $\mathbf{sym}(\mathcal{A})$ to the multilinear rank and the outer product rank of $\mathcal{A}$.

6.1. Multilinear Rank. The *multilinear rank* of $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is the d-tuple $\text{rank}_{\boxplus}(\mathcal{A}) = (r_1(\mathcal{A}), r_2(\mathcal{A}), \ldots, r_d(\mathcal{A}))$ where $r_i(\mathcal{A}) = \text{rank}(\mathcal{A}_{(i)})$. Note that if the tensor $\mathcal{C} \in \mathbb{R}^{N \times \cdots \times N}$ is symmetric, and $R = \text{rank}(\mathcal{C}_{(1)})$, then

$$\text{rank}_{\boxplus}(\mathcal{C}) \;=\; (R, R, \ldots, R) \tag{6.1}$$

because $\mathcal{C}_{(1)} = \mathcal{C}_{(2)} = \cdots = \mathcal{C}_{(d)}$. If $\mathcal{C} = \mathbf{sym}(\mathcal{A})$, then it is possible to connect $\text{rank}_{\boxplus}(\mathcal{C})$ to $\text{rank}_{\boxplus}(\mathcal{A})$.

THEOREM 6.1. *If* $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ *and* $\text{rank}_{\boxplus}(\mathcal{A}) = (r_1, \ldots, r_d)$, *then*

$$\text{rank}_{\boxplus}(\mathbf{sym}(\mathcal{A})) = (R, \ldots, R)$$

where $R = r_1 + \cdots + r_d$.

Proof. Suppose $\mathcal{C} = \mathbf{sym}(\mathcal{A})$ and $\mathcal{C}_{\mathbf{i}}$ is $\mathcal{C}$'s $\mathbf{i}$th block, $\mathbf{1} \le \mathbf{i} \le \mathbf{d}$. Let $\mathcal{C}^{(k)}$ be a $1 \times d \times \cdots \times d$ block tensor defined by

$$\mathcal{C}^{(k)}_{\mathbf{i}} \;=\; \mathcal{C}_{\mathbf{i}}$$

where $i_1 = k$ and $1 \le i_j \le d$ for $j = 2{:}d$. Note that if $\mathbf{i}(2{:}d)$ is a permutation of $[1{:}k-1 \;\; k+1{:}d]$, then

$$\mathcal{C}^{(k)}_{\mathbf{i}} \;=\; \mathcal{A}^{<[\,k\;\mathbf{i}(2:d)\,]>}.$$

It follows that

$$\text{range}(\mathcal{C}^{(k)}_{(1)}) = \text{range}(\mathcal{A}_{(k)}) \qquad k = 1{:}d \tag{6.2}$$

If

$$\mathcal{C}_{(1)} \;=\; \begin{bmatrix} C_1 \\ \vdots \\ C_d \end{bmatrix} \begin{matrix} \}\, n_1 \\ \vdots \\ \}\, n_d \end{matrix}$$

is a block row partitioning of $\mathcal{C}_{(1)}$, then C_k is a column permutation of $\mathcal{C}^{(k)}_{(1)}$ and so using (6.2) we have

$$\text{rank}(C_k) \;=\; \text{rank}(\mathcal{C}^{(k)}_{(1)}) \;=\; r_k. \tag{6.3}$$

If

$$v = \begin{bmatrix} v_1 \\ \vdots \\ v_d \end{bmatrix} \begin{matrix} \}\, n_1 \\ \vdots \\ \}\, n_d \end{matrix}$$

is a column of $\mathcal{C}_{(1)}$ then it is a a mode-1 fiber of $\mathcal{C}$ and thus can "pass through" at most one $\mathcal{C}$-block having an index that is a permutation of $1{:}d$. This means that at most one of v's subvectors is zero. It follows from (6.3) that

$$\text{rank}(\mathcal{C}_{(1)}) = \sum_{k=1}^{d} \text{rank}(C_k) = \sum_{k=1}^{d} r_k$$

completing the proof of the theorem. □

6.2. Outer Product Rank. The *outer product rank* of $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is the minimum number of rank-1 tensors, as defined in (5.1) and (5.2), that are needed to represent it as a sum

$$\text{rank}_{\otimes}(\mathcal{A}) \equiv \min \left\{ r : \mathcal{A} = \sum_{i=1}^{r} u_1^{(i)} \circ u_2^{(i)} \circ \cdots \circ u_d^{(i)}, \quad u_j^{(i)} \in \mathbb{R}^{n_j} \right\}.$$

For matrices $\text{rank}(\textbf{sym}(A)) = 2\ \text{rank}(A)$. Indeed, If $A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$ is the SVD of A,then

$$\textbf{sym}(A) = \sum_{i=1}^{r} \sigma_i \left(\begin{bmatrix} 0 \\ v_i \end{bmatrix} \begin{bmatrix} u_i^T & 0 \end{bmatrix} + \begin{bmatrix} u_i \\ 0 \end{bmatrix} \begin{bmatrix} 0 & v_i^T \end{bmatrix} \right).$$

Motivated by this expansion we make a definition.

Definition 6.2. *If $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is the rank-1 tensor $\mathcal{T} = t_1 \circ \cdots \circ t_d$ and $N = n_1 + \cdots n_d$, then $\mathcal{S} \in \mathbb{R}^{N \times \cdots \times N}$ is the rank-1 tensor*

$$\mathcal{S} = \pi(\mathcal{T}) = s_1 \circ \cdots \circ s_d$$

where

$$s_k = \begin{bmatrix} 0 \\ t_k \\ 0 \end{bmatrix} \begin{matrix} \}\, n_1 + \cdots + n_{k-1} \\ \}\, n_k \\ \}\, n_{k+1} + \cdots + n_d \end{matrix}$$

With this construction, we can produce an outer product expansion of $\textbf{sym}(\mathcal{A})$ given an outer product expansion of $\mathcal{A}$.

Theorem 6.3. *If $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ and*

$$\mathcal{A} = \sum_{i=1}^{r} u_i^{(1)} \circ u_i^{(2)} \circ \cdots \circ u_i^{(d)}$$

where $u_1^{(k)}, \ldots, u_r^{(k)} \in \mathbb{R}^{n_k}$, then

$$\textbf{sym}(\mathcal{A}) = \sum_{\mathbf{p} \in S_d} \sum_{i=1}^{r} \pi(u_i^{(1)} \circ u_i^{(2)} \circ \cdots \circ u_i^{(d)})^{<\mathbf{p}>} \qquad (6.4)$$

where S_d is the set of all permutations of $1{:}d$.

Proof. Let $\mathcal{C}$ be the sum on the right side of (6.4) and note that

$$\mathcal{C} \;=\; \sum_{\mathbf{p}\in S_d}\sum_{i=1}^{r}\pi(u_i^{(p_1)}\circ u_i^{(p_2)}\circ\cdots\circ u_i^{(p_d)}).$$

We must show that the $\mathbf{q}$th block of $\mathbf{sym}(\mathcal{A})$ equals the $\mathbf{q}$th block of $\mathcal{C}_{\mathbf{q}}$. If $\mathbf{q}$ is not a permutation of $1{:}d$, then these blocks are both zero. Otherwise

$$\mathcal{C}_{\mathbf{q}} \;=\; \sum_{i=1}^{r} u_i^{(q_1)}\circ u_i^{(q_2)}\circ\cdots\circ u_i^{(q_d)} \;=\; \left(\sum_{i=1}^{r} u_i^{(1)}\circ u_i^{(2)}\circ\cdots\circ u_i^{(d)}\right)^{<\mathbf{q}>} \;=\; \mathcal{A}^{<\mathbf{q}>}.$$

completing the proof of the theorem. ∎

Since the double summation in (6.4) involves $rd!$ terms, it follows that

$$\mathrm{rank}_{\otimes}(\mathbf{sym}(\mathcal{A})) \;\le\; d!\cdot\mathrm{rank}_{\otimes}(\mathcal{A}) \tag{6.5}$$

We conjecture that equality prevails. This is somewhat reminiscent of the direct sum conjecture [21], i.e. that $\mathrm{rank}_{\otimes}(\mathcal{A}\oplus\mathcal{B}) = \mathrm{rank}_{\otimes}(\mathcal{A})+\mathrm{rank}_{\otimes}(\mathcal{B})$. Intuitively, $\mathbf{sym}(\mathcal{A})$ contains $d!$ distinct copies of $\mathcal{A}$ in nonoverlapping index regions so if the matrix case were to generalize, any expansion of $\mathbf{sym}(\mathcal{A})$ into a sum of $\le d!r$ rank-1 terms could be reduced to (6.4) without adding terms, thus having exactly $d!r$ terms. We have so far been unable to prove this. Note that it can be shown that

$$d\cdot\mathrm{rank}_{\otimes}(\mathcal{A}) \;\le\; \mathrm{rank}_{\otimes}(\mathbf{sym}(\mathcal{A}))$$

using Lemma 3.5 in [7].

7. Conclusions. The symmetrization $\mathbf{sym}(\mathcal{A})$ can be used to connect algorithms for symmetric tensors and ones for general tensors. In this paper we have shown how algorithms such as the S-HOPM and SS-HOPM give rise to non-symmetric algorithms through the symmetrization in a way that preserves many convergence properties. In particular, the non-symmetric version of the SS-HOPM we derive is guaranteed to converge for an appropriately chosen shift $\alpha_{\mathcal{A}}$. Are there other tensor methods where the symmetrization could be used to spot new connections or derive useful algorithms?

The rank properties of the symmetrization in some ways mirror the matrix case, but fundamental questions regarding the outer product rank of $\mathbf{sym}(\mathcal{A})$ remain open. Resolution of these questions may help bridge the conceptual gap that exists between matrix rank and tensor rank.

8. Acknowledgements. The authors thank the referees for suggestions that led to an improved presentation.

REFERENCES

[1] D. Cartwright and B. Sturmfels, *The number of eigenvalues of a tensor*. arXiv:1004.4953v1, Apr. 2010.

[2] P. Comon, M. Sorensen, E. Tsigaridas, *Decomposing tensors with structured matrix factors reduces to rank-1 approximations*, Research report I3S/RR-2009-11, Sept. 8, 2009.

[3] P. Comon, *Tensors versus matrices : Usefulness and unexpected properties*, Statistical Signal Processing, 2009. SSP '09. IEEE/SP 15th Workshop on Statistical Signal Processing.

[4] L. De Lathauwer, B. De Moor and J. Vandewalle, *A multilinear singular value decomposition*, SIAM. J. Matrix Anal. & Appl. Vol. 21, No. 4 April 2000, pp. 1253-1278.

[5] L. De Lathauwer, B. de Moor, and J. Vandewalle, *On the best rank-1 and rank-*$(R_1, R_1, \ldots, R_N)$ *approximation of higher-order tensors*, SIAM J. Matrix Anal. Appl. Vol. 21, No. 4, pp. 1324-1342.

[6] L. De Lathauwer, P. Comom, B. de Moor, and J. Vandewalle, *Higher-order power method–Application in independent component analysis*, in proceedings of the International Symposium on Nonlinear Theory and Its Applications (NOLTA '95), Las Vegas, NV, 1995, pp. 91-96.

[7] V. De Silva, L.-H. Lim, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM Journal on Matrix Analysis and Applications, 30 (2008), no. 3, pp. 1084-1127.

[8] G. H. Golub, C. Van Loan, *Matrix Computations*. 3rd edition (1996), Johns Hopkins University Press.

[9] C. J. Hillar, L.-H. Lim, *Most tensor problems are NP hard.* arXiv:0911.1393v1 [cs.CC]

[10] E. Kofidis, P. A. Regalia, *On the best rank-1 approximation of higher-order supersymmetric tensors*, SIAM J. Matrix Anal. Appl. Vol. 23, pp. 863-884.

[11] E. Kofidis, P.A. Regalia. *The higher-order power method revisited: convergence proofs and effective initialization*. Proceedings of the Acoustics, Speech, and Signal Processing, 2000. on IEEE International Conference, Vol. 5, pp. 2709-2712.

[12] T. G. Kolda, *Multilinear operators for higher-order decompositions.* Technical Report Number SAND2006-2081, Sandia National Laboratories, Albuquerque, NM and Livermore, CA, April 2006.

[13] T. G. Kolda, J. R. Mayo, *Shifted power method for computing tensor eigenpairs.* arXiv:1007.1267v1, Jul. 2010.

[14] T. G. Kolda, B. W. Bader, *Algorithm 862: MATLAB Tensor Classes for Fast Algorithm Prototyping*, ACM Transactions on Mathematical Software, Vol. 32, No. 4, December 2006, pp. 635653.

[15] L.-H. Lim, *Singular values and eigenvalues of tensors : A variational approach*, Proceedings of the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 1 (2005), pp. 129-132.

[16] L. Qi, *Rank and eigenvalues of a supersymmetric tensor, the multivariate homogeneous polynomial and the algebraic hypersurface it defines*, Journal of Symbolic Computation 41 (2006) 1309-1327.

[17] L. Qi, *Eigenvalues and invariants of tensors*, J. Math. Anal. Appl. 325 (2007) 1363-1377.

[18] L. Qi, *Eigenvalues of a real supersymmetric tensor*, Journal of Symbolic Computation 40 (2005) 1302-1324.

[19] L. Qi, *Eigenvalues of a supersymmetric tensor an positive definiteness of an even degree multivariate form*, Department of Applied Mathematics, The Hong Kong Polytechnic University.

[20] B. Savas and L. L. Elden, *Krylov subspace methods for tensor computations.* Technical report, Department of Mathematics, Linkoping University, 2009. Available at `http://www.mai.liu.se/~laeld/`.

[21] V. Strassen, *Vermeidung yon Divisionen*,. J. Reine Angew. Math. 264 (1973), 184-202.

[22] C.F. Van Loan. *Computational Frameworks for the Fast Fourier Transform* (1992). SIAM Publications, Philadelphia, PA.

[23] Y. Wang, L. Qi, *On the successive supersymmetric rank-1 decomposition of higher-order supersymmetric tensors*, Numerical Linear Algebra with Applications, Vol. 14, No. 4, pp. 503-519.