Spectral partitioning of large and sparse 3-tensors using low-rank tensor approximation

Lars Eldén1 | Maryam Dehghan2

1Department of Mathematics, Linköping University, Linköping, Sweden
2Department of Mathematics, Persian Gulf University, Bushehr, Iran

Correspondence
Lars Eldén, Department of Mathematics, Linköping University, SE-58183 Linköping, Sweden.
Email: lars.elden@liu.se

Abstract
The problem of partitioning a large and sparse tensor is considered, where the tensor consists of a sequence of adjacency matrices. Theory is developed that is a generalization of spectral graph partitioning. A best rank-(2,2,λ) approximation is computed for λ = 1, 2, 3, and the partitioning is computed from the orthogonal matrices and the core tensor of the approximation. It is shown that if the tensor has a certain reducibility structure, then the solution of the best approximation problem exhibits the reducibility structure of the tensor. Further, if the tensor is close to being reducible, then still the solution of the exhibits the structure of the tensor. Numerical examples with synthetic data corroborate the theoretical results. Experiments with tensors from applications show that the method can be used to extract relevant information from large, sparse, and noisy data.

Keywords
low-rank approximation, perturbation theory, reducibility, sparse tensor, spectral partitioning, tensor

1 | INTRODUCTION

Spectral graph partitioning is a standard method in data science graph applications, see, for example, References 1-5. It is usually based on the computation of the two smallest eigenvalues and corresponding eigenvectors of the graph Laplacian or its normalized version. In view of the abundance in data science applications of sequences of graph data organized as tensors, it is natural to ask whether it is possible to generalize spectral graph partitioning to tensors. To our knowledge there is no natural generalization of the graph Laplacian to tensors. However, the normalized Laplacian approach is equivalent to a method based on the computation of the two largest eigenvalues and corresponding eigenvectors of the normalized adjacency matrix of the graph. If the graph is close to being disconnected, then this property shows in the structure of the eigenvectors, and can be used to partition the graph.

In this paper we present a generalization of graph partitioning to large and sparse 3-tensors. To present the basic ideas, we here give an example, where the tensor $A \in \mathbb{R}^{m \times m \times n}$ consists of a sequence of normalized adjacency matrices for undirected graphs over the same nodes; thus the matrices are symmetric, and the tensor is nonnegative. Assume that almost all the graphs are close to being disconnected in the same way. This means that there is a reordering of the nodes of the graphs, such that the reordered tensor can be written (for detailed definitions, see Section 2),

$$\hat{A} = A \cdot (P, P)_{1,2} = \begin{pmatrix} A_1 & E \\ E' & A_2 \end{pmatrix}, \quad A_i \in \mathbb{R}^{m \times m \times n}, \quad i = 1, 2, \quad m_1 + m_2 = m.$$
where $A \cdot (P, P)_{1,2}$ denotes multiplication of the tensor in modes $(1,2)$ by a permutation matrix $P$, and where $E$ is much sparser than $A_1$ and $A_2$. Let $\|A\|$ be the Frobenius norm. We prove that, if $\|E\| \ll \|A_i\|, i = 1,2$, then we can compute the permutation $P$ from the best rank-$(2,2,1)$ approximation

$$\min_{\text{rank}(C)=(2,2,1)} \|A - C\|. \tag{1}$$

Thus we are considering low-rank Tucker approximations$^6,7$ with best approximation properties.$^8$ Let the solution be

$$C = (U, U, w) \cdot F,$$

where $U^T U = I_2$ and $w^T w = 1$, and the core tensor $F \in \mathbb{R}^{2 \times 2 \times 1}$ is almost diagonal (note that the best approximation preserves the partial symmetry of $A$, see Proposition 1). The connectivity properties (the permutation $P$) of $A$ can be deduced from the structure of the solution: $U$ is not unique but its columns can be chosen as approximate indicator vectors for the partitioning. The magnitude of the elements of $F$ indicate the closeness of the graphs to being disconnected.

As in the matrix case, partitioning is related to the concept of reducibility. A 3-tensor can be reducible in different ways. We give examples of reducibility structure that can be computed from best rank-$(2,2,1)$ and rank-$(2,2,2)$ approximations. For instance, if the graphs are close to being bipartite (over the same nodes), this too can be found by computing a best rank-$(2,2,1)$ approximation.

The best approximation problems considered can be written, equivalently, as the the maximization problems,

$$\max_{X,Z} \|A \cdot (X, X, Z)\|, \quad (X, X, Z) \in \text{Gr}^3(2, 2, \lambda), \quad \lambda = 1, 2, \tag{2}$$

where $\text{Gr}^3(2, 2, \lambda)$ is a product of certain Grassmann manifolds. This is a generalization of Rayleigh quotient maximization of a symmetric matrix, and it is the analogy with graph partitioning based on a normalized adjacency matrix.

In Section 2 we define notation and preliminaries. To motivate the generalization to tensors, we give a brief introduction to spectral graph partitioning using the normalized adjacency matrix in Section 3. A graph is disconnected if the adjacency matrix is reducible. The corresponding tensor property is introduced in Section 4, where we also recall the concept of stationary points for the maximization problem (2). In Section 5 we give characterizations of the solutions of the best approximation of tensors with certain reducibility structures. These are used in a method for computing the reducibility structure of tensors that are close to reducible. We give examples of such tensors and compute their structure in Section 7. There we also show how the method can be used to analyze a real data tensor, which is far from being reducible.

The results of this paper are based on the perturbation theory for the best low-rank approximation of a tensor,$^9$ which shows that if the tensor is perturbed by a small amount, then the solution $(U, U, w)$, and $F$ of (1) is also perturbed by a small amount. The theory of spectral partitioning of a single graph can be based on the corresponding theory for matrices (which is actually a special case of that in Reference 9).

The main contributions of this paper are the following: (i) investigation of the structure of the solution of the best approximation problem for several reducible $(1,2)$-symmetric tensors, (ii) methods for computing the structure of the solution for tensors that are close to reducible, (iii) demonstration that for such tensors the solution vectors can be written as approximate indicator vectors, which can be used for spectral partitioning of the tensor.

Methods for computing best multilinear approximations of large tensors are given in References 10,11. Other approaches to tensor clustering are described in References 12-18, and in 19 clustering is done using HOSVD (Higher-Order Singular Value Decomposition). Our method is novel and different from all these, as it is based on one simple criterion (2), and can be considered as a generalization of graph partitioning for a single graph. In this paper we present the ideas behind the method; a study of the performance of our method compared to the others when applied to real data is outside the scope of this paper.

The present paper is the second in a series of three on the analysis of large and sparse tensor data. In the first$^{20}$ we develop a Krylov–Schur like algorithm for efficiently computing best low multilinear rank approximations of large and sparse tensors. In the third paper$^{21}$ we use the methods to analyze large tensors from data science applications.
2. TENSOR CONCEPTS AND PRELIMINARIES

Let \( \mathcal{A} \in \mathbb{R}^{l \times m \times n} \) be a three-dimensional (3D) array of real numbers. With the definitions below and the approximation problem (1), \( \mathcal{A} \) is a Cartesian tensor (tensor for short).22 Tensors will be denoted by calligraphic letters, for example, \( \mathcal{A}, \mathcal{B} \), matrices by capital roman letters and vectors by lowercase roman letters. In order not to burden the presentation with too much detail, we sometimes will not explicitly mention the dimensions of matrices and tensors, and assume that they are such that the operations are well-defined. The whole presentation will be in terms of tensors of order three, or equivalently 3-tensors.

The different “dimensions” of the tensor are referred to as modes. We will use both standard subscripts and MATLAB-like notation: a particular tensor element will be denoted in two equivalent ways, \( \mathcal{A}(i, j, k) = a_{ijk} \).

A subtensor obtained by fixing one of the indices is called a slice, for example, \( \mathcal{A}(i, :, :) \). In the case when the index is fixed in the first mode, we call the slice a 1-slice, and correspondingly for the other modes. A slice can be considered as a 3-tensor with a singleton mode, and also as a matrix. A fiber is a subtensor, where all indices but one are fixed. For instance, \( \mathcal{A}(i, :, k) \) denotes a mode-2 fiber.

For a given third-order tensor, there are three associated subspaces, one for each mode. These subspaces are given by

\[
\begin{align*}
\text{Range}\{\mathcal{A}(::, j, k) | j = 1 : m, k = 1 : n\}, \\
\text{Range}\{\mathcal{A}(i ::, : , k) | i = 1 : l, k = 1 : n\}, \\
\text{Range}\{\mathcal{A}(i, j ::) | i = 1 : l, j = 1 : m\}.
\end{align*}
\]

The multilinear rank of the tensor is said to be equal to \( (p, q, r) \) if the dimension of these subspaces are \( p, q, \) and \( r \), respectively.

It is customary in numerical linear algebra to write out column vectors with the elements arranged vertically, and row vectors with the elements horizontally. This becomes inconvenient when we are dealing with more than two modes. Therefore we will not make a notational distinction between mode-1, mode-2, and mode-3 vectors, and we will allow ourselves to write all vectors organized vertically. It will be clear from the context to which mode the vectors belong.

Since we will be dealing with 3-tensors only, it is still possible to display them graphically in figures and formulas. For instance, we will consider tensors,

\[
\mathbb{R}^{2 \times 2 \times n} \ni \begin{pmatrix} p_1 & p_3 \\ p_3 & p_2 \end{pmatrix}, \quad p_i \in \mathbb{R}^n, \quad i = 1, 2, 3.
\]

It may help the reader’s intuition, if the mode-3 fibers are thought of as extending perpendicular to the plane of the paper/screen.

2.1 Tensor-matrix multiplication

We define multilinear multiplication of a tensor by a matrix as follows. For concreteness we first present multiplication by one matrix along the first mode and later for all three modes simultaneously. The mode-1 product of a tensor \( \mathcal{A} \in \mathbb{R}^{l \times m \times n} \) by a matrix \( U \in \mathbb{R}^{p \times l} \) is defined

\[
\mathbb{R}^{p \times m \times n} \ni B = (U)_1 \cdot \mathcal{A}, \quad b_{ijk} = \sum_{\alpha=1}^{l} u_{\alpha i} a_{\alpha j k},
\]

This means that all mode-1 fibers in the 3-tensor \( \mathcal{A} \) are multiplied by the matrix \( U \). Similarly, mode-2 multiplication by a matrix \( V \in \mathbb{R}^{p \times m} \) means that all mode-2 fibers are multiplied by the matrix \( V \). Mode-3 multiplication is analogous. With a third matrix \( W \in \mathbb{R}^{r \times n} \), the tensor-matrix multiplication in all modes is given by

\[
\mathbb{R}^{p \times q \times r} \ni B = (U, V, W) \cdot \mathcal{A}, \quad b_{ijk} = \sum_{\alpha, \beta, \gamma=1}^{l, m, n} u_{\alpha i} v_{\beta j} w_{\gamma k} a_{\alpha \beta \gamma},
\]

where the mode of each multiplication is understood from the order in which the matrices are given.
It is convenient to introduce a separate notation for multiplication by a transposed matrix $\bar{U} \in \mathbb{R}^{l \times p}$:

\[
\mathbb{R}^{p \times m \times n} \ni C = (\bar{U}^T)_1 \cdot A = A \cdot (\bar{U})_1, \quad c_{ijk} = \sum_{a=1}^{l} a_{ajk} \bar{u}_{ai}.
\]  

Let $u \in \mathbb{R}^l$ be a vector and $A \in \mathbb{R}^{l \times m \times n}$ a tensor. Then

\[
\mathbb{R}^{1 \times m \times n} \ni B := (u^T)_1 \cdot A = A \cdot (u)_1 \equiv B \in \mathbb{R}^{m \times n}.
\]

Thus we identify a tensor with a singleton dimension with a matrix. Similarly, with $u \in \mathbb{R}^l$ and $w \in \mathbb{R}^n$, we will identify

\[
\mathbb{R}^{1 \times m \times 1} \ni C := A \cdot (u, w)_1, 3 \equiv c \in \mathbb{R}^m,
\]

that is, a tensor of order 3 with two singleton dimensions is identified with a vector, here in the second mode.

### 2.2 Inner product, norm, and contractions

Given two tensors $A$ and $B$ of the same dimensions, we define the inner product,

\[
\langle A, B \rangle = \sum_{a,\beta,\gamma} a_{a\beta\gamma} b_{a\beta\gamma}.
\]  

The corresponding tensor norm is

\[
\|A\| = \langle A, A \rangle^{1/2}.
\]

This *Frobenius norm* will be used throughout the paper. As in the matrix case, the norm is invariant under orthogonal transformations, that is,

\[
\|A\| = \|(U, V, W) \cdot A\| = \|A \cdot (P, Q, S)\|,
\]

for orthogonal matrices $U, V, W, P, Q,$ and $S$. This is obvious from the fact that multilinear multiplication by orthogonal matrices does not change the Euclidean length of the corresponding fibers of the tensor.

For convenience we will denote the inner product of vectors $x$ and $y$ in any mode (but, of course, the same) by $x^T y$.

The following well-known result will be needed.

**Lemma 1.** Let $A \in \mathbb{R}^{l \times m \times n}$ be given along with three matrices with orthonormal columns, $U \in \mathbb{R}^{p \times l}$, $V \in \mathbb{R}^{m \times q}$, and $W \in \mathbb{R}^{n \times r}$, where $p \leq l$, $q \leq m$, and $r \leq n$. Then the least squares problem

\[
\min_S \|A - (U, V, W) \cdot S\|,
\]

with $S \in \mathbb{R}^{p \times q \times r}$, has the unique solution

\[
S = (U^T, V^T, W^T) \cdot A = A \cdot (U, V, W).
\]

The elements of the tensor $S$ are given by

\[
s_{\lambda \mu \nu} = A \cdot (u_{\lambda}, v_{\mu}, w_{\nu}) \cdot 1 \leq \lambda \leq p, \quad 1 \leq \mu \leq q, \quad 1 \leq \nu \leq r.
\]  

The proof is a straightforward generalization of the corresponding proof for matrices, see Reference 8.

The inner product (8) can be considered as a special case of the *contracted product of two tensors*, which is a tensor (outer) product followed by a contraction along specified modes. Thus, if $A$ and $B$ are 3-tensors, we define, using essentially the notation of Reference 25,
\[ C = \langle A, B \rangle_1, \quad c_{jk'k} = \sum_a a_{ajk} b_{aj'k'}, \quad (4\text{-tensor}). \quad (11a) \]

\[ D = \langle A, B \rangle_{1,2}, \quad d_{kk'} = \sum_{a,\beta} b_{a\beta k} b_{a\beta k'}, \quad (2\text{-tensor}). \quad (11b) \]

\[ e = \langle A, B \rangle = \langle A, B \rangle_{1,2,3}, \quad e = \sum_{a,\beta,\gamma} a_{a\beta\gamma} b_{a\beta\gamma}, \quad (\text{scalar}). \quad (11c) \]

It is required that the contracted dimensions in the two tensors are equal. We will refer to the first two as partial contractions. It is also convenient to introduce a notation when contraction is performed in all but one mode. For example the product in (11b) may also be written

\[ \langle A, B \rangle_{1,2} \equiv \langle A, B \rangle_{-3}. \quad (12) \]

Two tensors \( A \) and \( B \) of equal dimensions are said to be \((i,j)\)-orthogonal, if

\[ \langle A, B \rangle_{ij} = 0. \]

In the case of 3-tensors, if two tensors \( A \) and \( B \) are \((1,2)\)-orthogonal, we also say that they are \(-3\)-orthogonal, that is, \( \langle A, B \rangle_{-3} = 0 \).

The definition of contracted products is valid also when the tensors are of different order. The only assumption is that the dimension of the correspondingly contracted modes are the same in the two arguments. The dimensions of the resulting product are in the order given by the noncontracted modes of the first argument followed by the noncontracted modes of the second argument.

### 2.3 Best rank-\((r_1, r_2, r_3)\) approximation

The problem of approximating tensor \( A \in \mathbb{R}^{b \times m \times n} \) by another tensor \( B \) of lower multilinear rank,

\[
\min_{\text{rank}(B) = (r_1, r_2, r_3)} \| A - C \|^2,
\]

is treated in References 8,26-28. It is shown in Reference 8 that (13) is equivalent to

\[
\max_{X,Y,Z} \| A \cdot (X, Y, Z) \|^2, \quad \text{subject to} \quad X^T X = I_{r_1}, \quad Y^T Y = I_{r_2}, \quad Z^T Z = I_{r_3}, \quad (14)
\]

where \( X \in \mathbb{R}^{b \times r_1}, Y \in \mathbb{R}^{m \times r_2}, \) and \( Z \in \mathbb{R}^{n \times r_3}. \) The problem (14) can be considered as a Rayleigh quotient maximization problem, in analogy with the matrix case.29 We will refer to the two equivalent problems as the the best rank-(\( r_1, r_2, r_3 \)) approximation problem. To simplify the terminology somewhat we will refer to a solution \((U, V, W)\) of the approximation problem as the best rank-(\( r_1, r_2, r_3 \)) approximation of the tensor, and often tacitly assume that the corresponding core tensor is \( F = A \cdot (U, V, W) \).

The constrained maximization problem (14) can be thought of as an unconstrained maximization problem on a product of Grassmann manifolds, see References 30,31. Thus a solution \((U, V, W)\) is a representative of an equivalence class of matrices \((UQ_1, VQ_2, WQ_3), \) where \( Q_i \in \mathbb{R}^{r_i \times r_i} \) are orthogonal. Note also that

\[
(U, V, W) \cdot F = (U_\theta, V_\theta, W_\theta) \cdot F_\theta, \quad (15)
\]

where \((U_\theta, V_\theta, W_\theta) = (UQ_1, VQ_2, WQ_3), \) and \( F_\theta = F \cdot (Q_1, Q_2, Q_3). \) For simplicity, in the sequel when we talk about the solution of the best approximation problem, it is understood as a representative of the solution equivalence class. Likewise, a unique solution means that the equivalence class is unique.

It is also important to keep in mind that as the optimization problem is nonconvex, we cannot guarantee that a local optimum is a global one, see Reference 32.
2.4 (1,2)-Symmetric tensors

A 3-tensor $A \in \mathbb{R}^{m \times m \times n}$ is called (1,2)-symmetric if all its 3-slices are symmetric, that is,

$$A(i, j, k) = A(j, i, k), \quad i, j = 1, 2, \ldots, m; \quad k = 1, 2, \ldots, n.$$ 

Henceforth in this paper we will assume that $A$ is (1,2)-symmetric.

We will partition the (1,2)-symmetric tensor with respect to the first two modes,

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A'_{12} & A_{22} \end{pmatrix},$$

where $A_{11} \in \mathbb{R}^{m_1 \times m_1 \times n}$ for $m_1 < m$, and the other “block tensors” have dimensions accordingly. Here we use the notation $A'_{12}$ to indicate that each three-slice is the transpose of the corresponding slice of $A_{12}$.

**Proposition 1.** Assume that the tensor $A$ is (1,2)-symmetric, and that the solution $(U, V, W)$ of the best approximation problem (13) with $r_1 = r_2$ is unique. Then representatives of the solution can be chosen such that $U = V$ and the core tensor $F = A \cdot (U, U, W)$ is (1,2)-symmetric.

**Proof.** $A$ is (1,2)-symmetric, and, to be optimal, so is the approximating tensor $(U, V, W) \cdot F$. Therefore $U$ and $V$ must span the same subspace, and we can choose $V = UQ$, for an orthogonal matrix $Q$. The rest follows from the Grassmann property of the problem, and from Lemma 1.

In this paper we will treat the best approximation problem

$$\max_{X, Z} \| A \cdot (X, X, Z) \|, \quad X^T X = I_{r_1}, \quad Z^T Z = I_{r_3},$$

and denote its solution $(U, U, W)$. We use the notation

$$\text{Gr}^3(r_1, r_1, r_3) = \{ (X, X, Z) | X^T X = I_{r_1}, \quad Z^T Z = I_{r_3} \},$$

in connection with (1,2)-symmetric minimization over a product of Grassmann manifolds.

2.5 A Krylov–Schur like method for computing best rank-$(r_1, r_2, r_3)$ approximations

To be able to solve large problems with sparse tensors it is necessary to use methods that do not apply transformations to the tensor itself (except possibly reorderings), because transformations would generally create fill-in. For problems with sparse matrices, restarted Krylov methods are standard. In References 20,35 we developed a block-Krylov type method for tensors, which accesses the tensor only in tensor-matrix multiplications, where the matrix consists of a relatively small number of columns. Thus the method is memory-efficient. In Reference 20 it was combined with the Krylov-Schur approach for the computation of the best rank-$(r_1, r_2, r_3)$ approximation of a large and sparse tensors. For the problems solved in Section 7 convergence is very fast. It is shown in Reference 20 that the block Krylov–Schur method is in general considerably faster than the Higher-Order Orthogonal Iteration.

3 BACKGROUND: SPECTRAL GRAPH PARTITIONING USING THE ADJACENCY MATRIX

In this section we give a brief overview of the ideas behind spectral graph partitioning, emphasizing the aspects that we will generalize to tensors. A survey of methods is given in Reference 2, and the mathematical theory is presented in Reference 1. The presentation here is based on the fact that graph partitioning using the normalized graph Laplacian is equivalent to partitioning using the normalized adjacency matrix.
Let $A_0$ be the (unnormalized) adjacency matrix of a connected undirected graph $G$. The Laplacian of the graph is

$$L_0 = D - A_0,$$

where $D$ is a diagonal matrix,

$$D = \text{diag}(d), \quad d = A_0 e,$$

and $e = (1 \ 1 \ \cdots \ 1)^T$. The normalized Laplacian is defined

$$L = I - A := I - D^{-1/2} A_0 D^{-1/2}.$$

The matrix $A$ can be called the normalized adjacency matrix, and it is symmetric. Both Laplacian matrices are positive semidefinite. Usually spectral partitioning is developed in terms of the Laplacian or the normalized Laplacian.\cite{1, 36} The smallest eigenvalue of either Laplacian is equal to zero, and the value of the second smallest determines how close the graph is to being disconnected. Obviously, the normalized Laplacian has the same eigenvectors as the adjacency matrix $A$ and the eigenvalues are related by

$$\lambda(L) = 1 - \lambda(A).$$

So instead of basing spectral partitioning on the smallest eigenvalues of $L$ (or $L_0$) we can consider the largest eigenvalues of $A$ and the corresponding eigenvectors. We will assume the ordering of the eigenvalues of the normalized adjacency matrix $A$,

$$1 = \lambda_1 \geq \lambda_2 \geq \cdots,$$

and denote the corresponding eigenvectors $v_1, v_2, \ldots$.

To illustrate the heuristics of spectral graph partitioning, we constructed a disconnected graph with 100 nodes, and a nearby graph, where four edges were added that made the graph connected. The respective normalized adjacency matrices $A$ and $\hat{A}$ are illustrated in Figure 1. Note that the matrix $A$ is reducible (actually in reduced form), whereas $\hat{A}$ is irreducible (see, e.g., Reference 37, section 6.2.21).

The following fundamental properties of the eigenvalue problem for symmetric matrices are the key to spectral partitioning. We give the results in somewhat loose form, for precise statements, see, for example, Reference 38, theorems 8.1.4, 8.1.10 of Reference 39, theorems 17.2, 17.3.

**Proposition 2.** If a symmetric matrix $A$ is perturbed by the amount $e$, then the eigenvalues are perturbed by approximately the same amount $e$. The eigenvectors are perturbed essentially by $e$ divided by a quantity that depends on the distance to nearby eigenvalues.

The adjacency matrices are normalized and therefore the two-subgraph matrix $A$ has a double eigenvalue equal to 1, which are the largest eigenvalues for the adjacency matrices of each subgraph. The two largest eigenvalues of $\hat{A}$ corresponding to the connected graph are 1 and 0.9913.

The first two eigenvectors of the two adjacency matrices are given in Figure 1. Since $A$ has a double eigenvalue, the eigenvectors are not unique, only the eigenspace. One can choose the basis vectors for the eigenspace as the eigenvectors for each of the two submatrices, which are nonnegative. For $\hat{A}$ the first eigenvector is positive, due to the fact that the matrix is irreducible and nonnegative (see Perron-Frobenius theory, e.g. in Reference 40, chapter 8). The second eigenvector, which is orthogonal to the first, must have a sign change, and that sign change occurs at the interface between the diagonal blocks. Since $\hat{A}$ is a small perturbation of $A$, the subspaces spanned by the first two eigenvectors are close; here the cosines of the angles between the subspaces are 1.0000 and 0.9997. In real applications, given a graph that is close to being disconnected, the nodes are not ordered so that the structure is apparent as in our example. In fact, the purpose of spectral partitioning is to find the node ordering that exhibits such structure. A simplified formulation of one step of spectral partitioning of a connected graph is given in Algorithm 1.
Algorithm 1. One step of spectral graph partitioning

1. Compute the first two eigenvectors $v_1$ and $v_2$ of the normalized adjacency matrix.
2. Order the elements of $v_2$ in decreasing order, and renumber the nodes correspondingly.
3. Given that ordering, compute the cost for cuts in the vicinity of the sign change of the reordered eigenvector, and partition the graph in two where the cost is the smallest.

“Cost” in the algorithm can be given a precise meaning, for example, in terms of the concept conductance, see for example Reference 1 or 5, table 1. Essentially it should be a measure of the number of edges that are broken at a cut, divided by a measure of the number of edges in the denser partitions.

Finally in this section we summarize a few facts about spectral graph partitioning that will be the basis for generalization to the tensor case.

A1. Using the normalized adjacency matrix is equivalent to using the normalized Laplacian.
A2. If the graph is connected, then the first eigenvector $v_1$ is positive, and the second has positive and negative components (due to orthogonality).
A3. The matrix $B = \lambda_1 v_1 v_1^T + \lambda_2 v_2 v_2^T$ is the best rank-2 approximation of the normalized adjacency matrix.
A4. In the disconnected case the normalization makes the two blocks each have the largest eigenvalue equal to 1, with nonnegative eigenvectors.

The arguments of this section can be modified so that they apply to two graphs that are bipartite and close to bipartite. For conciseness we omit them here; they can be deduced from Proposition 5 and the reasoning in Section 6.
4 | SPECTRAL PARTITIONING OF TENSORS

We first present two motivating examples, which generalize the graph concept. The objects of study are sparse (1,2)-symmetric and nonnegative tensors.

Example 1. Web link analysis. In Reference 41 a graph representing links between web pages is studied. Thus the first two modes represent the web pages. Instead of just registering the existence of links, a third mode is used to denote the key word (term) via which the two web pages are linked, in the style of Wikipedia. Thus, a 3-tensor \( \mathbf{A} \) is constructed as

\[
a_{ijk} = \begin{cases} 1 & \text{if page } i \text{ points to page } j \text{ using term } k, \\ 0 & \text{otherwise}. \end{cases}
\]

Assuming that the links are bidirectional, each 3-slice of \( \mathbf{A} \) represents the undirected graph of pages that are linked using one particular term.

Assuming that there are two different topics covered on two (almost) disjoint sets of web pages, can we find those topics and the corresponding groups of web pages? Can we also find the key words that are common to all web pages?

Example 2. Network evolving over time. In Reference 42 a subset of the Enron e-mail communication is analyzed over a period of 44 months. The first two tensor modes represent 184 employees communicating, and the third mode is time. Thus each 3-slice is equivalent to a graph of the email users who communicate during one time unit.

Assume that there are two groups of users, and that both groups communicate frequently internally during one period in time. Then at a later stage in time, all users communicate with all the others.

On the other hand, assume that there are two groups of users that both communicate only within their own group during the whole period studied.

As a third example, assume that during the first time period one group communicates with all others, while the other group does not communicate within its own group. Then during the second period, all users are communicating with all.

Can we detect such communication patterns by spectral partitioning?

In both examples, since we assume that the graphs are undirected, the tensor is (1, 2)-symmetric. Also, in both examples the nodes have no loops, that is, for all \( i \) and \( k \), \( a_{ijk} = 0 \).

4.1 | Reducibility and conductance

Reducibility and Perron–Frobenius theory for tensors are treated in References 44-46. From Example 2 one can see that we need a more specific reducibility concept than that in Reference 44. Thus we make the following definition for the case of (1,2)-symmetric 3-tensors.

Definition 1. A (1,2)-symmetric tensor \( \mathbf{A} \in \mathbb{R}^{m \times m \times n} \) is called (1,2)-reducible if there exists a nonempty proper subset \( I \subset \{1, 2, \ldots, m\} \) and a nonempty subset \( K \subseteq \{1, 2, \ldots, n\} \) such that

\[
a_{ijk} = a_{jik} = 0, \quad \forall i \in I, \quad \forall j \not\in I, \quad \forall k \in K.
\]

Similarly, it is called 3-reducible if

\[
a_{ijk} = 0, \quad \forall i, j \in I, \quad \forall k \in K.
\]

If \( \mathbf{A} \) is either (1,2)-reducible or 3-reducible we call it reducible, and if it is not reducible, we call it irreducible.

If the tensor is reducible, then we can permute its indices in a (1,2)-symmetric way so that \( I \) becomes \( \{1, 2, \ldots, m_1\} \), for some \( m_1 < m \), and also permute the indices in the third mode in a usually different way so that \( K \) becomes \( \{1, 2, \ldots, n_1\} \) for some \( n_1 \leq n \). A (1,2)-reducible tensor will after such a permutation have the structure

\[
\mathbf{A}(::; 1 : n_1) = \begin{pmatrix} \mathbf{A}_1 & 0 \\ 0 & \mathbf{A}_2 \end{pmatrix}, \quad \mathbf{A}(::; n_1 : n) = \begin{pmatrix} \mathbf{A}_3 & \mathbf{A}_4 \\ \mathbf{A}_4^T & \mathbf{A}_5 \end{pmatrix}.
\]
Note that we allow \( n_1 = n \), in which case the second subtensor in (18) is vacuous. A 3-reducible tensor has the structure
\[
\mathcal{A}(\cdot, \cdot, 1 : n_1) = \begin{pmatrix} A_1 & A_2 \\ A'_2 & 0 \end{pmatrix}, \quad \mathcal{A}(\cdot, \cdot, n_1 : n) = \begin{pmatrix} A_3 & A_4 \\ A'_4 & A_5 \end{pmatrix},
\]
for some \( n_1 \leq n \). (We here use the prime notation to indicate that \( A'_1(\cdot, \cdot, k)^T = A_d(\cdot, \cdot, k) \) for \( k = n_1 + 1, \ldots, n \). Note that (18) corresponds to the first case in Example 2 and (19) corresponds to the third case. If, in (19), \( n_1 = n \) and \( A_1 = 0 \), then each 3-slice represents a bipartite graph.

In the case of adjacency matrices for one single graph, closeness to reducibility can be measured in different ways, see for example, Reference 5, table 1. One common definition is conductance. Let the two sets of nodes be \( S \) and \( \overline{S} \), and define \( \text{vol}(S) \) as the sum of the degrees of the nodes in \( S \). The number of edges between the nodes in the two sets is denoted \( \text{cut}(S, \overline{S}) \). Then conductance is defined
\[
\text{cond} = \frac{\text{cut}(S, \overline{S})}{\min(\text{vol}(S), \text{vol}(\overline{S}))}.
\]
As we are dealing with different types of reducibility we must adapt the measure to the actual type. Moreover, as our method is based on perturbation theory, it is natural to define a measure using norms of the tensor blocks after permutations. For instance, when the permuted tensor is approximately of type (18),
\[
\mathcal{A}(\cdot, \cdot, 1 : n_1) = \begin{pmatrix} A_1 & \mathcal{E} \\ \mathcal{E}' & A_2 \end{pmatrix}, \quad \mathcal{A}(\cdot, \cdot, n_1 : n) = \begin{pmatrix} A_3 & A_4 \\ A'_4 & A_5 \end{pmatrix},
\]
where \( \|\mathcal{E}\| \) is small, we define conductance in terms of the Frobenius norm,
\[
\text{cond}_F = \frac{\|\mathcal{E}\|}{\min(\|A_1\|, \|A_2\|, \|\mathcal{A}(\cdot, \cdot, n_1 : n)\|)}.
\]
Naturally, measures based on generalizations of cut and vol could also be used. Our measure is equivalent to such measures in a similar way as different matrix norms are equivalent.

### 4.2 Normalization of 3-slices

In Section 3 we described the normalization of an adjacency matrix for a connected, undirected graph. There the normalization has three main effects: (i) the largest eigenvalue of the adjacency matrix is equal to 1, (ii) if the graph is close to being disconnected, then the second largest eigenvalue is close to 1, (iii) the eigenvectors corresponding to the two largest eigenvalues give information about the structure of the graph, and can be used to find a good partitioning.

We are not aware of any analogous normalization of a (1,2)-symmetric 3-tensor with analogous properties. However, if we assume that the 3-slices are adjacency matrices of undirected graphs, it is reasonable to normalize them separately. Thus we assume that the tensor has been normalized as in Algorithm 2. The vector \( e = (1 \ 1 \ \cdots \ 1)^T \in \mathbb{R}^m \).

#### Algorithm 2. Normalization of the 3-slices of (1,2)-symmetric \( \mathcal{A} \)

```plaintext
for i = 1 : n
  B := \mathcal{A}(\cdot, \cdot, i)
  d := Be
  D := \text{diag}(d)
  \mathcal{A}(\cdot, \cdot, i) := D^{-1/2}BD^{-1/2}
end
```
Thus we consider the special case when \( n \) is (1,2)-symmetric, then the best rank-\((1,1,1)\) approximation problem has a solution.

### 4.3 Stationary points

For later reference we here restate the first-order conditions for a stationary point for the maximization problem (16)\(^9\), Proposition 4.2.

Consider the solution \((U, U, W)\) of the (1,2)-symmetric best rank-\((r_1, r_1, r_3)\) approximation problem (i.e., \( U \in \mathbb{R}^{m \times r_1}\), and \( W \in \mathbb{R}^{n \times r_3}\)), and let \( U_\perp \) and \( W_\perp \) be matrices such that \((U U_\perp)\) and \((W W_\perp)\) are both square and orthogonal.

**Proposition 3.** Let \((U, U, W)\) be a stationary point for the (1,2)-symmetric best approximation problem, and put \( F = A \cdot (U, U, W) \). Then the following equations are satisfied

\[
\begin{align*}
\langle F_\perp^1, F \rangle_{-1} &= 0, \\
\langle F_\perp^2, F \rangle_{-2} &= 0, \\
\langle F_\perp^3, F \rangle_{-3} &= 0,
\end{align*}
\]

In our case, when \( A \) is (1,2)-symmetric, the two relations (21) and (22) are equivalent.

If we apply Proposition 3 to the best rank-\((1, 1, 1)\) approximation problem, we have the following result.\(^{26}\)

**Lemma 2.** A stationary point \((u, v, w)\) for the best rank-\(-(1,1,1)\) approximation of a (non-symmetric) tensor \( A \) satisfies the equations

\[
A \cdot (u, v)_{1,2} = rv, \quad A \cdot (u, w)_{1,3} = rw, \quad A \cdot (v, w)_{2,3} = ru,
\]

where \( r = A \cdot (u, v, w) \).

In the case when \( A \) is (1,2)-symmetric the solution \((u, u, w)\) satisfies

\[
A \cdot (u, u)_{1,2} = rv, \quad A \cdot (u, w)_{1,3} = A \cdot (u, w)_{2,3} = ru.
\]

The lemma follows from the orthogonality relations in Proposition 3, and the fact that \( F \) is scalar: then \( \mathbb{R}^{(m-1)\times 1 \times 1} \ni F_\perp^1 = 0 \), etc.

In the rest of the paper we will use the following lemma. The result seems to be well-known (at least the first part). For completeness we give a relatively elementary proof in Appendix A.

**Lemma 3.** Let the 3-tensor \( A \) be nonnegative. Then the best rank-\((1,1,1)\) approximation is nonnegative. If, in addition, the tensor is irreducible and (1,2)-symmetric, then the best rank-\((1,1,1)\) approximation problem has a solution \((u, u, w)\), where \( u > 0 \) and \( w > 0 \).

### 5 Best approximation problems for reducible tensors

We will now present the solution of best low-rank approximation problems for some specific (1,2)-symmetric reducible tensors, and we will show how the reducibility structure of the tensors is mirrored in the structure of the solution. Then, in Section 6 we will use perturbation theory and show how the the solution for a close-to-reducible tensor can be used to ascertain the structure of the tensor and compute a partitioning. We remind the reader that \((U, U, W)\) should be understood as a representative of an equivalence class, compared with the arguments around (15).

Assume that the (1,2)-symmetric tensor \( A \) is (1,2)-reducible with

\[
A = \begin{pmatrix}
A_1 & 0 \\
0 & A_2
\end{pmatrix}, \quad A_i \in \mathbb{R}^{m_i \times m_i \times n}, \quad v = 1, 2, \quad m_1 + m_2 = m.
\]

Thus we consider the special case when \( n_1 = n \) in (18).
Lemma 4. Assume that the (1,2)-symmetric tensor \( A \) has the structure (26), where the subtensors \( A_1 \) and \( A_2 \) are irreducible, and that \((U, U, W)\) is the unique solution of the best rank-(2,2,2) approximation problem. Assume further that

\[
\min_{Gr^3(2,2,2)} \|A - (X,X,Z) \cdot F\|^2 < \min_{Gr^3(2,2,2)} \|A_1 - (X,X,W) \cdot F\|^2 + \|A_2\|^2, \tag{27}
\]

and that the corresponding inequality holds with \( A_1 \) and \( A_2 \) interchanged. Then \( U \) has the structure

\[
\begin{pmatrix}
  u_{11} & 0 \\
  0 & u_{22}
\end{pmatrix}, \tag{28}
\]

where \( u_{11} \in \mathbb{R}^{m_1}, u_{22} \in \mathbb{R}^{m_2}, \) for \( i = 1, 2. \)

The same result holds for the best rank-(2,2,1) approximation of \( A. \)

Proof. Let \((X,X,Z) \in Gr^3(2,2,2)\), and let \( H \in \mathbb{R}^{2 \times 2 \times 2} \) be arbitrary (1,2)-symmetric. Put \( C = (X,X,Z) \cdot H, \) and partition

\[
C = \begin{pmatrix}
  C_1 & C_{12} \\
  C_{12}^T & C_2
\end{pmatrix}
\]

Clearly \( C \) is (1,2)-symmetric. Then

\[
\min_{Gr^3(2,2,2)} \|A - C\|^2 = \min_{Gr^3(2,2,2)} \left( \|A_1 - C_1\|^2 + 2\|C_{12}\|^2 + \|A_2 - C_2\|^2 \right) \\
\geq \min_{Gr^3(2,2,2)} \left( \|A_1 - C_1\|^2 + \|A_2 - C_2\|^2 \right) \tag{29}
\]

\[
= \min_{Gr^3(2,2,2)} \left\| \begin{pmatrix}
  A_1 & 0 \\
  0 & A_2
\end{pmatrix} - \begin{pmatrix}
  C_1 & 0 \\
  0 & C_2
\end{pmatrix} \right\|^2, \tag{30}
\]

with strict inequality if \( C_{12} \neq 0. \) We can write the rank-(2,2,2) approximation

\[
C = (z_1)_3 \cdot H_1 + (z_2)_3 \cdot H_2,
\]

where \( H_v = (X,X)_1, v = 1, 2, \) are symmetric rank-2 matrices. Due to the orthogonality of \( z_1 \) and \( z_2, \) the (1,2)-block of \( C \) is equal to zero if and only if the corresponding (1,2)-blocks of \( H_1 \) and \( H_2 \) are both zero. From Lemma 5 in Appendix B we then see that the solution \( U \) must have the structure (28) or

\[
\begin{pmatrix}
  u_{11} & u_{12} \\
  0 & 0
\end{pmatrix}, \quad \begin{pmatrix}
  0 & 0 \\
  u_{21} & u_{22}
\end{pmatrix}. \tag{31}
\]

If, we let \( C_1 \) have rank \((2,2,2)\), then, due to the structure of \( C, \) we must have \( C_2 = 0, \) and vice versa. So we can let \( C_1 \) have rank \((2,2,2)\), put \( C_2 = 0, \) and minimize the first term in (29). This corresponds to the first structure in (31). Similarly, minimizing only the second term in (29) corresponds to the second structure in (31). The assumption (27) rules out these two cases. Thus, we can let \( C_1 \) and \( C_2 \) each have rank \((1,1,2), \) and minimize both terms in (29) simultaneously. This corresponds to the structure in (28).

The proof for the best rank-(2,2,1) approximation is almost identical. □

If we normalize the 3-slices of \( A, \) then \( \|A_1\| \) and \( \|A_2\| \) will be approximately equal, and the technical condition (27) will be satisfied if both \( A_1 \) and \( A_2 \) are sufficiently far from being reducible. The condition is needed because we cannot guarantee that \( \|A_1\| \) and \( \|A_2\| \) are equal. We remind that uniqueness in the following proposition is in the sense of equivalence class, see Section 2.3.
**Proposition 4.** Let the (1,2)\textemdash symmetric, nonnegative tensor $\mathbf{A} \in \mathbb{R}^{m \times m \times n}$ be (1,2)-reducible with the structure

$$\mathbf{A} = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \quad A_i \in \mathbb{R}^{m_i \times m_i \times n}, \quad i = 1, 2, \quad m_1 + m_2 = n, \quad (32)$$

where $A_1$ and $A_2$ are irreducible. Assume that the assumptions of Lemma 4 are satisfied. Further assume that the solution $(U, U, W)$ with core tensor $F$, of the best rank-(2,2,2) approximation problem for $\mathbf{A}$, is unique. Then

$$U = \begin{pmatrix} u_1 & 0 \\ 0 & u_2 \end{pmatrix}. \quad (33)$$

The 3-slices of $F$ are diagonal. The corresponding result hold for the best rank-(2,2,1) approximation.

**Proof.** From Lemma 4 we that the solution must have the structure (33). Then the core tensor is

$$F = (A \cdot (U, U)_{1,2}) \cdot (W)_3 = \begin{pmatrix} A_1 \cdot (u_1, u_1)_{1,2} & 0 \\ 0 & A_2 \cdot (u_2, u_2)_{1,2} \end{pmatrix} \cdot (W)_3$$

$$= \begin{pmatrix} A_1 \cdot (u_1, u_1, W) & 0 \\ 0 & A_2 \cdot (u_2, u_2, W) \end{pmatrix} \in \mathbb{R}^{2 \times 2 \times 2},$$

thus the 3-slices of $F$ are diagonal.

The proof for the rank-(2,2,1) case is almost identical. □

**Corollary 1.** The solution vectors $u_1$ and $u_2$ in Proposition 4 are positive.

**Proof.** In the rank-(2,2,1) case the corollary can be proved by a modification of the proof of Proposition 3 in Appendix A. In the rank-(2,2,2) case, let $(\tilde{u}_i, \tilde{w}_i)$ be the positive solution vectors for the rank-(1,1,1) problems for $A_i$, $i = 1, 2$. Using (25) we have

$$A \cdot (U, U)_{1,2} = \begin{pmatrix} \tau_1 \tilde{w}_1 & 0 \\ 0 & \tau_2 \tilde{w}_2 \end{pmatrix} \in \mathbb{R}^{2 \times 2 \times n},$$

and it is not possible to make $\|A \cdot (U, U)_{1,2}\|$ larger with $U$ as in (28) (or any other structure). The solution matrix $W$ is the left singular matrix in the SVD,

$$\begin{pmatrix} \tau_1 \tilde{w}_1 & \tau_2 \tilde{w}_2 \end{pmatrix} = W \Sigma V^T = \begin{pmatrix} w_1 & w_2 \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \end{pmatrix}. $$

The core tensor is

$$F(:, :, 1) = \sigma_1 \begin{pmatrix} v_{11} & 0 \\ 0 & v_{21} \end{pmatrix}, \quad F(:, :, 2) = \sigma_2 \begin{pmatrix} v_{12} & 0 \\ 0 & v_{22} \end{pmatrix}. $$

Note that $\|F\|^2 = \tau_1^2 + \tau_2^2 = \sigma_1^2 + \sigma_2^2$. We see that (33) with these positive $u_1$ and $u_2$ is the $U$-part of the solution. □

We will refer to nonnegative vectors with the structure (33) as vectors in *indicator form*, as they exhibit the reducibility structure of the tensor.

Consider the special case when $\mathbf{A}$ has the structure

$$A(:, :, 1: n_1) = \begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix}, \quad A(:, :, n_1 + 1: n) = \begin{pmatrix} 0 & 0 \\ 0 & A_2 \end{pmatrix},$$

$A_i \in \mathbb{R}^{m_i \times m_i \times n}$, \quad $i = 1, 2, \quad m_1 + m_2 = m, \quad n_1 + n_2 = n,$
where the solutions of the rank-(1,1,1) problems for \( A_i \) are \((u_i, u_i, w_i)\) with core \( \tau_i, i = 1, 2 \). Then, under a technical assumption analogous to that in (27), the solution of the best rank-(2,2,2) problem for \( A \) is

\[
U = \begin{pmatrix} u_1 & 0 \\ 0 & u_2 \end{pmatrix}, \quad W = \begin{pmatrix} w_1 & 0 \\ 0 & w_2 \end{pmatrix}.
\]

\[
P(\cdot, \cdot, 1) = \begin{pmatrix} \tau_1 & 0 \\ 0 & 0 \end{pmatrix}, \quad P(\cdot, \cdot, 2) = \begin{pmatrix} 0 & 0 \\ 0 & \tau_2 \end{pmatrix}.
\]

Thus the solution mirrors exactly the structure of \( A \).

Next assume that the 3-slices of the tensor \( A \) are adjacency matrices of undirected bipartite graphs. This a special case of 3-reducibility, and the tensor has the structure as in the following proposition.

**Proposition 5.** Let the \((1,2)\)-symmetric, nonnegative tensor \( A \in \mathbb{R}^{m \times n \times n} \) have the structure

\[
A = \begin{pmatrix} 0 & B \\ B' & 0 \end{pmatrix}, \quad B \in \mathbb{R}^{m_1 \times n_1 \times n}, \quad m_1 + m_2 = m.
\]

Assume that the best rank-(1,1,1) approximation of \( B \) is unique, given by nonnegative \((p, q, w)\), and let \( B \cdot (p, q, w) = rw \). Define the matrix \( U \),

\[
U = \begin{pmatrix} 0 & p \\ q & 0 \end{pmatrix}.
\] (34)

Then the best rank-(2,2,1) approximation of \( A \) is given by \((U, U, w)\), and the core tensor is

\[
\mathbb{R}^{2 \times 2 \times 1} \ni F = \begin{pmatrix} 0 & r \\ r & 0 \end{pmatrix}.
\]

**Proof.** We will show that \((U, U, w)\) satisfies (21)–(23), and thus is a stationary point. Choose \( P_\perp, Q_\perp, \) and \( W_\perp \) such that \((pP_\perp), (qQ_\perp), \) and \((wW_\perp)\) are orthogonal matrices. Since for the rank-(1,1,1) problem the core tensor is a scalar, we see from (21) to (23) that

\[
P^1_{1_1} := B \cdot (P_\perp, q, w) = 0, \quad P^2_{2_1} := B \cdot (p, Q_\perp, w) = 0, \quad P^3_{3_1} := B \cdot (p, q, W_\perp) = 0.
\]

Clearly

\[
\mathbb{R}^{(m-2)\times m} \ni U_\perp = \begin{pmatrix} 0 & P_\perp \\ Q_\perp & 0 \end{pmatrix}.
\]

satisfies \(U^TU_\perp = 0\). Therefore

\[
P^1_\perp = A \cdot (U_\perp, U, w) = \begin{pmatrix} 0 & B \cdot (p, Q_\perp, w) \\ B ' \cdot (P_\perp, q, w) & 0 \end{pmatrix} = \begin{pmatrix} 0 & P^1_{1_1} \\ P^2_{2_1} & 0 \end{pmatrix} = 0,
\]

where we have used the identity \( B' \cdot (Q_\perp, p, w) = B \cdot (p, Q_\perp, w) \). Similarly one can prove \( P^3_\perp = 0 \). So (21)–(23) hold and \((U, U, w)\) is a stationary point. To prove that the point is optimal, let \( C = (X, X, z) \cdot H \), for arbitrary \( H \in \mathbb{R}^{2 \times 2 \times 1} \), and consider

\[
\left\| \begin{pmatrix} 0 & B \\ B' & 0 \end{pmatrix} - \begin{pmatrix} C_{11} & C_{12} \\ C'_{12} & C_{22} \end{pmatrix} \right\|^2 \geq \left\| \begin{pmatrix} 0 & B \\ B' & 0 \end{pmatrix} - \begin{pmatrix} 0 & C_{12} \\ C'_{12} & 0 \end{pmatrix} \right\|^2 = 2\|B - C_{12}\|^2.
\]
Here rank(C_{12}) must be equal to (1, 1, 1), because otherwise
\[
\begin{pmatrix}
0 & C_{12} \\
C'_{12} & 0
\end{pmatrix},
\]
would have rank higher than (2,2,1). Thus the optimum is achieved when \( C_{12} \) is the best rank-(1,1,1) approximation of \( B \).

The expression for \( F = A \cdot (U, U, w) \) follows directly from the identity \( B \cdot (p, q)_{1,2} = rw \).

It is possible to give heuristic arguments why, under certain assumptions, the best rank-(2,2,2) approximation problem for a tensor \( A \) that is (1,2)-reducible in the general sense of Definition 1, has a solution, where \( U \) and \( W \) are close to indicator form. However, such a discussion is rather complicated, and we refrain from giving the details here. Suffice it to say that if \( U \) and \( W \) are in indicator form, and the core tensor has the structure
\[
\begin{align*}
F(:, :, 1) &= \begin{pmatrix} f_{111} & 0 \\ 0 & f_{221} \end{pmatrix}, \\
F(:, :, 2) &= \begin{pmatrix} f_{112} & f_{122} \\ f_{122} & f_{222} \end{pmatrix},
\end{align*}
\]
then the approximating tensor \( C = (U, U, W) \cdot F \) is also (1,2)-reducible in the same way as \( A \). We refer to numerical examples in Section 7, which indicate that the optimal solution is approximately in indicator form.

## 6 Computing the Approximate Structure of Tensors That Are Close to Reducible

It is safe to assume that in real applications exactly reducible data tensors occur very seldom. In this section we will discuss how the results on reducible tensors can be used to ascertain the structure of tensors that are close to being reducible.

This is based on the theory of Section 5 and a perturbation result for the best rank-(\( r_1, r_2, r_3 \)) approximation for tensors that is an analogue to Proposition 2 for matrices. To avoid burdening the presentation with technical details, we give a rather loose formulation of the results.

**Proposition 6** (Reference 9, section 5). *If a tensor \( A \) is perturbed by \( \epsilon \), then the core tensor \( F \) of the best rank-(\( r_1, r_2, r_3 \)) approximation is perturbed by approximately the same amount \( \epsilon \). The solution matrices \((U, V, W)\) are perturbed essentially by \( \epsilon \) divided by a quantity that is similar to the distance to nearby eigenvalues in the eigenvalue case (the “gap”).*

It is shown in Reference 9 that the perturbation theory for the SVD and the symmetric eigenvalue problem are special cases of Proposition 6. Proposition 6 shows that, if a (1,2)-symmetric tensor is close to reducible, then the qualitative properties of the solution \((U, U, W)\) and \( F \) of the best rank-(\( r_1, r_1, r_3 \)) approximation are close to those of the exactly reducible nearby tensor.

Using the perturbation theory we can be more specific on how the core tensor is perturbed.

**Proposition 7.** Let \( A \geq 0 \) be as in Proposition 4, and let the perturbed tensor be
\[
\tilde{A} = \begin{pmatrix}
A_1 & \mathcal{E} \\
\mathcal{E}' & A_2
\end{pmatrix},
\]
where \( \mathcal{E} \geq 0 \) and \( \|A\| \gg \|\mathcal{E}\| =: \epsilon \). Then the perturbed core tensor of the best rank-(2,2,1) approximation is
\[
\tilde{F} = \begin{pmatrix}
f_1 & d \\
d & f_2
\end{pmatrix},
\]
where
\[
d = \mathcal{E} \cdot (u_1, u_2, w) + O(\epsilon^2).
\]

The proof is a straightforward, but somewhat lengthy, application of corollary 5.4 of Reference 9, and we omit it here. From Corollary 1 in this paper \( u_1, u_2, \) and \( w \) are positive, and therefore, for a tensor that is close to that in Proposition 7,
the quantity \(d\) is an indicator of how close the tensor is to being (1,2)-reducible (in a similar way as the second smallest eigenvalue of the Laplacian in spectral graph theory). The corresponding result can be stated for tensors with the structure in Proposition 5. Thus, \((U, U, W)\) and \(F\) can be used to ascertain the structure of the tensor, and partition it accordingly.

In the discussion on structure determination we first consider tensors that are exactly reducible. Clearly, we must handle the Grassmann nonuniqueness of the solution of the best approximation problem. Let the best rank-\((r_1, r_1, r_3)\) approximation be \(C = (U, U, W) \cdot F\). Then, for any orthogonal \(Q_1 \in \mathbb{R}^{r_1 \times r_1}\) and \(Q_3 \in \mathbb{R}^{r_3 \times r_3}\),

\[
C = (UQ_1, UQ_1, WQ_3) \cdot (F \cdot (Q_1, Q_1, Q_3)).
\]

For the reducible problems we have considered in Propositions 4 and 5, we have written the solution in indicator form, for example, (28). If we multiply a pair of vectors in indicator form by a rotation matrix, we get

\[
UQ_1 = \begin{pmatrix} u_1 & 0 \\ 0 & u_2 \end{pmatrix} \begin{pmatrix} c & s \\ -s & c \end{pmatrix} = \begin{pmatrix} cu_1 & su_1 \\ -su_2 & cu_2 \end{pmatrix}.
\]

Due to the Grassmann indeterminacy, we can expect the solution to be in the form of the right-hand side of (35). Since \(u_1 \geq 0\) and \(u_2 \geq 0\), one of the columns remains positive after multiplication, while the other has positive and negative elements (the same happens if we multiply by a rotation matrix with determinant \(-1\)). This implies that we can read off the structure of the tensor by looking at the positive and negative entries of one of the vectors. For the tensor in nonreduced form, say \((P_1, P_1, P_3) \cdot A\), for permutation matrices \(P_1\) and \(P_3\), the rows of (35) are permuted by \(P_1\); thus, the structure of the tensor can easily be found by reordering in increasing order the columns of \(UQ_1\) that has both negative and positive entries, and reordering the tensor accordingly.

But the positive-negative pattern of one of the vectors of \(U\) (and also \(W\)), is not enough. To determine if the tensor is reducible, we need to transform the computed \(U\) and \(W\) to indicator form, and transform the core \(F\) tensor accordingly; if the core has one of the structures in the propositions in Section 5, then the tensor is reducible. Thus we want to determine the rotation that restores the indicator form.

Assume that the rows of \(U\) have been reordered so that the negative elements of are at the bottom of the second column and that the first column is positive. We want to find the rotation that puts the vectors in indicator form:

\[
\begin{pmatrix} u_{11} & u_{12} \\ u_{21} & -u_{22} \end{pmatrix} \begin{pmatrix} c & -s \\ s & c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.
\]

A simple computation gives

\[
\hat{U}q := \begin{pmatrix} u_{12} & -u_{11} \\ u_{21} & -u_{22} \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.
\]

The vector \(q\) is the right singular vector corresponding to the smallest singular value of \(\hat{U}\). Thus, we can transform \(U\) to indicator form if the smallest singular value of \(\hat{U}\) is equal to zero, and the transformation is determined from the corresponding right singular vector.

The same procedure can be used if the tensor is close to reducible. Assume that we have determined orthogonal matrices \(Q_1\) and \(Q_3\) that transform the computed \(U\) and \(W\) to approximate indicator form (in the rank-(2,2,1) case \(Q_3 = 1\) and \(W\) is not transformed). The corresponding tensor is \(F \cdot (Q_1, Q_1, Q_3)\), and we will refer to it as the structure tensor.

An algorithm for computing the structure of a (1,2)-symmetric tensor can be based on the successive computation of best rank-(2,2,1), rank-(2,2,2), and rank-(2,2,3) approximations, along with indicator vectors and structure tensor, and using the characterisations of the solutions for different structures in Section 5. This is illustrated in Section 7.

7 | EXAMPLES

We first give four examples with synthetic data that illustrate the reducibility structures in the Propositions in Section 5, and describe how our algorithm finds the structure. Then we apply the method to a tensor with real data and show how topics in news text can be extracted using the partitioning. Further examples with real data are given in Reference 21.
7.1 Synthetic data

For the examples we created random, sparse, nonnegative and (1,2)-symmetric tensors in $\mathbb{R}^{200 \times 200 \times 200}$ with certain reducibility structures, and added a (1,2)-symmetric perturbation, sparser than the original tensor. The perturbation made the tensor irreducible. All tensor elements were put equal to 1, before we normalized the 3-slices as described in Section 4.2. We then performed a (1,2)-symmetric permutation, and, in some cases, another permutation in the third mode. We refer to this as the permuted tensor. To verify the theory in Sections 5 and 6 we then restored the reducibility pattern via a best low-rank approximation. The approximation was computed by the Krylov–Schur-like BKS algorithm with relative termination criterion $10^{-12}$. After having been initialized with one HOOI iteration for rank-$(1, 1, 1)$, the BKS method converged in three to seven iterations.

The experiments were performed using Matlab on a desktop computer with 8 GB primary memory. To give a rough idea about the amount of work, we mention that the execution time for Examples 3–6 was between 1.5 and 3.5 s.

In order to investigate how robust the reconstruction of reducibility structure is, we made the perturbed tensors relatively far from reducible, see the conductance measures in the examples. The indicator vectors always gave the correct structure. In Examples 3–4, when we made the perturbation smaller, the structure tensors became closer to those in Section 5, compare with Proposition 7. With no perturbation in Examples 3 and 4 the vectors were in indicator form to working precision.

Example 3. We constructed a tensor with the structure close to that in Proposition 4. The original tensor is illustrated in Figure 2, where we have used a 3D equivalent of the Matlab spy plot. To exhibit the sparsity, we also give a spy plot of a 3-slice. We computed the best rank-(2,2,1) approximation, reordered $U(:, 2)$ to make it decreasing, computed indicator vectors and structure tensor. The vectors are illustrated in Figure 3.

Applying the reordering of the rows of $U$ to the permuted tensor, we got the one illustrated in Figure 4. The structure is seen already in the reordered vectors, but before we can draw any safe conclusion must look at the structure tensor,

$$F_{\text{struct}} = \begin{pmatrix} 11.7 & 0.980 \\ 0.980 & 10.0 \end{pmatrix}.$$

This indicates that we have identified the correct structure. Further confirmation is given by the fact that the conductance that was 0.3227 before the initial permutation was the same to 16 digits after the reconstruction and partitioning. The norm of the core tensor and the original tensor were $\| F \| = 15.46$ and $\| A \| = 116.8$. To check if there was any undetected significant structure in the third mode we also computed the best rank-(2,2,2) approximation. The norm of the core tensor now was 15.48, which shows that practically no structure was present that required a rank-(2,2,2) approximation. Moreover, while the $U$ vectors remained similar to those in Figure 3, the computed indicator vectors for $W$ did not reveal
**FIGURE 3**  Example 3, rank-(2,2,1) approximation. The column vectors of the solution $U$ after reordering that makes $U(:, 2)$ decreasing (upper left), indicator vectors (upper right), and the vector $w$ (lower left). $U(:, 1)$ is solid and blue, $U(:, 2)$ is dashed and red.

**FIGURE 4**  Example 3. The reconstructed tensor

any third mode structure. Thus we can safely conclude that the correct structure was identified using the rank-(2,2,1) approximation.

**Example 4.** The tensor was constructed with structure close to that of Proposition 5. The results after computing a rank-(2,2,1) approximation are given in Figures 5 and 6.

The structure tensor was

$$F_{\text{struct}} = \begin{pmatrix} 1.56 & 9.90 \\ 9.90 & 0.850 \end{pmatrix}.$$
Example 4. Rank-(2,2,1) approximation. The reordered column vectors of $U$ (upper left), indicator vectors for $U$ (upper right), $w$ (lower left).

Figure 6. Example 4. The reconstructed tensor.

which is a clear indication for the structure of Proposition 5. The norm of the original tensor $A$ was 122.7 and the norm of $F$ was 14.11; the norm of the core tensor in a rank-(2,2,2) approximation was exactly the same. Further, for the rank-(2,2,2) the smallest singular value of the matrix $\hat{W}$ in the computation of indicator vectors was 0.48 (as compared to 0.02 for $\hat{U}$). This shows that there was no significant mode-3 structure that was missed in the rank-(2,2,1) approximation. The conductance was 0.7206 before and after reconstruction, which shows that the correct structure is computed.

Example 5. We constructed a tensor close to a general (1,2)-reducible structure (18). For visibility reasons we do not give a spy plot of the tensor but of two slices, see Figure 7. Rank-(2,2,1) and rank-(2,2,2) approximations were computed with $\|F\| = 13.81$ and $\|F\| = 14.64$, respectively. Clearly there was significant mode-3 structure that was visible in the rank-(2,2,2) approximation. However, with rank-(2,2,3) the norm of the core tensor did not become larger.
After computing the rank-(2,2,2) approximation, we reordered the vectors and applied the reorderings to the tensor. The vectors are illustrated in Figure 8. The structure tensor was

\[ F_{\text{struct}}(:, :, 1) = \begin{pmatrix} 5.83 & 4.75 \\ 4.75 & 3.86 \end{pmatrix}, \quad F_{\text{struct}}(:, :, 2) = \begin{pmatrix} 8.26 & 0.685 \\ 0.685 & 7.15 \end{pmatrix}. \]

The singular values of the matrix \( \hat{U} \) in (36) were 1.414 and 0.040, and those of the corresponding \( \hat{W} \) were 1.414 and 0.052. The original conductance was 0.3381 and that of the reconstructed and partitioned tensor was exactly the same.

We also ran a similar example, where the tensor was exactly reducible. The structure tensor was

\[ F_{\text{struct}}(:, :, 1) = \begin{pmatrix} 5.77 & 4.79 \\ 4.79 & 3.77 \end{pmatrix}, \quad F_{\text{struct}}(:, :, 2) = \begin{pmatrix} 8.52 & 0.0047 \\ 0.0047 & 7.50 \end{pmatrix}. \]

This shows clearly that for the exactly reducible tensor the computed indicator vectors are only in approximate indicator form.

**Example 6.** A tensor with almost 3-reducible structure close to (19) was constructed and analyzed. The results were analogous to those in Example 5, and we therefore omit them here.

### 7.2 Real data

As we remarked earlier it is probably rare that tensor data are close to reducible. In Reference 21 we perform experiments, where we investigate if best low-rank approximation can be used for the analysis of real data by clustering. Here a brief account is given of an example from that paper.

**Example 7. Clustering of news texts.** The tensor is a collection of graphs constructed from a corpus of news texts published by Reuters during 66 days after the September 11 attack on the United States. The vertices are words; there is an
edge between two words if and only if they appear in the same text unit (sentence). The weight of an edge is its frequency. The tensor has dimension $13,332 \times 13,332 \times 66$ and is $(1,2)$-symmetric. Naturally it is impossible to discern any structure by visual inspection of a spy plot of the tensor.

We computed a best rank-(2,2,2) approximation of the tensor using the BKS method, where the outer iterations were intialized with five HOOI iterations for rank-(1,1,1). With relative termination criterion $10^{-13}$, 60 iterations were required, and the execution time was about 67 s. The reordered $U$ and $W$ vectors are shown in Figure 9. The core tensor was

$$F(\cdot, \cdot, 1) = \begin{pmatrix} 5.06 & -0.0021 \\ -0.0021 & 3.42 \end{pmatrix}, \quad F(\cdot, \cdot, 2) = \begin{pmatrix} -0.133 & 0.240 \\ 0.240 & 0.211 \end{pmatrix},$$

and $\|F\| = 6.12$, $\|F(\cdot, \cdot, 1)\| = 6.11$, and $\|F(\cdot, \cdot, 2)\| = 0.42$. As we can write the approximating tensor (in outer product form)

$$C \approx (u_1, u_1, w_1) \cdot 5.06 + (u_2, u_2, w_1) \cdot 3.42 + (U, U, w_2) \cdot F(\cdot, \cdot, 2),$$

where 5.06 and 3.42 are to be considered as tensors, we see that the first two terms dominate. The shape of $u_1$ and $u_2$ show that the main “mass” of $C$ is at the four corners in modes 1 and 2. From the shape of $w_1$ we see that the approximation is relatively constant in time.

After the reordering of the vectors (that represent words), the two main topics of the texts are at the beginning and at the end, see Table 1. Note the difference between the two topics: Topic 1 deals with the New York–Washington scene, and Topic 2 with the perpetrators.

A large number of insignificant words are at the middle (corresponding to small components of the vector $u_1$). Note that these words can not be considered as stop words in the sense of information retrieval, but most of them are certainly insignificant with respect to the task of finding the main topics in the texts.

The reordered tensor is shown in Figure 10. It is seen that the “mass” is concentrated to the four corners in term modes, which correspond to Topics 1 and 2 in Table 1. We computed the norms of the subtensors of dimensions $1000 \times 1000 \times 66$ at the corners (cf, the shape of $U$ in Figure 9), and they were 50.6, 25.5, 25.5, and 35.0, respectively. As the norm of $A$
FIGURE 9  The column vectors of $U$ and $W$ after reordering

TABLE 1  Beginning terms, middle terms, and end terms after reordering. The table is organized so that the most significant terms appear at the top of the first column and the bottom of the third (consistent with the reordered vectors). The second column shows a few terms at the middle, which are relatively insignificant with respect to the topics in the other two columns.

| Topic 1          | Insignificant | Topic 2  |
|------------------|---------------|----------|
| world_trade_ctr  | encouragement | hand     |
| pentagon         | continuous    | ruler    |
| new_york         | asymmetric    | saudi    |
| attack           | phrase        | guest    |
| hijack           | secrecy       | mullah   |
| plane            | larkin        | movement |
| airliner         | nebraska      | camp     |
| tower            | windsor       | harbor   |
| twin             | humidity      | shelter  |
| washington       | airfare       | group    |
| sept             | irby          | organization |
| suicide          | x-ray         | leader   |
| pennsylvania     | at&t          | rule     |
| people           | motorcycle    | exile    |
| passenger        | coolness      | guerrilla|
| jet              | N1082         | afghanistan|
| mayor            | cooperate     | fugitive |
| 110-story        | bangladeshi   | dissident|
| mayor_giuliani   | forget        | taliban  |
| hijacker         | plea          | islamic  |
| commercial       | investigations| network |
| aircraft         | biowarfare    | al_quaeda|
| assault          | washington-new_york | militant |
| airplane         | account       | saudi-born|
| miss             | grant         | bin_laden|
was 128.6, about 55% of the total mass was concentrated at those corners. This shows that the terms of Topics 1 and 2 co-occurred to a large extent with terms within the same topic. In addition, there was considerable co-occurrence between terms from the two main topics.

We concluded above that the time dependence of occurrence of the two topics was minor. This is confirmed by the fact that the norm of the core tensor for a rank-(2,2,1) approximation was 6.11, and that the vectors $u_1, u_2,$ and $w_1$ were very similar to the rank-(2,2,2) case. The topics were almost identical.

Clearly, the tensor of this example is far from reducible. Still, the example indicates that the partitioning given by the rank-(2,2,2) approximation orders the vertices of the co-occurrence graphs so that the dense parts of the graphs are at both ends. The middle part is sparser, corresponding to terms that can be considered as noise (when we removed the 9000 middle terms, and repeated the analysis, the same topics were produced).

It is interesting to apply the procedure to text corpuses with more than two topics. In Reference 21 we apply the partitioning recursively to the Reuters corpus. Our experiments indicate that it is possible to find subtopics of the two main topics.

8 | CONCLUSIONS

In this paper we have presented a generalization of spectral graph partitioning to sparse tensors, where the 3-slices are normalized adjacency matrices of undirected graphs. We have shown, theoretically and experimentally, that for some cases cases of tensors that are close to reducible, it is possible to determine the reducibility structure by computing a low-rank best multilinear approximation of the tensor. The algorithm was applied to synthetic data and it was shown that the reducibility structure is accurately computed.

In this paper we applied the methods only to (1,2)-symmetric tensors. The applicability to nonsymmetric tensors will be investigated in our future research. We will also investigate the applicability of the methods to the partitioning of 4-tensors that are (1,2,3)-symmetric, compare with Reference 5.

Due to the presence of noise, it is probably rare to find tensors in applications that are close to reducible. To illustrate the performance of the algorithm for more general tensors, we gave one example with data from a text analysis application, where most of the tensor data are noise, and there is no sharp distinction between useful information and noise. Here the best low-rank approximation gave a clustering that separated relevant information from noise. This property is likely to make the methods of this paper useful in applications, but more research is needed to investigate the properties of the methods for general problems. In Reference 21 we analyze tensors from a few different data science applications. It is shown that the methods can be used to extract information from large tensors with noisy data.
ACKNOWLEDGEMENTS
The authors would like to thank the referees for useful comments that helped to improve the paper. This study does not have any conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available in Pajek datasets at http://vlado.fmf.uni-lj.si/pub/networks/data/CRA/terror.htm. These data were derived from the following resources available in the public domain: - Days, http://vlado.fmf.uni-lj.si/pub/networks/data/CRA/DaysAll.zip

ENDNOTES
*The notation (3) and (4) was suggested in Reference 23. An alternative notation was earlier given in Reference 24. Our $(X)_{d} \cdot A$ is the same as $A \times_{d} X$ in the latter system.

†The standard notation would be $Gr(r_1, m) \times Gr(r_2, m) \times Gr(r_3, n)$. As the dimensions of the vector spaces $(m$ and $n$) are implicit in the problem formulation, we use (17), for simplicity.

‡In Reference 43 such temporal dynamics is discovered using the time factors of a CP decomposition of the tensor.

§That is, so that (1,2)-symmetry is preserved.

¶Note that the counterexamples in Reference 45 do not imply the nonexistence of a nonnegative solution of the best rank-(1,1,1) approximation problem for a nonnegative tensor.

#Using our notation for tensor-matrix multiplication, we have $u \otimes v \otimes w = (u, v, w) \cdot \mathbb{1}$, where $\mathbb{R}^{1\times1\times1} \ni \mathbb{1}$ is a tensor.

ORCID
Lars Eldén https://orcid.org/0000-0003-2281-856X

REFERENCES
1. Chung F. Spectral graph theory. Proceedings of the CBMS Regional Conference Series in Mathematics Number 92. American Mathematical Society; 1997.
2. von Luxburg U. A tutorial on spectral clustering. Stat Comput 2007;17:395–416. 10.1007/s11222-007-9033-z.
3. Martin JG, Canfield ER. Ranks and representations for spectral graph bisection. SIAM J Sci Comput. 2009;31(5):3529–46. Available from. http://link.aip.org/link/?SCE/31/3529/1
4. Urschel JC, Zikatanov LT. Spectral bisection of graphs and connectedness. Linear Algebra Appl. 2014;449:1–16. Available from. http://www.sciencedirect.com/science/article/pii/S0024379514000743
5. Ge Y, Peng P, Lu H. Mixed-order spectral clustering for complex networks. Pattern Recognit. 2021;117:107964. Available from. https://www.sciencedirect.com/science/article/pii/S0031320320302442
6. Hitchcock FL. Multiple invariants and generalized rank of a p-way matrix or tensor. J Math Phys Camb. 1927;7:39–70.
7. Tucker LR. The extension of factor analysis to three-dimensional matrices. In: Gulliksen H, Frederiksen N, editors. Contributions to mathematical psychology. New York, NY: Holt, Rinehart and Winston; 1964. p. 109–27.
8. De Lathauwer L, Moor BD, Vandewalle J. On the best rank-1 and rank-$$(R_1, R_2, \ldots, R_N)$$ approximation of higher-order tensors. SIAM J Matrix Anal Appl. 2000;21:1324–42.
9. Eldén L, Savas B. Perturbation theory and optimality conditions for the best multilinear rank approximation of a tensor. SIAM J Matrix Anal Appl. 2011;32:1422–50. Available from. http://epubs.siam.org/doi/full/10.1137/110829250
10. Kolda T, Sun J. Scalable tensor decompositions for multi-aspect data mining. Proceedings of the 2008 8th IEEE International Conference on Data Mining; Pisa, Italy. 2008. p. 363–72.
11. Xiao C, Yang C, Li M. Efficient alternating least squares algorithms for low multilinear rank approximation of tensors. J Sci Comput. 2021;87:67. https://doi.org/10.1007/s10915-021-1493-0
12. Jegelka S, Sra S, Banerjee A. Approximation algorithms for tensor clustering. Proceedings of the 20th International Conference on Algorithmic Learning Theory (ALT); Porto, Portugal, 2009. p. 368–83.
13. Cao X, Wei X, Han Y, Yang Y, Lin D. Robust tensor clustering with non-greedy maximization. Proceedings of the 23rd International Joint Conference on Artificial Intelligence IJCAI’13; Beijing, 2013. p. 1254–9.
14. Benson AR, Gleich DF, Leskovec J. Tensor spectral clustering for partitioning higher-order network structures. Proceedings of the 2015 SIAM International Conference on Data Mining. SIAM; Vancouver, 2015. p. 118–26. https://doi.org/10.1137/1.9781611974010.14.
15. Wei Z, Zhao H, Zhao L, Yan H. Multiscale co-clustering for tensor data based on canonical polyadic decomposition and slice-wise factorization. Inf Sci. 2019;503:72–91. Available from. http://www.sciencedirect.com/science/article/pii/S0020025519305833
16. Chen Y, Xiao X, Zhou Y. Multi-view subspace clustering via simultaneously learning the representation tensor and affinity matrix. Pattern Recognit. 2020;106:107441. Available from. http://www.sciencedirect.com/science/article/pii/S0031320320302442
17. Xu H, Zhang X, Xia W, Gao Q, Gao X. Low-rank tensor constrained co-regularized multi-view spectral clustering. Neural Netw. 2020;132:245–52. Available from. http://www.sciencedirect.com/science/article/pii/S0893608020303087
18. Boutalbi R, Labiod L, Nadif M. Implicit consensus clustering from multiple graphs. Data Mining Knowl Discov. 2021;35:2313–40. https://doi.org/10.1007/s10618-021-00788-y
19. Liu X, De Lathauwer L, Janssens F, De Moor B. Hybrid clustering of multiple information sources via HOSVD. In: Zhang L, Lu BL, Kwok J, editors. Advances in neural networks. Lecture Notes in Computer Science. Volume 6064. Berlin, Heidelberg: Springer; 2010. p. 337–45. https://doi.org/10.1007/978-3-642-13318-3_42

20. Eldén L, and Dehghan M. 2020. A Krylov-Schur like method for computing the best rank-$(r_1, r_2, r_3)$ approximation of large and sparse tensors; 2012. 07595. arXiv, math.NA.

21. Eldén L, and Dehghan M. 2020. Analyzing large and sparse tensor data using spectral low-rank approximation; 2012. 07754. arXiv, math.NA.

22. Lim LH. Tensors in computations. Acta Numer. 2021;30:555–764.

23. de Silva V, Lim LH. Tensor rank and the ill-posedness of the best low-rank approximation problem. SIAM J Matrix Anal Appl. 2008;30(3):1084–127. Available from. http://link.aip.org/link/?SML/30/1084/1

24. De Lathauwer L, Moor BD, Vandewalle J. A multilinear singular value decomposition. SIAM J Matrix Anal Appl. 2000;21:1253–78.

25. Bader, Kolda T. Algorithm 862: MATLAB tensor classes for fast algorithm prototyping. ACM Trans Math Softw. 2006;32:635–53. Available from. http://csmr.ca.sandia.gov/~tgkolda/TensorToolbox/

26. Zhang T, Golub GH. Rank-one approximation to higher order tensors. SIAM J Matrix Anal Appl. 2002;23:534–50.

27. Eldén L, Savas B. A Newton-Grassmann method for computing the best multi-linear rank-$(r_1, r_2, r_3)$ approximation of a tensor. SIAM J Matrix Anal Appl. 2009;31:248–71.

28. Ishteva M, Lathauwer LD, Absil PA, Huffel SV. Differential-geometric Newton method for the best rank-$(R_1, R_2, R_3)$ approximation of tensors. Numer Algorithms. 2008;51:179–94.

29. Absil PA, Mahoney R, Sepulchre R, Dooren PV. A Grassman-Rayleigh quotient iteration for computing invariant subspaces. SIAM Rev. 2002;44:57–73.

30. Edelman A, Arias T, Smith ST. The geometry of algorithms with orthogonality constraints. SIAM J Matrix Anal Appl. 1998;20:303–53.

31. Absil PA, Mahony R, Sepulchre R. Optimization algorithms on matrix manifolds. Princeton, NJ: Princeton University Press; 2007.

32. Frandsen A, Ge R. Optimization landscape of Tucker decomposition. Math Program Ser B. 2020.

33. Lehoucq R, Sorensen D, Yang C. Arpack users’ guide: solution of large scale eigenvalue problems with implicitly restarted Arnoldi methods. Philadelphia, PA: SIAM; 1998.

34. Stewart GW. A Krylov-Schur algorithm for large eigenproblems. SIAM J Matrix Anal Appl. 2002;23(3):601–14. Available from. http://link.aip.org/link/?SML/23/601/1

35. Savas B, Eldén L. Krylov-type methods for tensor computations I. Linear Algebra Appl. 2013;438:891–918. Available from. http://www.sciencedirect.com/science/article/pii/S0024379511007877

36. Fiedler M. Algebraic connectivity of graphs. Czech Mat J. 1973;23:298–305.

37. Horn RA, Johnson CR. Matrix analysis. Cambridge, UK: Cambridge University Press; 1985.

38. Golub GH, Van Loan CF. Matrix computations. 4th ed. SIAM: Philadelphia, PA; 2013.

39. Eldén L. Matrix methods in data mining and pattern recognition. 2nd ed. SIAM: Philadelphia, PA; 2019.

40. Meyer CD. Matrix analysis and applied linear algebra. Philadelphia, PA: SIAM; 2000.

41. Kolda T, Bader B. The TOPHITS model for higher-order web link analysis. Proceedings of the Workshop on Link Analysis, Counterterrorism and Security; Bethesda, Maryland, 2006.

42. Bader BW, Harshman RA, and Kolda TG. Temporal analysis of semantic graphs using ASALSAN. Proceedings of the 7th IEEE International Conference on Data Mining (ICDM 2007); Omaha, NE, 2007. p. 33–42.

43. Dunlavy DM, Kolda TG, Acar E. Temporal link prediction using matrix and tensor factorizations. ACM Trans Knowl Discov Data. 2011;5(2):10:27 pages.

44. Chang KC, Pearson K, Zhang T. Perron-Frobenius theorem for nonnegative tensors. Commun Math Sci. 2008;6(2):507–20.

45. Friedland S, Gaubert S, Han L. Perron-Frobenius theorem for nonnegative multilinear forms and extensions. Linear Algebra Appl. 2013;438(2):738–49. Available from. http://www.sciencedirect.com/science/article/pii/S0024379511001765

46. Gautier A, Hein M. Tensornorm and maximalsingular vectorsofnonnegativetensors-APerron-Frobeniustheorem,aCollatz-Wielandt characterization and a generalized power method. Linear Algebra Appl. 2016;505:313–43. Available from. http://www.sciencedirect.com/science/article/pii/S0024379516301239

47. Baeza-Yates R, Ribeiro-Neto B. Modern Information Retrieval. New York, Addison-Wesley: ACM Press; 1999.

48. Berry MW, Browne M. Understanding search engines. mathematical modeling and text retrieval. 2nd ed. Philadelphia, PA: SIAM; 2005.

---

**How to cite this article:** Eldén L, Dehghan M. Spectral partitioning of large and sparse 3-tensors using low-rank tensor approximation. Numer Linear Algebra Appl. 2022;29(5):e2435. https://doi.org/10.1002/nla.2435
APPENDIX A. NONNEGATIVITY OF THE BEST RANK-(1,1,1) SOLUTION FOR NONNEGATIVE TENSORS

Using Perron–Frobenius theory for nonnegative matrices it can be proved that the eigenvector corresponding to the dominant eigenvalue is nonnegative, and that for a nonnegative irreducible matrix it is positive (Reference 40, chapter 8). Generalizations to tensors are given in References 45,46, where the norms are $p$-norms with restrictions on the value of $p$. We here give a proof of Lemma 3 that the best rank-(1,1,1) approximation of a nonnegative 3-tensor is nonnegative, and that the same problem for an irreducible, (1,2)-symmetric tensor has a positive solution.

**Proof.** We first show that the solution is nonnegative. Here we do not assume (1,2)-symmetry or irreducibility. Let $(u, v, w)$ be a solution. Without loss of generality we can assume that

$$\max_{\|x\| = \|y\| = \|z\| = 1} A \cdot (x, y, z) = A \cdot (u, v, w) = 1,$$

and that

$$u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \quad v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \quad w = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix},$$

where $u_1 > 0, v_1 > 0, w_1 > 0$, and $u_2 \leq 0, v_2 \leq 0$, and $w_2 \leq 0$. Assume that the tensor is partitioned accordingly in the three modes and denote the subtensors $A_{\lambda\mu\nu}$, for $\lambda, \mu, \nu = 1, 2$. Let $|x|$ denote the vector of absolute values of the components of $x$, and let $u \otimes v \otimes w \in \mathbb{R}^{k \times k \times k}$ denote the tensor (outer) product of the vectors. Then

$$\|A - u \otimes v \otimes w\|^2 = \|A_{111} - u_1 \otimes v_1 \otimes w_1\|^2 + \|A_{121} + u_1 \otimes |v_2| \otimes w_1\|^2 + \cdots$$

$$+ \|A_{212} - |u_2| \otimes v_1 \otimes |w_2|\|w_1\|^2$$

$$+ \|A_{222} + |u_2| \otimes |v_2| \otimes |w_2|\|^2.$$  \hspace{1cm} (A1)

where we have omitted four similar terms. Since all subtensors are nonnegative, we have, for example,

$$\|A_{121} + u_1 \otimes |v_2| \otimes w_1\|^2 \geq \|A_{212} - |u_2| \otimes v_1 \otimes |w_2|\|w_1\|^2,$$  \hspace{1cm} (A2)

The only situation when strict inequality can occur is if a component of $A_{121}$ is positive, and the corresponding component of $v_2$ is nonzero; however, that would be a contradiction to the assumption that $(u, v, w)$ is a solution. Thus, we have equality in (A3). It follows that we can replace all + in (A1) and (A2) by −, and have

$$\|A - u \otimes v \otimes w\|^2 = \|A - x \otimes y \otimes z\|^2,$$  \hspace{1cm} (A4)

where

$$x = \begin{pmatrix} u_1 \\ |u_2| \end{pmatrix}, \quad y = \begin{pmatrix} v_1 \\ |v_2| \end{pmatrix}, \quad z = \begin{pmatrix} w_1 \\ |w_2| \end{pmatrix}.$$

Thus we have a nonnegative solution $(x, y, z)$.

In the proof of the second statement we use the elementary fact that if a function $F(\alpha)$ that is defined in $[0, 1]$ and differentiable in $[0, 1)$, satisfies $F'(\alpha) < 0$ for $\alpha \in [a_1, 1)$ for some $a_1$, then there exists $\bar{\alpha} \in (a_1, 1)$ such that $F(\bar{\alpha}) > F(1)$.

Let

$$u = \begin{pmatrix} u_0 \\ 0 \end{pmatrix}, \quad \|u_0\| = 1, \quad u_0 > 0,$$

$$v = \begin{pmatrix} v_0 \\ 0 \end{pmatrix}, \quad \|v_0\| = 1, \quad v_0 > 0.$$
be arbitrary nonnegative vectors (it is no restriction to assume that they have this structure: just reorder the tensor). Let $u_1 > 0$ and $w_1 > 0$ be arbitrary of suitable dimensions, with $\|u_1\| = \|w_1\| = 1$, and define the unit length vectors

$$
x(a) = \left(\frac{a u_0}{\sqrt{1 - a^2}} u_1\right), \quad a \in [0, 1], \quad z(\rho) = \left(\frac{\rho w_0}{\sqrt{1 - \rho^2}} w_1\right), \quad \rho \in [0, 1].
$$

Partition $\mathcal{A}$ in three ways, conforming with the structure of $x$ and $z$.

$$
\mathcal{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^t & A_{22} \end{pmatrix} = \begin{pmatrix} A_1 & A_2 \\ A_2^t & A_2 \end{pmatrix} = \begin{pmatrix} A_{111} & A_{112} & A_{112} \\ A_{121} & A_{122} & A_{122} \\ A_{122} & A_{122} & A_{222} \end{pmatrix};
$$

here $(A_1 | A_2)$ denotes a partitioning in the third mode. Define $F(a, \rho) = \mathcal{A} \cdot (x, x, z)$ and write

$$
F(a, \rho) = \rho A_1 \cdot (x, x, w_0) + \sqrt{1 - \rho^2} A_2 \cdot (x, x, w_1),
$$

and

$$
\frac{\partial F}{\partial \rho}(a, \rho) = A_1 \cdot (x, x, w_0) - \frac{\rho}{\sqrt{1 - \rho^2}} A_2 \cdot (x, x, w_1).
$$

Thus

$$
\frac{\partial F}{\partial \rho}(1, \rho) = A_{111} \cdot (u_0, u_0, w_0) - \frac{\rho}{\sqrt{1 - \rho^2}} A_{112} \cdot (u_0, u_0, w_0).
$$

As $\mathcal{A}$ is irreducible, $A_{112} \cdot (u_0, u_0, w_1) > 0$ (because otherwise $\mathcal{A}$ would be 3-reducible). Therefore $\frac{\partial F}{\partial \rho}(1, \rho) < 0$ for $\rho$ close to 1, and there is an $\bar{\rho}$ such that $F(1, \bar{\rho}) > F(1, 1)$.

In a similar vein we can write

$$
F(a, \rho) = a^2 A_{11} \cdot (u_0, u_0, z) + 2a \sqrt{1 - a^2} A_{12} \cdot (u_0, u_1, z) + (1 - a^2) A_{22} \cdot (u_1, u_1, z),
$$

and

$$
\frac{\partial F}{\partial a}(a, 1) = 2(a A_{11} \cdot (u_0, u_0, w_0) + \frac{1 - 2a^2}{\sqrt{1 - a^2}} A_{12} \cdot (u_0, u_1, w_0)

- a A_{22} \cdot (u_1, u_1, w_0)).
$$

Irreducibility implies that $A_{121} \cdot (u_0, u_1, w_0) > 0$ (because otherwise $\mathcal{A}$ would be (1,2)-reducible); therefore, $\frac{\partial F}{\partial a}(a) < 0$ for $a$ close to 1, and there is an $\bar{a}$ such that $F(\bar{a}, 1) > F(1, 1)$.

It follows that the solution must be positive.

**APPENDIX B. LEMMA FOR A RANK-2 MATRIX**

**Lemma 5.** Let $U \in \mathbb{R}^{m \times 2}$, $m > 2$, satisfying $U^T U = I$, and let $A \in \mathbb{R}^{2 \times 2}$ be symmetric. Put

$$
Z = UAU^T = \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{12}^t & Z_{22} \end{pmatrix}, \quad Z_{12} \in \mathbb{R}^{m_1 \times m_2}, \quad m_1 + m_2 = m,
$$

and let the columns of $U$ be partitioned accordingly.

Then $Z_{12} = 0$ and $\text{rank}(Z) = 2$ if and only if

$$
U = \begin{pmatrix} u_1 & 0 \\ 0 & u_2 \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}, \quad \text{rank}(A) = 2,
$$

(B1)
or

\[
U = \begin{pmatrix}
  u_{11} & u_{12} \\
  0 & 0
\end{pmatrix}, \quad \text{rank}(A) = 2, \quad (B2)
\]

or

\[
U = \begin{pmatrix}
  0 & 0 \\
  u_{21} & u_{22}
\end{pmatrix}, \quad \text{rank}(A) = 2. \quad (B3)
\]

**Proof.** First note that, for any orthogonal \(Q \in \mathbb{R}^{2 \times 2}\),

\[
Z = (UQ)(Q^TAQ)(UQ)^T, \quad (B4)
\]

so the factors in the product are not unique. Let

\[
A = \begin{pmatrix}
  a_{11} & a_{12} \\
  a_{12} & a_{22}
\end{pmatrix}, \quad U = \begin{pmatrix}
  x_1 & y_1 \\
  x_2 & y_2
\end{pmatrix},
\]

with the obvious partitioning of the column vectors in \(U\). By straightforward computation

\[
Z_{12} = a_{11}x_1x_2^T + a_{22}y_1y_2^T + a_{12}(y_1x_2^T + x_1y_2^T). \quad (B5)
\]

Thus, any of the \(U\)-structures in (B1) or (B2) implies \(Z_{12} = 0\).

Then assume that \(Z_{12} = 0, x_1 \neq 0\) and \(x_2 \neq 0\). By multiplying (B5) from the left and right by \(x_1^T\) and \(x_2\), respectively, we see that

\[
y_1 = r_1x_1, \quad y_2 = r_2x_2,
\]

for some \(r_1\) and \(r_2\). The requirements that the column vectors are orthogonal and have length 1, give

\[
U = \begin{pmatrix}
  x_1 & \tau x_1 \\
  x_2 & -\frac{1}{\tau}x_2
\end{pmatrix}, \quad \tau = \frac{\|x_2\|}{\|x_1\|}.
\]

Inserting this in (B5) we get

\[
0 = Z_{12} = \gamma x_1x_2^T := \left( a_{11} - a_{22} + a_{12}\left( \tau - \frac{1}{\tau} \right) \right) x_1x_2^T,
\]

which implies \(\gamma = 0\). So we have

\[
Z = UAU^T = \begin{pmatrix}
  x_1 & \tau x_1 \\
  x_2 & -\frac{1}{\tau}x_2
\end{pmatrix} \begin{pmatrix}
  a_{11} & a_{12} \\
  a_{12} & a_{22}
\end{pmatrix} \begin{pmatrix}
  x_1^T \\
  x_2^T
\end{pmatrix} = \begin{pmatrix}
  x_1^T \\
  x_2^T
\end{pmatrix} \begin{pmatrix}
  a_{11} & a_{12} \\
  a_{12} & a_{22}
\end{pmatrix} \begin{pmatrix}
  x_1 \\
  x_2
\end{pmatrix}.
\]

We now insert the orthogonal matrix \(Q\),

\[
Z = (UQ)(Q^TAQ)(UQ)^T, \quad Q = \begin{pmatrix}
  \|x_1\| & \|x_2\| \\
  \|x_2\| & -\|x_1\|
\end{pmatrix},
\]

and put \(\hat{U} = UQ\) and \(\hat{A} = Q^TAQ\). Straightforward computation gives

\[
\hat{U} = \begin{pmatrix}
  \frac{1}{\|x_1\|} x_1 & 0 \\
  0 & \frac{1}{\|x_2\|} x_2
\end{pmatrix}, \quad \hat{A} = \begin{pmatrix}
  a_{11} & \hat{a}_{12} \\
  \hat{a}_{12} & a_{22}
\end{pmatrix},
\]

where \(\hat{a}_{12} = \|x_1\| \|x_2\| \left( a_{11} - a_{22} + a_{12}\left( -\frac{1}{\tau} + \tau \right) \right) = 0\). This means that \(U\) and \(A\) have the structure (B1).
Assume that $Z_{12} = 0$ and $\text{rank}(Z) = 2$, $x_1 \neq 0$ and $y_1 \neq 0$, and $x_2 = y_2 = 0$. Then it is easy to show that (B2) holds. Analogously one can easily prove the case of (B3).

Finally we must check the following case: assume that $Z_{12} = 0$, $x_1 \neq 0$ and $y_1 \neq 0$, $x_2 = 0$, and $y_2 \neq 0$. From orthogonality of the columns of $U$ we have $x_1^Ty_1 = 0$, and using

$$0 = Z_{12} = (a_{22}y_1 + a_{12}x_1)y_2^T,$$

we get $a_{22} = a_{12} = 0$. Therefore, in this case $Z$ is independent of $y_1$ and $y_2$, and

$$Z = \begin{pmatrix} a_{11}x_1x_1^T & 0 \\ 0 & 0 \end{pmatrix},$$

which means that $\text{rank}(Z) = 1$. The same is true three similar cases.

Note that the case when

$$U = \begin{pmatrix} 0 & u_{12} \\ u_{21} & 0 \end{pmatrix},$$

is equivalent to (B1) due to the property (B4).

In the case (B1) the blocks $Z_{11}$ and $Z_{22}$ both have rank 1. In the cases (B2) and (B3), $Z_{11}$ and $Z_{22}$ have rank 2, respectively.