Multi-subspace power method for decomposing all tensors

Kexin Wang, João M. Pereira, Joe Kileel, and Anna Seigal

Abstract. We present an algorithm for decomposing low rank tensors of any symmetry type, from fully asymmetric to fully symmetric. It generalizes the recent subspace power method from symmetric tensors to all tensors. The algorithm transforms an input tensor into a tensor with orthonormal slices. We show that for tensors with orthonormal slices and low rank, the summands of their decomposition are in one-to-one correspondence with the partially symmetric singular vector tuples (pSVTs) with singular value one. We use this to show correctness of the algorithm. We introduce a shifted power method for computing pSVTs and establish its global convergence. Numerical experiments demonstrate that our decomposition algorithm achieves higher accuracy and faster runtime than existing methods.

1. Introduction

A tensor is a multi-indexed array of numbers. The space of real tensors of format $m_1 \times \cdots \times m_d$ is denoted by $\mathbb{R}^{m_1} \otimes \cdots \otimes \mathbb{R}^{m_d}$. We assume that $m_i \geq 2$ for all $i = 1, \ldots, d$. A tensor is partially symmetric if its entries are unchanged under certain permutations of indices. For example, we consider $\mathscr{T} \in \mathbb{R}^m \otimes \mathbb{R}^m \otimes \mathbb{R}^n$ that satisfy $\mathscr{T}_{ijk} = \mathscr{T}_{jik}$ for all $i, j \in [m]$. More generally, let $\mathfrak{S}_d$ denote the symmetry group on d letters and consider a tensor $\mathscr{T} \in (\mathbb{R}^{m_1})^{\otimes d_1} \otimes \cdots \otimes (\mathbb{R}^{m_\ell})^{\otimes d_\ell}$. We say $\mathscr{T}$ has partial symmetry of type $(d_1, \ldots, d_\ell)$ if

$$\mathscr{T}_{i_1,\ldots,i_d} = \mathscr{T}_{\sigma(i_1),\ldots,\sigma(i_d)}$$

for all $\sigma \in \mathfrak{S}_{d_1} \times \cdots \times \mathfrak{S}_{d_\ell}$; that is, for all permutations that act within each block of indices. We denote the space of tensors with this symmetry type by $S^{d_1}(\mathbb{R}^{m_1}) \otimes \cdots \otimes S^{d_\ell}(\mathbb{R}^{m_\ell})$. We say that a tensor is (fully) symmetric if $\ell = 1$ and (fully) asymmetric if $d_1 = \cdots = d_\ell = 1$.

Partially symmetric tensors arise in a range of settings. In statistics and signal processing, covariance matrices, higher-order moments, and cumulants are symmetric tensors that encode distributional information and are central to PCA, Mixture Models, and ICA [**5, 9, 41, 57**]. When two datasets are analyzed jointly, such as in contrastive PCA and contrastive ICA [**1, 55**], it yields partially symmetric structure. In elastic material analysis, the fourth-order stiffness (or compliance) tensor $\mathscr{C}$ encodes a map from symmetric strain to symmetric stress. Physical and energetic considerations impose the symmetries $\mathscr{C}_{ijkl} = \mathscr{C}_{jikl} = \mathscr{C}_{ijlk} = \mathscr{C}_{klij}$. This tensor encodes the constitutive law of the material, and its M-eigenvectors, critical points of $f(\mathbf{x}, \mathbf{y}) = \mathscr{C}(\mathbf{x}, \mathbf{y}, \mathbf{x}, \mathbf{y})$, identify directional wave speeds and certify strong ellipticity [**21, 35**]. In quantum many-body systems, the Hilbert space (space of wave functions) of N identical bosons on a d-level lies in $S^N(\mathbb{C}^d)$. With two distinguishable bosonic species, the Hilbert space is $S^{N_1}(\mathbb{C}^{d_1}) \otimes S^{N_2}(\mathbb{C}^{d_2})$, symmetric within each

species but not across species [**17, 14**]. In algebraic geometry, the generic rank and identifiability of partially symmetric tensors are active research topics [**2, 3**]. Partially symmetric tensors also appear in computer vision [**38**], computing linear sections of Segre-Veronese secant varieties [**28**] and in training polynomial neural networks [**52**].

A CP decomposition writes a tensor as a sum of outer products of vectors. An order three tensor $\mathscr{T} \in \mathbb{R}^{m_1} \otimes \mathbb{R}^{m_2} \otimes \mathbb{R}^{m_3}$ has a decomposition $\mathscr{T} = \sum_{i=1}^{r} \mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}_i$, where $\mathbf{a}_i \in \mathbb{R}^{m_1}$, $\mathbf{b}_i \in \mathbb{R}^{m_2}$, and $\mathbf{c}_i \in \mathbb{R}^{m_3}$, for sufficiently large r. We normalize the vectors to lie on the unit sphere and include a scalar coefficient λ_i, to write:

$$\mathscr{T} = \sum_{i=1}^{r} \lambda_i \, \mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}_i.$$

The smallest r for which such a decomposition holds is the rank of $\mathscr{T}$. Each summand is a rank-one tensor. CP decompositions recover interpretable structure, for example, they fit observed data to mixture models [**5**] and encode the complexity of linear operators [**34**].

If the tensor has partial symmetry, we seek a decomposition that respects the symmetry. A decomposition where each rank-one term inherits the same symmetry is called a *partially symmetric CP decomposition*. It writes the tensor as a sum of rank-one tensors with partial symmetry of type $(d_1, \ldots, d_\ell)$, i.e. tensors $\lambda(\mathbf{v}^{(1)})^{\otimes d_1} \otimes (\mathbf{v}^{(2)})^{\otimes d_2} \otimes \cdots \otimes (\mathbf{v}^{(\ell)})^{\otimes d_\ell}$. By enforcing partial symmetry, we use fewer parameters and obtain decompositions that are easier to interpret in the model. One can ignore the symmetry and apply a general CP decomposition, but the partial symmetry in the factors may be lost. Our numerical experiments show that respecting the partial symmetry leads to computational speed-ups.

We introduce an algorithm (Algorithm 4) for the decomposition of tensors with any partial symmetry. We call it the *Multi-Subspace Power Method (MSPM)*. It finds rank-one tensors in a subspace and uses additional subspaces to extend these partial factors into full summands of the decomposition. The implementation is available at `https://github.com/joaompereira/SPM`. It finds one component at a time via the partially symmetric higher-order power method (PS-HOPM). The algorithm generalizes the subspace power method (SPM) [**29**] from symmetric tensors to all tensors. It is competitive for usual asymmetric CP decompositions. The algorithm has four parts:

(1) Form a tensor with last slices orthonormal: Given $\mathscr{T} \in S^{d_1}(\mathbb{R}^{m_1}) \otimes \cdots \otimes S^{d_\ell}(\mathbb{R}^{m_\ell})$, compute its flattening with rows of symmetry type $(f_1, \ldots, f_\ell)$, where $0 \leq f_i \leq d_i$. Compute an orthonormal basis for its column space (e.g. via SVD). Stack the basis vectors to obtain a tensor $\mathscr{T}^{(f)}$ whose last slices are orthonormal. See Figure 1.
(2) Rank-one recovery: Apply PS-HOPM (Section 5) to $\mathscr{T}^{(f)}$ to compute partially symmetric singular vector tuples (pSVTs) with singular value one. Each yields a partially symmetric rank-one component $\bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes f_i}$.
(3) Completion and deflation: Construct the complementary tensor $\bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes (d_i - f_i)}$ (extra flattenings are needed when some $f_i = 0$), estimate its scalar coefficient λ, and subtract $\lambda \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes d_i}$ from the current residual tensor.
(4) Iteration until full decomposition: Repeat the rank-one recovery step and the completion and deflation step until all r components are recovered.

One can work with noisy tensors by truncating the SVD in the first step of the algorithm.

We show that tensors with orthonormal slices have decompositions in terms of their partially symmetric singular vector tuples, provided the rank is small enough. This proves the correctness of our algorithm and builds on work linking critical points of the best rank-one

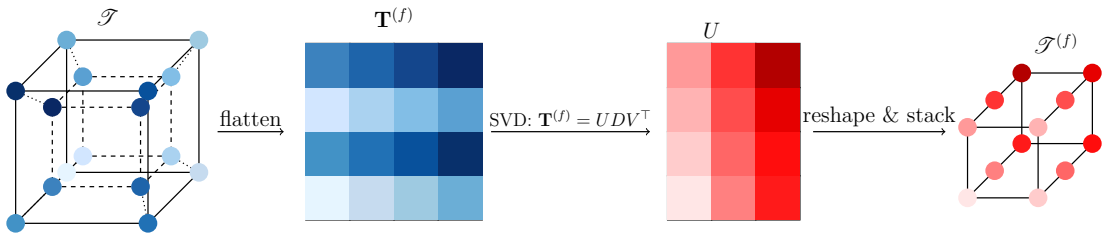


FIGURE 1. Building a tensor with last slices orthonormal from a general tensor. A fourth-order tensor $\mathscr{T}$ is represented as a 4D hypercube. Flattening with $f = (1, 1, 0, 0)$ combines the first two directions to be columns in the flattening matrix $\mathbf{T}^{(f)}$. The SVD yields the left singular vectors U. The columns of U are reshaped into tensors and stacked to form the tensor $\mathscr{T}^{(f)}$.

approximation of a tensor to the rank-one summands of a decomposition [**15, 46, 53**]. An alternative to PSPM is to compute the pSVT corresponding to the best rank-one approximation of the original tensor, subtract it off, and repeat. However, this fails in general: the best rank-one approximation need not be a component of the decomposition and subtracting it off can even increase the tensor rank [**13, 27, 46, 50**].

To compute partially symmetric singular vector tuples, we introduce the partially symmetric Higher-Order Power Method (PS-HOPM), which generalizes both the Higher-Order Power Method [**11**] and the shifted power method for symmetric tensors [**32**], and we prove global convergence and local linear convergence of this algorithm in a special case. We demonstrate through numerical experiments on noisy tensors that our algorithm achieves higher accuracy and faster run times than existing approaches.

In Section 2 we give background on partially symmetric tensors, their flattenings, and singular vector tuples. Section 3 presents our first main result, showing how tensors with orthonormal slices can be decomposed using their singular vector tuples. Section 4 discusses the choice of flattening and provides rank bounds that guarantee uniqueness. Section 5 describes the *Partially Symmetric Higher-Order Power Method* and establishes its global convergence and local linear convergence. Section 6 develops the deflation and factor completion procedures needed for full decompositions. Section 7 reports numerical experiments that compare our method with existing approaches. Section 8 concludes the paper.

2. Background

A partially symmetric tensor in $S^{d_1}(\mathbb{R}^{m_1}) \otimes \cdots \otimes S^{d_\ell}(\mathbb{R}^{m_\ell})$ has a decomposition

$$\mathscr{T} = \sum_{i=1}^{r} \lambda_i (\mathbf{v}_i^{(1)})^{\otimes d_1} \otimes (\mathbf{v}_i^{(2)})^{\otimes d_2} \otimes \cdots \otimes (\mathbf{v}_i^{(\ell)})^{\otimes d_\ell},$$

where $\mathbf{v}_i^{(j)}$ lies in the unit sphere $\mathbb{S}^{m_j - 1}$ for $j = 1, \ldots, \ell$ and r is sufficiently large. Let $\|\cdot\|$ denote the Frobenius norm (the square root of the sum of the squares of the entries). In practice, computing a partially symmetric decomposition amounts to solving

$$\min_{\{\mathbf{v}_i^{(j)}, \lambda_i\}} \left\| \mathscr{T} - \sum_{i=1}^{r} \lambda_i \left(\mathbf{v}_i^{(1)}\right)^{\otimes d_1} \otimes \left(\mathbf{v}_i^{(2)}\right)^{\otimes d_2} \otimes \cdots \otimes \left(\mathbf{v}_i^{(\ell)}\right)^{\otimes d_\ell} \right\|.$$

Tensor decomposition is challenging, both theoretically and practically. It is in general NP hard [**24, 25**]. There are numerical challenges that arise from the fact that the set of tensors of rank at most r is not closed [**13**] and the optimization landscape is non-convex [**20**].

DEFINITION 2.1 (Contraction). Let $\mathscr{T}$ be a tensor in $S^{d_1}(\mathbb{R}^{m_1}) \otimes \cdots \otimes S^{d_\ell}(\mathbb{R}^{m_\ell})$. For vectors $\mathbf{v}^{(i)} \in \mathbb{R}^{m_i}$ and non-negative integers $f_1 \leq d_1, \ldots, f_\ell \leq d_\ell$, the contraction of $\mathscr{T}$ with $\bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes f_i}$ produces a tensor in $S^{d_1-f_1}(\mathbb{R}^{m_1}) \otimes \cdots \otimes S^{d_\ell-f_\ell}(\mathbb{R}^{m_\ell})$. We denote it by

$$\mathscr{T} \cdot \bigotimes_{i=1}^{\ell} \left(\mathbf{v}^{(i)}\right)^{\otimes f_i}.$$

Its entries are

$$\left(\mathscr{T} \cdot \bigotimes_{i=1}^{\ell} \left(\mathbf{v}^{(i)}\right)^{\otimes f_i}\right)_{\mathbf{k}_1 \cdots \mathbf{k}_\ell} = \sum_{\mathbf{j}_1 \in [m_1]^{f_1}} \cdots \sum_{\mathbf{j}_\ell \in [m_\ell]^{f_\ell}} \mathscr{T}_{(\mathbf{j}_1,\mathbf{k}_1)\cdots(\mathbf{j}_\ell,\mathbf{k}_\ell)} \prod_{i=1}^{\ell} (\mathbf{v}^{(i)})_{\mathbf{j}_i}^{\otimes f_i},$$

where $[m] = \{1, \ldots, m\}$ and $\mathbf{k}_i \in [m_i]^{d_i - f_i}$.

Let $\delta_{i,j}$ denote the Kronecker delta, which equal to 1 if $i = j$ and 0 otherwise. In the special case $f_i = d_i - \delta_{i,j}$, the contraction produces a vector in $\mathbb{R}^{m_j}$.

DEFINITION 2.2 (Partially symmetric singular vector tuple (pSVT), see [**18**]). Fix a tensor $\mathscr{T} \in \bigotimes_{i=1}^{\ell} S^{d_i}(\mathbb{R}^{m_i})$. A *partially symmetric singular vector tuple* is a tuple of vectors

$$(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(\ell)}) \in \mathbb{S}^{m_1-1} \times \cdots \times \mathbb{S}^{m_\ell-1}$$

such that, for $j = 1, \ldots, \ell$, the following equation holds:

$$\mathscr{T} \cdot \bigotimes_{i=1}^{\ell} \left(\mathbf{v}^{(i)}\right)^{\otimes d_i - \delta_{ij}} = \sigma \mathbf{v}^{(j)}, \tag{1}$$

The scalar $\sigma \in \mathbb{R}$ is called the *singular value*, which can be computed explicitly as the full contraction $\sigma = \mathscr{T} \cdot \bigotimes_{i=1}^{\ell} \left(\mathbf{v}^{(i)}\right)^{\otimes d_i}$.

Throughout the paper, when we talk about pSVTs, we mean real vector tuples. PSVTs specialize to usual singular vector tuples (SVTs) [**36**] for when $d_i = 1$ for all i, and to tensor eigenvectors when $\ell = 1$. The pSVT pairs $(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(\ell)}, \sigma)$ of $\mathscr{T}$ are the critical points of the squared distance

$$\left\| \mathscr{T} - \sigma \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes d_i} \right\|^2,$$

see [**36**]. For matrices, singular values can always be assumed to be non-negative by flipping the sign of a singular vector. For partially symmetric tensors with some d_i odd, the same approach works to assume $\sigma \geq 0$. When all exponents d_i are even, the singular value is invariant under sign flips of the vectors (just like for eigenvectors of matrices).

PROPOSITION 2.3 (See [**18**, Theorem 12]). *A generic partially symmetric tensor $\mathscr{T}$ has finitely many pSVTs.*

A tensor can be flattened to a matrix in various ways.

DEFINITION 2.4 (Flattenings of a partially symmetric tensor). Fix $\mathscr{T} \in \bigotimes_{i=1}^{\ell} S^{d_i}(\mathbb{R}^{m_i})$, and a tuple $f = (f_1, \ldots, f_\ell)$ with $0 \le f_i \le d_i$. The f-flattening of $\mathscr{T}$, denoted $\mathbf{T}^{(f)}$, is a matrix with rows and columns indexed by entries of partially symmetric tensors in

$$\bigotimes_{i=1}^{\ell} S^{f_i}(\mathbb{R}^{m_i}) \quad \text{and} \quad \bigotimes_{i=1}^{\ell} S^{d_i - f_i}(\mathbb{R}^{m_i}),$$

respectively. When the tuple f is $(d_1, \ldots, d_\ell)$, the flattening $\mathbf{T}^{(f)}$ is a vector, denoted $\operatorname{vect}(\mathscr{T})$ and called the vectorization of $\mathscr{T}$.

For example, when $\mathscr{T} = (\mathbf{v}^{(1)})^{\otimes d_1} \otimes \cdots \otimes (\mathbf{v}^{(\ell)})^{\otimes d_\ell}$,

$$\mathbf{T}^{(f)} = \operatorname{vect}\left(\left(\mathbf{v}^{(1)}\right)^{\otimes f_1} \otimes \cdots \otimes \left(\mathbf{v}^{(\ell)}\right)^{\otimes f_\ell}\right) \otimes \operatorname{vect}\left(\left(\mathbf{v}^{(1)}\right)^{\otimes (d_1 - f_1)} \otimes \cdots \otimes \left(\mathbf{v}^{(\ell)}\right)^{\otimes (d_\ell - f_\ell)}\right).$$

3. A decomposition from singular vector tuples

In this section we show how tensors with orthonormal slices can be decomposed using their pSVTs, provided the rank is not too large.

DEFINITION 3.1 (Orthonormal slices). Given $\mathscr{T} \in \bigotimes_{k=1}^{\ell} \mathbb{R}^{m_k}$, its *$j$-slices* are a collection of m_j order $\ell - 1$ tensors in $\mathbb{R}^{m_1} \otimes \cdots \otimes \mathbb{R}^{m_{j-1}} \otimes \mathbb{R}^{m_{j+1}} \otimes \cdots \otimes \mathbb{R}^{m_\ell}$ obtained by fixing the value of the j-th index of $\mathscr{T}$. We call the ℓ-slices the *last slices*. We say the j-slices of $\mathscr{T}$ are *orthonormal* if their vectorizations are orthonormal vectors.

We relate CP decomposition to pSVTs, for tensors with orthonormal slices. We see that pSVTs with singular value appear in the summands of the CP decomposition.

EXAMPLE 3.2 (A $2 \times 2 \times 2$ tensor). Consider the tensor

$$\mathscr{T} = \mathbf{u} \otimes \mathbf{u} \otimes \mathbf{v} \; + \; \mathbf{v} \otimes \mathbf{v} \otimes \mathbf{u}, \quad \text{where} \quad \mathbf{u} := \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \quad \mathbf{v} := \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

The tensor $\mathscr{T}$ is odeco [**47, 48**]. The third slices of $\mathscr{T}$ are orthonormal. We compute the pSVTs of $\mathscr{T}'$ by solving polynomial equations (1) in `HomotopyContinuation.jl` [**6**]. The polynomial system has six complex solutions, of which four are the real pSVTs. Among them, the two tuples $(\mathbf{u}, \mathbf{u}, \mathbf{v})$ and $(\mathbf{v}, \mathbf{v}, \mathbf{u})$ have singular value one. They are the two rank-one summands of $\mathscr{T}$.

EXAMPLE 3.3. Let $\mathbf{u}$ be as in the previous example and consider

$$\mathscr{T}' = \mathbf{u} \otimes \mathbf{u} \otimes \begin{pmatrix} \frac{1}{\sqrt{3}} \\ 1 \end{pmatrix} + \mathbf{e}_1 \otimes \mathbf{e}_1 \otimes \begin{pmatrix} -\frac{2}{\sqrt{3}} \\ 0 \end{pmatrix}, \quad \text{where} \quad \mathbf{e}_1 := \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The decomposition is unique, by Kruskal's criterion [**33**]. The tensor $\mathscr{T}'$ is not odeco since $\mathbf{u}$ is not orthogonal to $\mathbf{e}_1$, but its last slices

$$\begin{pmatrix} \frac{-\sqrt{3}}{2} & \frac{1}{2\sqrt{3}} \\ \frac{1}{2\sqrt{3}} & \frac{1}{2\sqrt{3}} \end{pmatrix}, \quad \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

are orthonormal. We compute the pSVTs of $\mathscr{T}'$ by solving polynomial equations (1) in `HomotopyContinuation.jl` [**6**]. This system has six complex solutions and all of them are real. Two have singular value one:

$$(\mathbf{e}_1, \mathbf{e}_1, \begin{pmatrix} -\frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{pmatrix}), \quad (\mathbf{u}, \mathbf{u}, \mathbf{e}_2).$$

The first two factors of each pSVT, $\mathbf{e}_1 \otimes \mathbf{e}_1$ and $\mathbf{u} \otimes \mathbf{u}$, appear in the decomposition of $\mathscr{T}'$. The third vectors in the pSVTs do not match the decomposition factors.

We next prove a one-to-one correspondence observed in Examples 3.2 and 3.3 between pSVTs with singular value one and the rank-one summands in the decomposition, for a tensor with orthonormal slices. PSVTs with singular value one exist for tensors with orthonormal slices and sufficiently low rank. The question of how high the rank can be will be addressed in the next section. First we recall a related result for matrices.

LEMMA 3.4. *Fix $m \geq n$. Let $\mathbf{M} \in \mathbb{R}^{m \times n}$ be a matrix with orthonormal columns. The singular values of $\mathbf{M}$ are 0 or 1. A vector $\mathbf{u} \in \mathbb{S}^{m-1}$ is a left singular vector of $\mathbf{M}$ with singular value one if and only if it lies in the column span of $\mathbf{M}$.*

PROOF. We have $\mathbf{M}^\mathsf{T}\mathbf{M} = \mathbf{I}_n$, since $\mathbf{M}$ has orthonormal columns. Hence, the eigenvalues of $\mathbf{M}^\mathsf{T}\mathbf{M}$ are all one, so the nonzero singular values of $\mathbf{M}$ are all one. Suppose $(\mathbf{u}, \mathbf{v})$ is a singular vector pair of $\mathbf{M}$ with singular value one. It follows that $\mathbf{u} = \mathbf{M}\mathbf{v}$ so $\mathbf{u} \in \operatorname{colspan} \mathbf{M}$. Conversely, let $\mathbf{u} \in \operatorname{colspan}(\mathbf{M})$ be a unit vector. Define $\mathbf{v} = \mathbf{M}^\mathsf{T}\mathbf{u}$. Using that $\mathbf{M}\mathbf{M}^\mathsf{T}$ is orthogonal projection onto $\operatorname{colspan} \mathbf{M}$, we have $\mathbf{M}\mathbf{v} = \mathbf{M}\mathbf{M}^\mathsf{T}\mathbf{u} = \mathbf{u}$ and $\|\mathbf{v}\|^2 = \mathbf{u}^\mathsf{T}\mathbf{M}\mathbf{M}^\mathsf{T}\mathbf{u} = \|\mathbf{u}\|^2 = 1$. So $(\mathbf{u}, \mathbf{v})$ is a singular vector pair with singular value one. □

THEOREM 3.5. *Fix $\mathscr{T} \in \left(\bigotimes_{i=1}^{\ell} S^{d_i}(\mathbb{R}^{m_i})\right) \otimes \mathbb{R}^m$. Let $\mathcal{A} \subseteq \bigotimes_{i=1}^{\ell} S^{d_i}(\mathbb{R}^{m_i})$ be the span of the last slices of $\mathscr{T}$, which we assume to be orthonormal. Then, the pSVTs of $\mathscr{T}$ have singular values in the interval $[0, 1]$, and the rank-one tensors in $\bigotimes_{i=1}^{\ell} S^{d_i}(\mathbb{R}^{m_i})$ corresponding to pSVTs with singular value one are the rank-one tensors in $\mathcal{A}$.*

Suppose further that $\mathscr{T}$ has rank $r := \dim \mathcal{A}$ and decomposition

$$\mathscr{T} = \sum_{i=1}^{r} \lambda_i \left(\mathbf{v}_i^{(1)}\right)^{\otimes d_1} \otimes \cdots \otimes \left(\mathbf{v}_i^{(\ell)}\right)^{\otimes d_\ell} \otimes \mathbf{u}_i. \tag{2}$$

Then for each summand in the decomposition of $\mathscr{T}$, there exists $\mathbf{w}_i \in \mathbb{S}^{m-1}$ such that the tuple $(\mathbf{v}_i^{(1)}, \ldots, \mathbf{v}_i^{(\ell)}, \mathbf{w}_i)$ is a pSVT of $\mathscr{T}$ with singular value one. A linear transformation computed from $\mathscr{T}$ maps each $\mathbf{w}_i$ to $\mathbf{u}_i$.

PROOF. Let $\mathbf{M}$ denote the flattening of $\mathscr{T}$ using the tuple $(d_1, \ldots, d_\ell, 0)$. It has orthonormal columns, since the last slices of $\mathscr{T}$ are orthonormal. Their span is $\mathcal{A}$. Let $(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(\ell)}, \mathbf{w})$ be a pSVT of $\mathscr{T}$ with singular value σ. After changing the sign of $\mathbf{w}$ if necessary, we can assume σ is non-negative.

The maximum singular value of $\mathscr{T}$ is $\max_{\mathbf{v}^{(i)} \in \mathbb{S}^{m_i - 1}, \mathbf{u} \in \mathbb{S}^{m-1}} \mathscr{T} \cdot \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes d_i} \otimes \mathbf{u}$. Define $\mathscr{R}_s = \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes d_i}$. Then $\operatorname{vect}(\mathscr{R}_s) \in \mathbb{S}^{M-1}$, where $M = \prod_{i=1}^{\ell} m_i^{d_i}$, and

$$\mathscr{T} \cdot \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes d_i} \otimes \mathbf{u} = \operatorname{vect}(\mathscr{R}_s)^\mathsf{T} \mathbf{M} \mathbf{u}.$$

The matrix $\mathbf{M}$ has maximum singular value one, by Lemma 3.4. Thus, the singular values of $\mathscr{T}$ lie in $[0, 1]$.

Suppose $(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(\ell)}, \mathbf{w})$ is a pSVT with singular value one. Then $(\operatorname{vect}(\mathscr{R}_s), \mathbf{w})$ is a singular vector pair of $\mathbf{M}$ with singular value one. We obtain $\operatorname{vect}(\mathscr{R}_s) = \mathbf{M}\mathbf{w} \in \operatorname{colspan}(\mathbf{M})$, by Lemma 3.4. Hence $\mathscr{R}_s \in \mathcal{A}$. For the converse, suppose $\mathscr{R}_s \in \mathcal{A}$. Then $(\operatorname{vect}(\mathscr{R}_s), \mathbf{w})$ is a singular vector pair of $\mathbf{M}$ with singular value one, again by Lemma 3.4. This implies

$$\mathscr{T}(*, \ldots, *, \mathbf{w}) = \mathscr{R}_s, \quad \text{and} \quad \mathscr{T} \cdot \mathscr{R}_s = \mathbf{w}.$$

Therefore, $(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)}, \mathbf{w})$ is a pSVT with singular value one, since $(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)})$ is a pSVT of $\mathscr{R}_s$ and $\mathbf{w} = \mathscr{T} \cdot \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes d_i}$.

Suppose $\operatorname{rank}(\mathscr{T}) = \dim(\mathcal{A})$ and that (2) holds. Let $\mathbf{U} = \begin{pmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_r \end{pmatrix}$, $\mathbf{W} = \begin{pmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_r \end{pmatrix}$, $\mathbf{M}_{\mathscr{R}} = \begin{pmatrix} \mathscr{R}^{(1)} \\ \vdots \\ \mathscr{R}^{(r)} \end{pmatrix}$ (where $\mathscr{R}^{(i)} = \bigotimes_{j=1}^{\ell} (\mathbf{v}_i^{(j)})^{\otimes d_j}$), and let $\Lambda = \operatorname{Diag}(\lambda_1, \dots, \lambda_r)$. We have $\mathscr{T} = \sum_{i=1}^{r} \lambda_i \mathscr{R}^{(i)} \otimes \mathbf{u}_i$, and it follows that

$$\mathbf{M} = \sum_{i=1}^{r} \lambda_i \operatorname{vect}(\mathscr{R}^{(i)}) \otimes \mathbf{u}_i = \mathbf{M}_{\mathscr{R}}^{\mathsf{T}} \mathbf{\Lambda}\, \mathbf{U}.$$

The linear space $\mathcal{A}$ is spanned by $\operatorname{vect}(\mathscr{R}^{(1)}), \dots, \operatorname{vect}(\mathscr{R}^{(r)})$, since $r = \dim \mathcal{A}$. Hence, $\mathscr{R}^{(i)} \in \mathcal{A}$ and $(\mathbf{v}_i^{(1)}, \dots, \mathbf{v}_i^{(\ell)}, \mathbf{w}_i)$ is a pSVT with singular value one, where $\mathbf{w}_i = \mathscr{T} \cdot \mathscr{R}^{(i)}$. We now relate $\mathbf{U}$ and $\mathbf{W}$. By construction, we have $\mathscr{T} \cdot \mathscr{R}^{(i)} = \mathbf{w}_i$, so $\mathbf{M}_{\mathscr{R}}^{\mathsf{T}} \mathbf{M} = \mathbf{W}$. Combining this with $\mathbf{M} = \mathbf{M}_{\mathscr{R}} \mathbf{\Lambda} \mathbf{U}$ yields

$$\mathbf{W} = \mathbf{M}_{\mathscr{R}}^{\mathsf{T}} \mathbf{M}_{\mathscr{R}}\, \mathbf{\Lambda}\, \mathbf{U}, \quad \text{so} \quad \mathbf{U} = \mathbf{\Lambda}^{-1} \big(\mathbf{M}_{\mathscr{R}}^{\mathsf{T}} \mathbf{M}_{\mathscr{R}} \big)^{-1} \mathbf{W}. \qquad \square$$

REMARK 3.6. *With the same argument, one can show a stronger statement: any SVT of $\mathscr{T}$ with singular value one is a pSVT of $\mathscr{T}$. Moreover, when the tensors $\{\mathscr{R}^{(i)}\}_{i=1}^{n}$ are orthogonal, the decomposition of $\mathscr{T}$ consists of rank-one components that are pSVTs, since $\mathbf{M}_{\mathscr{R}}^{\mathsf{T}} \mathbf{M}_{\mathscr{R}} = \mathbf{I}$ and $\mathbf{U} = \mathbf{W}$. The vectors $\{\mathbf{w}_1, \dots, \mathbf{w}_r\}$ in the second part of Theorem 3.5 are also orthogonal since $\mathbf{M}_{\mathscr{R}} \mathbf{M} = \mathbf{W}$.*

4. The choice of flattening

In Section 3, we studied tensors with orthonormal slices. We established a correspondence between the summands in their decomposition and the pSVTs with singular value one. For a partially symmetric tensor $\mathscr{T} \in \bigotimes_{i=1}^{\ell} S^{d_i}(\mathbb{R}^{m_i})$, we build a tensor with orthonormal slices, as follows (see Figure 1).

(1) Given any $f = (f_1, \dots, f_\ell)$ with $0 \le f_i \le d_i$ for all i, build the flattening $\mathbf{T}^{(f)}$ of $\mathscr{T}$.
(2) Compute an orthonormal basis for its column space.
(3) Reshape each basis vector into a partially symmetric tensor in $\bigotimes_{i=1}^{\ell} S^{f_i}(\mathbb{R}^{m_i})$
(4) Stack these to form the last slices of tensor $\mathscr{T}^{(f)}$ with symmetry $(f_1, \dots, f_\ell, 1)$.

If the tensor $\mathscr{T}$ has rank r, then provided that the new smaller tensor $\mathscr{T}^{(f)}$ has a unique rank r decomposition, we can use the decomposition of $\mathscr{T}^{(f)}$ to build part of the decomposition of $\mathscr{T}$. Moreover, if the rank of $\mathscr{T}$ is $r = \operatorname{rank} \mathbf{T}^{(f)}$, we can use the pSVTs of $\mathscr{T}^{(f)}$ with singular value one to decompose $\mathscr{T}^{(f)}$ by Theorem 3.5. The condition that $\mathscr{T}^{(f)}$ has unique rank $r = \operatorname{rank} \mathbf{T}^{(f)}$ decomposition holds for generic rank-one summands of $\mathscr{T}$ if r is smaller than a quantity related to f. We now define this quantity.

DEFINITION 4.1 (Maximal rank $r(f)$). Let $\mathscr{T} \in \bigotimes_{i=1}^{\ell} S^{d_i}(\mathbb{R}^{m_i})$ be a partially symmetric tensor of rank r with generic rank-one summands. Let $\mathscr{T}^{(f)}$ be the tensor with orthonormal slices obtained from a flattening f of $\mathscr{T}$, as described above. We define $r(f)$ to be the largest integer r such that $r = \operatorname{rank} \mathbf{T}^{(f)}$ and $\mathscr{T}^{(f)}$ has a unique decomposition of rank r.

The notion $r(f)$ is well-defined, as the uniqueness of the decomposition of $\mathscr{T}^{(f)}$ does not depend on the orthonormal basis. Suppose $\mathscr{T} \in \bigotimes_{i=1}^{\ell} S^{d_i}(\mathbb{R}^{m_i})$. The rank $r(f)$ cannot exceed

the number of rows and columns of the flattening matrix $\mathbf{T}^{(f)}$, which we denote by

$$n^f_{\text{row}} := \prod_{i=1}^{\ell} \binom{m_i + f_i - 1}{f_i} = \dim\left(\bigotimes_{i=1}^{\ell} S^{f_i}(\mathbb{R}^{m_i})\right), \qquad n^f_{\text{col}} := \prod_{i=1}^{\ell} \binom{m_i + d_i - f_i - 1}{d_i - f_i}. \tag{3}$$

Here n^f_{row} and n^f_{col} count the number of distinct rows and columns, after removing repetitions due to partial symmetry.

Each column of $\mathbf{T}^{(f)}$ is (the vectorization of) a partially symmetric tensor in $\bigotimes_{i=1}^{\ell} S^{f_i}(\mathbb{R}^{m_i})$. The set of rank-one tensors (up to scale) in this space is the Segre–Veronese variety of $\mathbb{P}^{m_1-1} \times \cdots \times \mathbb{P}^{m_\ell - 1}$ embedded via multi-degree $\mathcal{O}(f_1, \ldots, f_\ell)$, which we denote by $X \subseteq \mathbb{P}^{N-1}$, where $N = \prod_{i=1}^{\ell} \binom{m_i+f_i-1}{f_i} = n^f_{\text{row}}$. The dimension of X is $\sum_{f_i \neq 0}(m_i - 1)$ and its codimension is

$$n^f_{\text{codim}} := \prod_{i=1}^{\ell} \binom{m_i + f_i - 1}{f_i} - 1 - \sum_{f_i \neq 0} (m_i - 1). \tag{4}$$

Since $X \subset \mathbb{P}^{n^f_{\text{row}}-1}$ and $m_i \geq 2$, we have $n^f_{\text{codim}} < n^f_{\text{row}}$. The roles of the rows and columns in $\mathbf{T}^{(f)}$ are not on equal footing: we seek rank-one tensors in the column span.

We will see that $\mathscr{T}^{(f)}$ has a unique r decomposition when $\operatorname{colspan} \mathbf{T}^{(f)}$ contains exactly r rank-one tensors (up to scale) with the prescribed partial symmetry. This holds whenever $r \leq n^f_{\text{codim}}$. It may hold with positive Lebesgue measure when $r = n^f_{\text{codim}} + 1$, see [**43**]. We formalize the above discussion of $r(f)$ in the following theorem.

THEOREM 4.2. *Let $\mathscr{T} \in \bigotimes_{i=1}^{\ell} S^{d_i}(\mathbb{R}^{m_i})$ be a generic rank r tensor, with flattening $\mathbf{T}^{(f)}$ from tuple $f = (f_1, \ldots, f_\ell)$. Form the tensor $\mathscr{T}^{(f)}$ by stacking an orthonormal basis of $\mathbf{T}^{(f)}$, and define $n^f_{\text{row}}, n^f_{\text{col}}, n^f_{\text{codim}}$ as in* (3) *and* (4)*. Then the largest r for which $\mathscr{T}^{(f)}$ admits a unique rank $r = \operatorname{rank} \mathbf{T}^{(f)} = \operatorname{rank} \mathscr{T}$ decomposition of $\mathscr{T}$ is*

$$r(f) = \begin{cases} \min\{\, n^f_{\text{col}},\, n^f_{\text{codim}} + 1 \,\}, & \text{if } f \text{ has two entries } i, j \text{ equal to } 1 \text{ (all others } 0) \\ & \quad \text{with } \min\{m_i, m_j\} = 2, \\ \min\{\, n^f_{\text{col}},\, n^f_{\text{codim}} \,\}, & \text{otherwise.} \end{cases}$$

PROOF. Consider a rank r tensor $\mathscr{T} \in \bigotimes_{i=1}^{\ell} S^{d_i}(\mathbb{R}^{m_i})$ with decomposition

$$\mathscr{T} = \sum_{i=1}^{r} \lambda_i \bigotimes_{j=1}^{\ell} \big(\mathbf{v}_i^{(j)}\big)^{\otimes d_j}, \qquad \lambda_i \in \mathbb{R} \smallsetminus \{0\},\ \mathbf{v}_i^{(j)} \in \mathbb{S}^{m_j - 1},$$

where the coefficients λ_i and the factors $\{\mathbf{v}_i^{(j)}\}$ are generic. For a fixed tuple $f = (f_1, \ldots, f_\ell)$, the flattening $\mathbf{T}^{(f)}$ is

$$\mathbf{T}^{(f)} = \sum_{i=1}^{r} \lambda_i \underbrace{\operatorname{vect}\Big(\bigotimes_{j=1}^{\ell} (\mathbf{v}_i^{(j)})^{\otimes f_j}\Big)}_{=\mathbf{a}_i} \otimes \underbrace{\operatorname{vect}\Big(\bigotimes_{j=1}^{\ell} (\mathbf{v}_i^{(j)})^{\otimes (d_j - f_j)}\Big)}_{=\mathbf{b}_i}.$$

If $r \leq \min\{n^f_{\text{row}}, n^f_{\text{col}}\}$, then the vectors $\{\mathbf{a}_i\}_{i=1}^r$ and $\{\mathbf{b}_i\}_{i=1}^r$ are linearly independent, since this is a Zariski open condition on the factors $\mathbf{v}_i^{(j)}$ and it is nonempty because the Segre-Veronese varieties are nondegenerate. Hence $\operatorname{rank}(\mathbf{T}^{(f)}) = r$ and

$$\operatorname{colspan}(\mathbf{T}^{(f)}) = \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_r\} = \operatorname{span}\Big\{\operatorname{vect}\Big(\bigotimes_{j=1}^{\ell} (\mathbf{v}_i^{(j)})^{\otimes f_j}\Big) : i = 1, \ldots, r\Big\}.$$

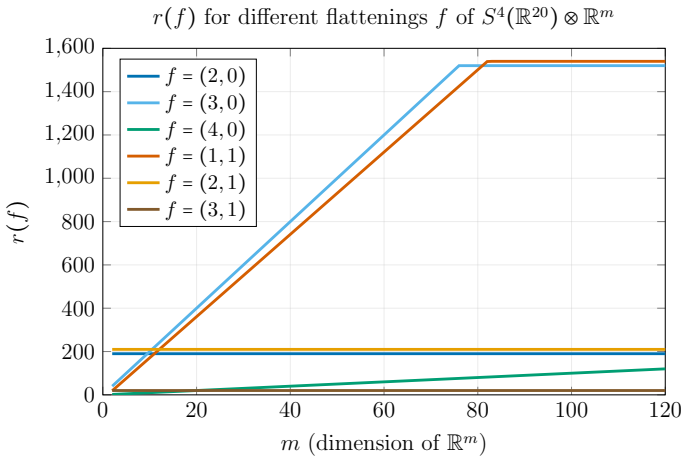
If $r \le r(f)$, we show that the tensor $\mathscr{T}^{(f)}$ has a unique rank r decomposition, since $\bigotimes_{j=1}^{\ell}(\mathbf{v}_i^{(j)})^{\otimes f_j}$ for $i = 1, \dots, r$ are the only rank-one tensors in colspan $\mathbf{T}^{(f)}$ up to scale. Let $X \subseteq \mathbb{P}^{N-1}$ denote the the Segre–Veronese variety parametrizing rank-one tensors in $\bigotimes_{j=1}^{\ell} S^{f_j}(\mathbb{R}^{m_j})$ up to scale. If

$$\dim \operatorname{colspan}(\mathbf{T}^{(f)}) + \dim X < N - 1 = \prod_{i=1}^{\ell} \binom{m_i + f_i - 1}{f_i} - 1,$$

which is equivalent to $r \le n^f_{\text{codim}}$, then the intersection $\operatorname{colspan}(\mathbf{T}^{(f)}) \cap X$ consists of the r points $\bigotimes_{j=1}^{\ell}(\mathbf{v}_i^{(j)})^{\otimes f_j}$ for $i = 1, \dots, r$, by the Generalized Trisecant Lemma [**8**, Proposition 2.6], When f has two entries i, j equal to one, with $\min\{m_i, m_j\} = 2$, and all other entries zero, we have $X \cong \mathbb{P}^{m_i - 1} \times \mathbb{P}^{m_j - 1}$ with degree $\deg X = \max\{m_i, m_j\}$. When $r = n^f_{\text{codim}} + 1 = \max\{m_i, m_j\}$, the linear space $\operatorname{colspan}(\mathbf{T}^{(f)})$ has complementary dimension to X, so by Bézout's theorem [**22**, Theorem 18.3] the intersection consists of exactly $\deg X = r$ points. For all other tuples f, we have $\deg X > r$ and there exist linear space of projective dimension $r - 1$ that intersects X in more than r real points, see [**43**, Theorem 1.6].

If $r > r(f)$, there is an Euclidean open set of linear spaces $\operatorname{colspan}(\mathbf{T}^{(f)})$ that intersect X in more than r real rank-one tensors [**43**, Theorem 1.6]. We can choose a different basis of $\operatorname{colspan}(\mathbf{T}^{(f)})$ consisting of rank-one tensors. This yields another rank r decomposition of $\mathscr{T}^{(f)}$, contradicting uniqueness. □

The above theorem gives a formula for $r(f)$. The optimal flattening depends nontrivially on the size of the tensor. Figure 2 illustrates the behaviour for $S^4(\mathbb{R}^{20}) \otimes \mathbb{R}^m$ as m varies from 2 to 120: we compare $r(f)$ across all possible flattenings.



,

FIGURE 2. The maximal rank $r(f)$ for $S^4(\mathbb{R}^{20}) \otimes \mathbb{R}^m$ with $m \in [2, 120]$ for all flattenings f. The quantity $r(f)$ is piecewise linear in m in this case.

5. A partially symmetric power method

In this section we consider a power method for computing pSVTs of partially symmetric tensors, which unifies two classical algorithms: the Higher-Order Power Method (HOPM) [**11**] and the shifted power method for symmetric tensors [**32**]. We call it the

partially symmetric higher order power method (PS-HOPM). PS-HOPM is described in Algorithm 1, with two implementations of power method iteration in Algorithms 2a and 2b. It is applicable to general partially symmetric tensors, and may be of independent interest. For partially symmetric CP decompositions, we will apply PS-HOPM to tensors with orthonormal last slices $\mathscr{T}^{(f)}$ and their symmetrized analogs $\widetilde{\mathscr{T}^{(f)}}$, see Definition 5.3.

Algorithm 1 Partially Symmetric Higher Order Power Method (PS-HOPM)

Input: Partially symmetric tensor $\mathscr{T} \in \bigotimes_{i=1}^{\ell} S^{a_i}(\mathbb{R}^{m_i})$
Input: Shift parameters $\gamma_1, \dots, \gamma_\ell \geq 0$, stopping criterion
Output: A pSVT $(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)})$ with singular value σ
  for $i = 1, \dots, \ell$ **do**
    $\mathbf{v}^{(i)} \leftarrow$ random unit vector in $\mathbb{S}^{m_i - 1}$

  repeat
    $(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)}) \leftarrow \mathrm{PMIe}(\mathscr{T}; \mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)})$ (or apply PMI)
  until stopping criterion is met
  $\sigma \leftarrow \mathscr{T} \cdot \bigotimes_{i=1}^{\ell} \big(\mathbf{v}^{(i)}\big)^{\otimes a_i}$

Algorithm 2a PMI – Power Method Iteration

Input: Partially symmetric tensor

$$\mathscr{T} \in \bigotimes_{i=1}^{\ell} S^{a_i}(\mathbb{R}^{m_i})$$

Input: Unit norm vectors $(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)})$
Input: Shift parameters $\gamma_1, \dots, \gamma_\ell \geq 0$
Output: Vectors $(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)})$ after a power method iteration on $\mathscr{T}$.
  for $i = 1, \dots, \ell$ **do**
    $\tilde{\mathbf{v}}^{(i)} \leftarrow \mathscr{T} \cdot \bigotimes_{j=1}^{\ell} \big(\mathbf{v}^{(j)}\big)^{\otimes a_j - \delta_{i,j}} + \gamma_i \mathbf{v}^{(i)}$
    if $\|\tilde{\mathbf{v}}^{(i)}\| > 0$ **then**
      $\mathbf{v}^{(i)} \leftarrow \tilde{\mathbf{v}}^{(i)} / \|\tilde{\mathbf{v}}^{(i)}\|$

Algorithm 2b PMIe – Efficient Power Method Iteration

Input: Partially symmetric tensor

$$\mathscr{T} \in \bigotimes_{i=1}^{\ell} S^{a_i}(\mathbb{R}^{m_i})$$

Input: Unit norm vectors $(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)})$
Input: Shift parameters $\gamma_1, \dots, \gamma_\ell \geq 0$
Output: Vectors $(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)})$ after a power method iteration on $\mathscr{T}$.
  if $\ell = 1$ **then**
    $\tilde{\mathbf{v}}^{(1)} \leftarrow \mathrm{PMI}(\mathscr{T}; \mathbf{v}^{(1)})$
  else
    $k \leftarrow \lfloor \frac{\ell}{2} \rfloor$
    $\mathscr{U} \leftarrow \mathscr{T} \cdot \bigotimes_{j=k+1}^{\ell} \big(\mathbf{v}^{(j)}\big)^{\otimes a_j}$
    $(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(k)}) \leftarrow \mathrm{PMIe}(\mathscr{U}; \mathbf{v}^{(1)}, \dots, \mathbf{v}^{(k)})$
    $\mathscr{W} \leftarrow \mathscr{T} \cdot \bigotimes_{j=1}^{k} \big(\mathbf{v}^{(j)}\big)^{\otimes a_j}$
    $(\mathbf{v}^{(k+1)}, \dots, \mathbf{v}^{(\ell)}) \leftarrow \mathrm{PMIe}(\mathscr{W}; \mathbf{v}^{(k+1)}, \dots, \mathbf{v}^{(\ell)})$

In the fully asymmetric case, with $\gamma_1 = \cdots = \gamma_\ell = 0$, Algorithm 1 reduces to the classical HOPM, known to converge globally to a second-order critical point of $(\mathbf{v}_1, \dots, \mathbf{v}_\ell) \mapsto \mathscr{T} \cdot \bigotimes_{i=1}^{\ell} \mathbf{v}_i^{a_i}$, see [**51**]. In the fully symmetric case, it specializes to the shifted power method, which also has global convergence to a second-order critical point [**29**]. A related work in the partially symmetric setting is the recent paper [**39**], whose alternating algorithm is equivalent to our update scheme in Algorithm 2a. The authors show monotonic increase of the function values, but do not address convergence of the iterates, the choice of shifts, or local convergence rates. We address these three aspects in the following subsections.

We relate Algorithms 2a and 2b from the point of view of computational complexity. Algorithm 2a requires ℓ tensor contractions with $\mathscr{T}$, which requires ℓM multiplications, where $M = \prod_{i=1}^{\ell} m_i^{a_i}$ is the number of entries of $\mathscr{T}$. In contrast, Algorithm 2b rearranges the computation to be more efficient, as follows. It only requires two large tensor contractions with $\mathscr{T}$, which requires $2M$ multiplications, followed by contractions with two smaller tensors, which have costs M_1 and M_2 where $M_1 = \prod_{i=1}^{k} m_i^{a_i}$ and $M_2 = \prod_{i=k+1}^{\ell} m_i^{a_i}$, and so on recursively. If $M_1, M_2 = O(\sqrt{M})$, then the cost of the efficient method is $2M + O(\sqrt{M})$.

5.1. Global convergence. We show that Algorithm 1 has global convergence to a pSVT for suitable shifts $\gamma_1, \ldots, \gamma_\ell$. We refer to the variables at one symmetric factor $S^{a_i}(\mathbb{R}^{m_i})$ as a block. A block update means we update the vector in one block, fixing the other factors. Let $\|\cdot\|_2$ denote the spectral norm of a tensor, the largest absolute value of its singular values.

THEOREM 5.1. *Fix $\mathscr{T} \in \bigotimes_{i=1}^{\ell} S^{a_i}(\mathbb{R}^{m_i})$. Define the function $F : \mathbb{S}^{m_1-1} \times \cdots \times \mathbb{S}^{m_\ell-1} \to \mathbb{R}$ by*

$$F(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(\ell)}) = \Big\langle \mathscr{T}, \bigotimes_{i=1}^{\ell} \big(\mathbf{v}^{(i)}\big)^{\otimes a_i} \Big\rangle + \sum_{i=1}^{\ell} \gamma_i \|\mathbf{v}^{(i)}\|^{a_i}.$$

Then the pSVTs of $\mathscr{T}$ are the first-order critical points of F. There exists $\gamma_i \geq 0$ such that the restriction of F to the i-th block is convex over $\mathbb{R}^{m_i}$ for all values of the other blocks. A sufficient choice is $\gamma_i \geq (a_i - 1)\|\mathscr{T}\|_2$. If all block restrictions are convex over $\mathbb{R}^{m_i}$, then Algorithm 1 applied to $\mathscr{T}$ converges globally to a pSVT of $\mathscr{T}$. The convergence rate is at least algebraic.

PROOF. The first-order critical points of F are the first-order critical points of

$$\Big\langle \mathscr{T}, \bigotimes_{i=1}^{\ell} \big(\mathbf{v}^{(i)}\big)^{\otimes a_i} \Big\rangle,$$

which are the pSVTs of $\mathscr{T}$, by [**36**]. For $i \in [\ell]$, let F_i be the restriction of F to the i-th block with the other blocks fixed. The gradient and Hessian are

$$\nabla F_i(\mathbf{v}^{(i)}) = a_i\, \mathscr{T} \cdot \bigotimes_{j=1}^{\ell} \big(\mathbf{v}^{(j)}\big)^{\otimes (a_j - \delta_{i,j})} + a_i \gamma_i \|\mathbf{v}^{(i)}\|^{a_i - 2} \mathbf{v}^{(i)},$$

(5)

$$\nabla^2 F_i(\mathbf{v}^{(i)}) = a_i(a_i - 1)\, \mathscr{T} \cdot \bigotimes_{j=1}^{\ell} \big(\mathbf{v}^{(j)}\big)^{\otimes (a_j - 2\delta_{i,j})} + a_i(a_i - 2)\gamma_i \|\mathbf{v}^{(i)}\|^{a_i - 4} (\mathbf{v}^{(i)})^{\otimes 2} + a_i \gamma_i \|\mathbf{v}^{(i)}\|^{a_i - 2} I.$$

For convexity of F_i, we require $\mathbf{y}^{\mathsf{T}} \nabla^2 F_i(\mathbf{v}^{(i)}) \mathbf{y} \geq 0$ for all $\mathbf{y} \in \mathbb{S}^{m_i - 1}$. If $a_i \geq 2$, this is bounded from below by $-a_i(a_i - 1)\|\mathscr{T}\|_2 + a_i \gamma_i$, so it is non-negative if $\gamma_i \geq (a_i - 1)\|\mathscr{T}\|_2$, using (5) and $\|\mathbf{v}^{(i)}\| = 1$. If $a_i = 1$, the inequality holds by Cauchy-Schwarz.

We show that each block update in Algorithm 1 increases the value of F. The update for $\mathbf{v}^{(i)}$ is

$$\mathbf{v}^{(i)\prime} = \frac{\mathscr{T} \cdot \bigotimes_{j=1}^{\ell} \big(\mathbf{v}^{(j)}\big)^{\otimes a_j - \delta_{i,j}} \; + \; \gamma_i \mathbf{v}^{(i)}}{\big\| \mathscr{T} \cdot \bigotimes_{j=1}^{\ell} \big(\mathbf{v}^{(j)}\big)^{\otimes a_j - \delta_{i,j}} \; + \; \gamma_i \mathbf{v}^{(i)} \big\|} = \frac{\nabla F_i(\mathbf{v}^{(i)})}{\|\nabla F_i(\mathbf{v}^{(i)})\|}, \quad \text{for } \nabla F_i(\mathbf{v}^{(i)}) \neq 0.$$

If $\nabla F_i(\mathbf{v}^{(i)}) = 0$, we set $\mathbf{v}^{(i)\prime} = \mathbf{v}^{(i)}$. When F_i is convex, updating a unit vector $\mathbf{v}^{(i)}$ by normalizing the gradient of F_i ensures that F increases, unless $\mathbf{v}^{(i)} = \mathbf{v}^{(i)\prime}$, see e.g. [**30, 44**]

for a proof and [**32**] for a tensor application. We quantify the increase. Let $\mathbf{v} = (\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(\ell)})$ and $\mathbf{v}' = (\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(i)\prime}, \ldots, \mathbf{v}^{(\ell)})$. Convexity of F_i gives

$$\begin{aligned} F(\mathbf{v}') - F(\mathbf{v}) &\geq \langle \nabla F_i(\mathbf{v}),\, \mathbf{v}^{(i)\prime} - \mathbf{v}^{(i)} \rangle \\ &= \|\nabla F_i(\mathbf{v})\| \langle \mathbf{v}^{(i)\prime}, \mathbf{v}^{(i)\prime} - \mathbf{v}^{(i)} \rangle \\ &= \tfrac{1}{2} \|\nabla F_i(\mathbf{v})\| \, \|\mathbf{v}^{(i)\prime} - \mathbf{v}^{(i)}\|^2. \end{aligned} \tag{6}$$

The Riemannian gradient of F_i on the sphere $\mathbb{S}^{m_i-1}$ is $\nabla_{\mathbb{S}^{m_i-1}} F_i(\mathbf{v}) = (I - \mathbf{v}^{(i)}\mathbf{v}^{(i)\mathsf{T}})\nabla F_i(\mathbf{v})$. Substituting $\nabla F_i = \|\nabla F_i\| \mathbf{v}^{(i)\prime}$ yields

$$\|\nabla_{\mathbb{S}^{m_i-1}} F_i(\mathbf{v})\|^2 = \|\nabla F_i(\mathbf{v})\|^2 \big(1 - \langle \mathbf{v}^{(i)}, \mathbf{v}^{(i)\prime} \rangle^2\big) \;\leq\; \|\nabla F_i(\mathbf{v})\|^2 \, \|\mathbf{v}^{(i)\prime} - \mathbf{v}^{(i)}\|^2. \tag{7}$$

Combining (6) and (7) gives

$$F(\mathbf{v}') - F(\mathbf{v}) \;\geq\; \tfrac{1}{2} \|\nabla_{\mathbb{S}^{m_i-1}} F_i(\mathbf{v})\| \, \|\mathbf{v}^{(i)\prime} - \mathbf{v}^{(i)}\|. \tag{8}$$

We now prove the convergence of Algorithm 1. The proof adapts [**49**, Theorem 2.3] to blockwise updates. Let

$$\mathbf{v}_{(k,i)} = \big(\mathbf{v}^{(1,k+1)}, \ldots, \mathbf{v}^{(i-1,k+1)}, \mathbf{v}^{(i,k+1)}, \mathbf{v}^{(i+1,k)}, \ldots, \mathbf{v}^{(\ell,k)}\big)$$

be the tuple in the k-th iteration after updating block i and set $\mathbf{v}_{(k,\ell+1)} = \mathbf{v}_{(k+1,1)}$. The sequence $\{\mathbf{v}_{(k,i)}\}_{k\geq 0, 1\leq i\leq \ell}$ has a subsequence that converges to a point

$$\mathbf{v}_* = \big(\mathbf{v}_*^{(1)}, \ldots, \mathbf{v}_*^{(\ell)}\big),$$

since the ambient space $\mathbb{S}^{m_1-1} \times \cdots \times \mathbb{S}^{m_\ell - 1}$ is compact. Each F_i is real-analytic on $\mathbb{S}^{m_i-1}$. Hence there exist constants $\theta_i \in [\frac{1}{2}, 1)$, $\Lambda_i > 0$, and $\sigma > 0$ such that for all $\mathbf{v}^{(i)}$ with $\|\mathbf{v}^{(i)} - \mathbf{v}_*^{(i)}\| \leq \sigma$, the Łojasiewicz inequality [**37**, p92] holds:

$$\big|F_i(\mathbf{v}^{(i)}) - F_i(\mathbf{v}_*^{(i)})\big|^{\theta_i} \;\leq\; \Lambda_i \big\|\nabla_{\mathbb{S}^{m_i-1}} F_i(\mathbf{v}^{(i)})\big\|. \tag{9}$$

Set $\theta = \max_i \theta_i$ and $\Lambda = \max_i \Lambda_i$. The sequence $\{F(\mathbf{v}_{(k,i)})\}_{k\geq 0}$ increases monotonically, so it converges to $F(\mathbf{v}_*)$. Suppose $\mathbf{v}_{(k,i)}$ is close enough to $\mathbf{v}_*$ and k is large so that

$$\|\mathbf{v}_{(k,i)} - \mathbf{v}_*\| \leq \tfrac{\sigma}{2}, \quad \big(F(\mathbf{v}_*) - F(\mathbf{v}_{(k,i)})\big)^{1-\theta} \leq \min\{\tfrac{\sigma}{4\Lambda}, \frac{1}{2}\}. \tag{10}$$

Then equation (9) holds for $\mathbf{v}_{(k,i)}$ and together with (8), we obtain

$$\big(F(\mathbf{v}_*) - F(\mathbf{v}_{(k,i)})\big)^{\theta_i} \|\mathbf{v}^{(i,k+1)} - \mathbf{v}^{(i,k)}\| \;\leq\; 2\Lambda_i (F(\mathbf{v}_{(k,i+1)}) - F(\mathbf{v}_{(k,i)})).$$

We define $G(\mathbf{v}) \coloneqq F(\mathbf{v}_*) - F(\mathbf{v})$, whose function value is non-increasing after each update. If $G(\mathbf{v}_{(k,i)}) = 0$, then $\mathbf{v}_* = \mathbf{v}_{k,i}$ and the proof is complete. Otherwise, we have

$$\begin{aligned} \|\mathbf{v}^{(i,k+1)} - \mathbf{v}^{(i,k)}\| \;&\leq\; 2\Lambda_i \frac{G(\mathbf{v}_{(k,i)}) - G(\mathbf{v}_{(k,i+1)})}{G(\mathbf{v}_{(k,i)})^{\theta_i}} \\ &\leq\; 2\Lambda \big(G(\mathbf{v}_{(k,i)})^{1-\theta_i} - G(\mathbf{v}_{(k,i+1)})^{1-\theta_i}\big) \\ &\leq\; 2\Lambda \big(G(\mathbf{v}_{(k,i)})^{1-\theta} - G(\mathbf{v}_{(k,i+1)})^{1-\theta}\big) \\ &\leq\; 2\Lambda G(\mathbf{v}_{(k,i)})^{1-\theta} \leq \tfrac{\sigma}{2}, \end{aligned} \tag{11}$$

where the second last inequality follows from the fact that the function $f(x) = x^{1-\theta} - x^{1-\theta_i}$ takes non-negative values and is monotonically increasing for small $x > 0$ and the last inequality follows from (10). By the triangle inequality, $\mathbf{v}_{(k,i+1)}$ also satisfies $\|\mathbf{v}_{(k,i+1)} - \mathbf{v}_*\| \le \sigma$, so the Łojasiewicz inequality holds for $\mathbf{v}_{(k,i+1)}$. Hence all iterates $\mathbf{v}_{(k',j)}$ after $\mathbf{v}_{(k,i)}$ satisfy

$$\|\mathbf{v}_{(k',j)} - \mathbf{v}_*\| \le \frac{\sigma}{2},$$

by induction, so the Łojasiewicz inequality holds for all $\mathbf{v}_{(k',j)}$ around $\mathbf{v}_*$. Summing over iterates, we obtain from (11) that

$$\begin{aligned}\sum_{(k,i)<(m,n)\le(k',j)} \|\mathbf{v}_{(m,n)} - \mathbf{v}_{(m,n-1)}\| &\le \sum_{(k,i)<(m,n)\le(k',j)} 2\Lambda\big(G(\mathbf{v}_{(m,n-1)})^{1-\theta} - G(\mathbf{v}_{(m,n)})^{1-\theta}\big)\\ &\le 2\Lambda\, G(\mathbf{v}_{(k,i)})^{1-\theta} < \tfrac{\sigma}{2}.\end{aligned}$$

Therefore, $\{\mathbf{v}_{(k,i)}\}$ is Cauchy and converges to $\mathbf{v}_* \in \mathbb{S}^{m_1-1} \times \cdots \times \mathbb{S}^{m_\ell-1}$. The convergence rate is at least algebraic by [**49**, Theorem 2.3].

We are left to check that $\mathbf{v}_*$ is a pSVT of $\mathscr{T}$. Taking limits on the update rule of Algorithm 1, we have

$$\mathbf{v}_*^{(i)} \| \mathscr{T} \cdot \bigotimes_{j=1}^{\ell} (\mathbf{v}_*^{(j)})^{\otimes(a_i - \delta_{i,j})} + \gamma_i \mathbf{v}_*^{(i)} \| = \mathscr{T} \cdot \bigotimes_{j=1}^{\ell} (\mathbf{v}_*^{(j)})^{\otimes(a_i - \delta_{i,j})} + \gamma_i \mathbf{v}_*^{(i)},$$

which implies $\sigma \mathbf{v}_*^{(i)} = \mathscr{T} \cdot \bigotimes_{j=1}^{\ell} (\mathbf{v}_*^{(j)})^{\otimes(a_i - \delta_{i,j})}$, where $\sigma = \mathscr{T} \cdot \bigotimes_{j=1}^{\ell} (\mathbf{v}_*^{(j)})^{\otimes a_i}$. □

REMARK 5.2. *In the above proof, we update the blocks in a fixed order in each outer iteration. Global convergence still holds for any order with every block updated at least once per outer iteration.*

Our partially symmetric tensor decomposition computes pSVTs with singular value one of a tensor $\mathscr{T}^{(f)}$ with last slices orthonormal. To speed up convergence, we introduce a new tensor $\widetilde{\mathscr{T}}^{(f)}$ such that the pSVTs of $\mathscr{T}^{(f)}$ and $\widetilde{\mathscr{T}}^{(f)}$ are in one-to-one correspondence. The proof of this correspondence is in the appendix. The tensor $\widetilde{\mathscr{T}}^{(f)}$ is a symmetrized combination of the orthonormal slices of $\mathscr{T}^{(f)}$. The shift parameters for $\widetilde{\mathscr{T}}^{(f)}$ can be chosen to be less than one, which is smaller than the shifts in Theorem 5.1. Smaller shifts imply a faster increase in the function value and hence faster convergence. For numerical computations, it is not necessary to form $\widetilde{\mathscr{T}}^{(f)}$, since we only require its contractions with rank-one tensors, which can be computed using $\mathscr{T}^{(f)}$.

DEFINITION 5.3 (Partial symmetrization P_{sym} and tensor $\widetilde{\mathscr{T}}^{(f)}$). Fix a partially symmetric tensor $\mathscr{T}^{(f)} \in \bigotimes_{i=1}^{\ell} S^{f_i}(\mathbb{R}^{m_i}) \otimes \mathbb{R}^r$ whose last slices $\mathscr{S}_1, \ldots, \mathscr{S}_r$ are orthonormal. Define the *partial symmetrization operator* P_{sym} to be the linear projection

$$P_{\text{sym}} \colon \Big(\bigotimes_{i=1}^{\ell} S^{f_i}(\mathbb{R}^{m_i})\Big)^{\otimes 2} \to \bigotimes_{i=1}^{\ell} S^{2f_i}(\mathbb{R}^{m_i}).$$

Define $\widetilde{\mathscr{T}}^{(f)} := \sum_{i=1}^{r} P_{\text{sym}}(\mathscr{S}_i \otimes \mathscr{S}_i)$.

Notice the order of $\widetilde{\mathscr{T}}^{(f)}$ is double the order of $\mathscr{T}^{(f)}$. The next proposition gives shifts γ_i that ensure global convergence of Algorithm 1 applied to $\widetilde{\mathscr{T}}^{(f)}$. Its proof is in the appendix.

PROPOSITION 5.4. *Fix a partially symmetric tensor $\mathscr{T}^{(f)} \in \bigotimes_{i=1}^{\ell} S^{f_i}(\mathbb{R}^{m_i}) \otimes \mathbb{R}^r$ whose last slices $\mathscr{S}_1, \ldots, \mathscr{S}_r$ are orthonormal. Let $\widetilde{\mathscr{T}}^{(f)}$ be defined as in Definition 5.3. Then Algorithm 1 applied to $\widetilde{\mathscr{T}}^{(f)}$ converges globally to a pSVT of $\widetilde{\mathscr{T}}^{(f)}$ with shifts*

(a) $\gamma_i = 0$ *when* $f_i = 1$.
(b) $\gamma_i = \sqrt{\frac{f_i - 1}{f_i}}\, h(\nu)$ *when* $f_i > 1$, *where* $F(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)}) \le \nu$ *and*

$$h(\nu) = \begin{cases} 1 - \frac{\nu}{2}, & \nu \le \frac{2}{3}, \\ \sqrt{2\nu(1-\nu)}, & \nu > \frac{2}{3}. \end{cases}$$

The shifts for $\widetilde{\mathscr{T}(f)}$ are strictly less than 1 and can be smaller than the shifts in Theorem 5.1. To connect the construction of $\widetilde{\mathscr{T}(f)}$ with our optimization problem, we introduce a projection operator $P_{\mathcal{A}}$ and objective function $F_{\mathcal{A}}$, which measures how close a rank-one tensor is to the subspace $\mathcal{A}$. We then relate $F_{\mathcal{A}}$ to the tensors $\mathscr{T}(f)$ and $\widetilde{\mathscr{T}(f)}$.

Definition 5.5 (Projection and objective function). Let $\mathscr{T}(f) \in \bigotimes_{i=1}^{\ell} S^{f_i}(\mathbb{R}^{m_i}) \otimes \mathbb{R}^r$ be a partially symmetric tensor with last slices orthonormal that span $\mathcal{A} \subseteq \bigotimes_{i=1}^{\ell} S^{f_i}(\mathbb{R}^{m_i})$.

- The projection $P_{\mathcal{A}}$ is the orthogonal projection onto $\mathcal{A}$ of a tensor or its vectorization.
- The objective function $F_{\mathcal{A}}$ quantifies proximity to $\mathcal{A}$:

$$F_{\mathcal{A}} \colon \prod_{i=1}^{\ell} \mathbb{S}^{m_i - 1} \to \mathbb{R}, \quad F_{\mathcal{A}}(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)}) = \Big\| P_{\mathcal{A}} \Big(\bigotimes_{i=1}^{\ell} \big(\mathbf{v}^{(i)}\big)^{\otimes f_i} \Big) \Big\|^2.$$

5.2. Local convergence. In the following result, we show local linear convergence of Algorithm 1 to certain pSVTs of a tensor with two blocks (i.e. $\ell = 2$). This generalizes [**29**, Theorem 4.10], which is the case $\ell = 1$. For our application to CP decomposition, this provides a guarantee for computing pSVTs of $\widetilde{\mathscr{T}(f)}$ defined in Definition 5.3 with two blocks.

Proposition 5.6. *Let* $\mathscr{T} \in S^{d_1}(\mathbb{R}^{m_1}) \otimes S^{d_2}(\mathbb{R}^{m_2})$ *be a partially symmetric tensor whose pSVTs have singular value at most one. Let* $\gamma_1, \gamma_2 \ge 0$ *be such that*

$$F(\mathbf{v}^{(1)}, \mathbf{v}^{(2)}) = \big\langle \mathscr{T}, (\mathbf{v}^{(1)})^{\otimes d_1} \otimes (\mathbf{v}^{(2)})^{\otimes d_2} \big\rangle + \gamma_1 \|\mathbf{v}^{(1)}\|^{d_1} + \gamma_2 \|\mathbf{v}^{(2)}\|^{d_2}$$

is convex in each block. Then, critical points of F *are pSVTs of* $\mathscr{T}$. *PS–HOPM on* $\mathscr{T}$ *with shifts* γ_1, γ_2 *has local linear convergence around every pSVT that is a second-order critical point of* F *with positive singular value on the product of spheres* $\mathbb{S}^{m_1 - 1} \times \mathbb{S}^{m_2 - 1}$.

Proof. Let F_1, F_2 denote the restrictions of F to the first and second blocks, respectively. They are convex, by assumption. Denote the Euclidean gradient of F_i over the i-th block by $\nabla F_i(\mathbf{v}^{(i)})$. The PS–HOPM update in block 1 is the normalized gradient map

$$\Psi_1(\mathbf{v}^{(1)}, \mathbf{v}^{(2)}) = \left(\frac{\nabla F_1(\mathbf{v}^{(1)})}{\|\nabla F_1(\mathbf{v}^{(1)})\|}, \mathbf{v}^{(2)} \right),$$

and similarly for block 2 and Ψ_2. The overall update is the composition $\Psi = \Psi_2 \circ \Psi_1$. A second-order critical point $\mathbf{v}_\star = (\mathbf{v}^{(1)}, \mathbf{v}^{(2)})$ with singular value σ of F is a fixed point of this map. Local linear convergence follows if the Jacobian

$$J\Psi = J\Psi_2(\mathbf{v}_\star)\, J\Psi_1(\mathbf{v}_\star)$$

has spectral radius strictly less than one, by the center–stable manifold theorem (see, e.g., [**45**, Theorem 3.5]).

We compute the Jacobians $J\Psi_2, J\Psi_1$ and relate them to the Riemannian Hessian of F at $\mathbf{v}_\star$. Let $\mathbf{H}_{i,j}$ denote the (i,j)–block of the Riemannian Hessian of F, where $i, j \in \{1, 2\}$ and define $\mathbf{P}_i = \mathbf{I} - \mathbf{v}^{(i)} \mathbf{v}^{(i)\mathsf{T}}$ and $\alpha_i = \|\nabla F_i(\mathbf{v}^{(i)})\|$. The gradient of F_i is $\nabla F_i(\mathbf{v}^{(i)}) = d_i\, \mathscr{T} \cdot$

$\bigotimes_{j=1}^{2}\big(\mathbf{v}^{(j)}\big)^{\otimes(d_j-\delta_{i,j})} + d_i\gamma_i\mathbf{v}^{(i)} = d_i(\sigma+\gamma_i)\mathbf{v}^{(i)}$, so $\alpha_i = d_i(\sigma+\gamma_i) > 0$ by the assumption $\sigma > 0$. We have

$$\frac{\partial\Psi_i(\mathbf{v}^{(i)})}{\partial\mathbf{v}_j^{(i)}} = \frac{\frac{\partial}{\partial\mathbf{v}_j^{(i)}}\nabla F_i(\mathbf{v}^{(i)})}{\|\nabla F_i(\mathbf{v}^{(i)})\|} - \frac{\nabla F_i(\mathbf{v}^{(i)})}{\|\nabla F_i(\mathbf{v}^{(i)})\|^2}\frac{\big\langle\nabla F_i(\mathbf{v}^{(i)}),\ \frac{\partial}{\partial\mathbf{v}_j^{(i)}}\nabla F_i(\mathbf{v}^{(i)})\big\rangle}{\|\nabla F_i(\mathbf{v}^{(i)})\|}.$$

It follows that the $(1,1)$–block of the Jacobian $J\Psi_1(\mathbf{v}^{(1)})$ (as a linear operator on the tangent space $T_{\mathbf{v}^{(1)}}\mathbb{S}^{m_1-1}$) is

$$(J\Psi_1)_{1,1} = \mathbf{P}_1\left(\frac{\nabla^2F_1(\mathbf{v}^{(1)})}{\alpha_1} - \frac{\nabla^2F_1(\mathbf{v}^{(1)})\nabla F_1(\mathbf{v}^{(1)})\nabla F_1(\mathbf{v}^{(1)})^\intercal}{\alpha_1^3}\right)\mathbf{P}_1 = \mathbf{P}_1\frac{\nabla^2F_1(\mathbf{v}^{(1)})}{\alpha_1}\mathbf{P}_1, \tag{12}$$

where the second term is zero since $\nabla F_1(\mathbf{v}^{(1)}) = \alpha_1\mathbf{v}^{(1)}$ at the fixed point $\mathbf{v}_\star$. The $(1,1)$ Riemannian Hessian block is

$$\mathbf{H}_{1,1} = \mathbf{P}_1\big(\nabla^2F_1(\mathbf{v}^{(1)}) - \langle\nabla F_1(\mathbf{v}^{(1)}),\mathbf{v}^{(1)}\rangle\mathbf{I}\big)\mathbf{P}_1 = \alpha_1(J\Psi_1)_{1,1} - \alpha_1\mathbf{P}_1,$$

so we have

$$(J\Psi_1)_{1,1} = \frac{\mathbf{H}_{1,1}}{\alpha_1} + \mathbf{P}_1. \tag{13}$$

Similarly, the $(1,2)$–block of the Jacobian $J\Psi_1$ is

$$(J\Psi_1)_{1,2} = \frac{\mathbf{H}_{1,2}}{\alpha_1}.$$

Altogether, the Jacobian of the full iteration is

$$J\Psi = \begin{pmatrix}\mathbf{I} & 0\\ \frac{\mathbf{H}_{2,1}}{\alpha_2} & \frac{\mathbf{H}_{2,2}}{\alpha_2}+\mathbf{P}_2\end{pmatrix}\begin{pmatrix}\frac{\mathbf{H}_{1,1}}{\alpha_1}+\mathbf{P}_1 & \frac{\mathbf{H}_{1,2}}{\alpha_1}\\ 0 & \mathbf{I}\end{pmatrix}.$$

The matrix $(J\Psi_1)_{1,1}$ is symmetric and its eigenvalues are at least 0 by (12) and F_1 being convex, and they are at most 1 by (13) and since $\mathbf{H}_{1,1}$ is negative definite, being a submatrix of the Hessian at a maximizer.

To analyze the spectrum, suppose $(\mathbf{v},\mathbf{w})$ is an normalized eigenvector of $J\Psi$ with eigenvalue λ. We will show that λ is real. Expanding $J\Psi(\mathbf{v},\mathbf{w}) = \lambda(\mathbf{v},\mathbf{w})$ gives

$$\lambda\mathbf{v} = \left(\frac{\mathbf{H}_{1,1}}{\alpha_1}+\mathbf{P}_1\right)\mathbf{v} + \frac{\mathbf{H}_{1,2}}{\alpha_1}\mathbf{w}, \qquad \lambda\mathbf{w} = \frac{\mathbf{H}_{2,1}}{\alpha_2}\lambda\mathbf{v} + \left(\frac{\mathbf{H}_{2,2}}{\alpha_2}+\mathbf{P}_2\right)\mathbf{w}. \tag{14}$$

Multiplying the equations (14) with the Hermitian transposes $\mathbf{v}^H$ and $\mathbf{w}^H$ respectively, and setting

$$\beta_1 = \mathbf{v}^H\Big(\frac{\mathbf{H}_{1,1}}{\alpha_1}+\mathbf{P}_1\Big)\mathbf{v}, \quad \beta_2 = \mathbf{w}^H\Big(\frac{\mathbf{H}_{2,2}}{\alpha_2}+\mathbf{P}_2\Big)\mathbf{w}, \quad \mu = \mathbf{v}^H\mathbf{H}_{1,2}\mathbf{w}, \tag{15}$$

we obtain

$$\lambda = \beta_1 + \frac{\mu}{\alpha_1}, \qquad \lambda = \frac{\lambda}{\alpha_2}\bar{\mu} + \beta_2. \tag{16}$$

Note that β_1,β_2 are real non-negative numbers with $\beta_1+\beta_2 \le 1$ since the symmetric matrices $\frac{\mathbf{H}_{1,1}}{\alpha_1}+\mathbf{P}_1, \frac{\mathbf{H}_{2,2}}{\alpha_2}+\mathbf{P}_2$ have eigenvalues in $[0,1]$ and $\|\mathbf{v}\|^2+\|\mathbf{w}\|^2 = 1$. Multiplying the first equation in (16) by $\alpha_1\bar{\lambda}$, the second equation by α_2 and taking the sum, we have

$$\alpha_1\|\lambda\|^2 + \lambda\alpha_2 = \beta_1\alpha_1\bar{\lambda} + \alpha_2\beta_2 + \mu\bar{\lambda} + \bar{\mu}\lambda.$$

It follows that $\beta_1\alpha_1\bar{\lambda} - \lambda\alpha_2$ is real. Then either λ is real or $\alpha_2 + \beta_1\alpha_1 = 0$. The second case cannot happen since $\alpha_2, \alpha_1 > 0, \beta_1 \geq 0$. Therefore, $\lambda \in \mathbb{R}$ and so is μ.

We now show that $0 \leq \lambda < 1$. If $\mu < 0$, then $\lambda = \beta_1 + \frac{\mu}{\alpha_1} < \beta_1 \leq 1$. Assume $\lambda < 0$, the relation $\lambda = \beta_2 + \frac{\lambda}{\alpha_2}\mu$ implies $\frac{\lambda}{\alpha_2}\mu < 0$ and hence $\mu > 0$, a contradiction. Thus, $0 \leq \lambda < 1$. If $\mu = 0$, then $0 \leq \lambda = \beta_1 = \beta_2 \leq \frac{1}{2}$, since $\beta_1, \beta_2 \geq 0$ and $\beta_1 + \beta_2 = 1$. If instead $\mu > 0$, then $\lambda = \beta_1 + \frac{\mu}{\alpha_1} > 0$. Then

$$2\mu + \alpha_1\beta_1 + \alpha_2\beta_2 \leq \alpha_1\|\mathbf{v}\|^2 + \alpha_2\|\mathbf{w}\|^2 \leq \max\{\alpha_1, \alpha_2\},$$

by (15) and the negative definiteness of the Hessian of F. Hence $\max\{\alpha_1, \alpha_2\} - \mu > 0$. Multiplying the equations of (16) by α_1 and α_2, respectively, and by the above inequality, we obtain

$$(\alpha_1 + \alpha_2)\lambda = (\lambda - 1)\mu + \alpha_1\beta_1 + \alpha_2\beta_2 + 2\mu < \max\{\alpha_1, \alpha_2\} + (\lambda - 1)\mu.$$

Hence, $\lambda < \frac{\max\{\alpha_1,\alpha_2\}-\mu}{\alpha_1+\alpha_2-\mu} < 1$. □

We have shown local linear convergence to a global maximizer of $F_{\mathcal{A}}$. In practice, PS-HOPM may converge to a critical point of $F_{\mathcal{A}}$ that is not globally optimal. We show that a critical point of $F_{\mathcal{A}}$ has low objective value or is close to an optimal rank-one component.

THEOREM 5.7. *Fix $\mathscr{T}^{(f)} = \sum_{i=1}^r \lambda_i \bigotimes_{j=1}^{\ell} (\mathbf{v}_i^{(j)})^{\otimes f_j} \otimes \mathbf{u}_i$, with last slices orthonormal with span $\mathcal{A}$. Define the rank-one tensors*

$$\mathscr{R}^{(i)} := \bigotimes_{j=1}^{\ell} (\mathbf{v}_i^{(j)})^{\otimes f_j}, \qquad i = 1, \ldots, r.$$

Assume that $\mathcal{A} = \operatorname{span}\{\mathscr{R}^{(1)}, \ldots, \mathscr{R}^{(r)}\}$. For $(\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(\ell)}) \in \mathbb{S}^{m_1-1} \times \cdots \times \mathbb{S}^{m_\ell-1}$, let $\mathscr{T}_{\mathbf{w}} := \bigotimes_{j=1}^{\ell} (\mathbf{w}^{(j)})^{\otimes f_j}$. Then the objective function $F_{\mathcal{A}}$ satisfies

$$\max_i |\langle \mathscr{R}^{(i)}, \mathscr{T}_{\mathbf{w}} \rangle|^2 \;\leq\; F_{\mathcal{A}}(\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(\ell)}) \;\leq\; C \max_i |\langle \mathscr{R}^{(i)}, \mathscr{T}_{\mathbf{w}} \rangle|,$$

where $C = (1-\rho)^{-1}\Big(\prod_{k=1}^{\ell}(1 + (r-1)\rho_k)\Big)^{1/\ell}$, where $\rho := \max_{i \neq j} |\langle \mathscr{R}^{(i)}, \mathscr{R}^{(j)} \rangle|$ and, for $k = 1, \ldots, \ell$, $\rho_k := \max_{i \neq j} |\langle \mathbf{v}_i^{(k)}, \mathbf{v}_j^{(k)} \rangle|$.

PROOF. We have

$$F_{\mathcal{A}}(\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(\ell)}) = \|P_{\mathcal{A}}(\mathscr{T}_{\mathbf{w}})\|^2 \;\geq\; \max_i \|P_{\operatorname{span}\{\mathscr{T}^{(i)}\}}(\mathscr{T}_{\mathbf{w}})\|^2 = \max_i |\langle \mathscr{R}^{(i)}, \mathscr{T}_{\mathbf{w}} \rangle|^2.$$

This gives the lower bound. For the upper bound, there exist scalars $\mu_1, \ldots, \mu_r$ such that $P_{\mathcal{A}}(\mathscr{T}_{\mathbf{w}}) = \sum_{i=1}^r \mu_i \mathscr{R}^{(i)}$, so that $F_{\mathcal{A}}(\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(\ell)}) = \sum_{i=1}^r \mu_i \langle \mathscr{R}^{(i)}, \mathscr{T}_{\mathbf{w}} \rangle$. Since $P_{\mathcal{A}}(\mathscr{R}^{(i)}) = \mathscr{R}^{(i)}$, we have

$$\langle P_{\mathcal{A}}(\mathscr{T}_{\mathbf{w}}), \mathscr{R}^{(i)} \rangle = \langle \mathscr{T}_{\mathbf{w}}, \mathscr{R}^{(i)} \rangle = \mu_i + \sum_{j \neq i} \mu_j \langle \mathscr{R}^{(j)}, \mathscr{R}^{(i)} \rangle.$$

We obtain

$$|\mu_i - \langle \mathscr{R}^{(i)}, \mathscr{T}_{\mathbf{w}} \rangle| \leq \rho\|\mu\|_\infty \Rightarrow |\mu_i| \leq \rho\|\mu\|_\infty + |\langle \mathscr{R}^{(i)}, \mathscr{T}_{\mathbf{w}} \rangle|,$$

by comparing these two equalities. This yields an upper bound

$$\|\mu\|_\infty \leq \frac{\max_i |\langle \mathscr{R}^{(i)}, \mathscr{T}_{\mathbf{w}} \rangle|}{1 - \rho}.$$

Plugging this into the expression for $F_{\mathcal{A}}$, we obtain

$$F_{\mathcal{A}}(\mathbf{w}^{(1)},\dots,\mathbf{w}^{(\ell)}) \le \frac{\sum_{i=1}^{r}|\langle \mathscr{R}^{(i)},\mathscr{T}_{\mathbf{w}}\rangle|}{1-\rho}\max_i|\langle \mathscr{R}^{(i)},\mathscr{T}_{\mathbf{w}}\rangle|. \tag{17}$$

We have

$$\sum_{i=1}^{r}|\langle \mathscr{R}^{(i)},\mathscr{T}_{\mathbf{w}}\rangle| = \sum_{i=1}^{r}\prod_{j=1}^{\ell}|\langle \mathbf{v}_i^{(j)},\mathbf{w}^{(j)}\rangle|^{f_j} \le \left(\prod_{j=1}^{\ell}\sum_{i=1}^{r}|\langle \mathbf{v}_i^{(j)},\mathbf{w}^{(j)}\rangle|^{\ell f_j}\right)^{1/\ell}, \tag{18}$$

by Hölder's inequality. Let $\mathbf{M}_j$ be the $r\times m_j$ matrix with rows $\mathbf{v}_1^{(j)},\dots,\mathbf{v}_r^{(j)}$. Then,

$$\sum_{i=1}^{r}|\langle \mathbf{v}_i^{(j)},\mathbf{w}^{(j)}\rangle|^{\ell f_j} \le \sum_{i=1}^{r}|\langle \mathbf{v}_i^{(j)},\mathbf{w}^{(j)}\rangle|^2 = \|\mathbf{M}_j\mathbf{w}^{(j)}\|_2^2 \le \|\mathbf{M}_j\mathbf{M}_j^{\mathsf{T}}\|_2 \le 1+(r-1)\rho_j, \tag{19}$$

where the last inequality follows from Gershgorin's circle theorem [**26**]. Hence, we obtain the upper bound

$$F_{\mathcal{A}}(\mathbf{w}_1,\dots,\mathbf{w}_\ell) \le \frac{\left(\prod_{j=1}^{\ell}(1+(r-1)\rho_j)\right)^{\frac{1}{\ell}}}{1-\rho}\max_i|\langle \mathscr{R}^{(i)},\mathscr{T}_{\mathbf{w}}\rangle|,$$

from (17), (18), and (19). □

Remark 5.8. *When all the rows of each* $\mathbf{M}_j$ *are orthonormal, we have* $\rho=\rho_1=\cdots=\rho_\ell=0$, *and hence* $\max_i|\langle \mathscr{R}^{(i)},\mathscr{T}_{\mathbf{w}}\rangle|^2 \le F_{\mathcal{A}}(\mathbf{w}^{(1)},\dots,\mathbf{w}^{(\ell)}) \le \max_i|\langle \mathscr{R}^{(i)},\mathscr{T}_{\mathbf{w}}\rangle|$.

6. Completing the decomposition

We have seen how part of the rank-one summands in a decomposition of $\mathscr{T}$ can be obtained by computing pSVTs of the tensor $\mathscr{T}^{(f)}$ or equivalently $\overline{\mathscr{T}}^{(f)}$. To recap, let

$$\mathscr{T}=\sum_{i=1}^{r}\lambda_i\bigotimes_{j=1}^{\ell}(\mathbf{v}_i^{(j)})^{\otimes d_j}$$

where $\lambda_i\in\mathbb{R}$, $\mathbf{v}_i^{(j)}\in\mathbb{S}^{m_j-1}$, we choose a tuple $f=(f_1,\dots,f_\ell)$ and form a flattening matrix $\mathbf{T}^{(f)}$ whose column span is $\mathcal{A}$. Assume $r\le r(f)$. Then the summand $\mathscr{R}_{(f)}:=\bigotimes_{i=1}^{\ell}(\mathbf{v}_1^{(i)})^{\otimes f_i}$ is one of the rank-one tensors in $\mathcal{A}$. We now describe how to recover its complement

$$\mathscr{R}_{(d-f)}=\bigotimes_{i=1}^{\ell}(\mathbf{v}_1^{(i)})^{\otimes(d_i-f_i)}.$$

The complement is determined immediately if $f_i>0$ for all i. It remains to consider when $f_i=0$ for some i. This is unavoidable when decomposing fully asymmetric tensors, and in the partially symmetric case can be advantageous, as it may lead to a flattening that can decompose tensors of higher rank.

Definition 6.1. Let $d=(d_1,\dots,d_\ell)$ and $f=(f_1,\dots,f_\ell)$ with $d_i>0$ and $0\le f_i\le d_i$. The tuple f is *symmetry breaking* for d if f there exists an index i with $f_i>0$ and $d_i-f_i>0$.

We call such a tuple symmetry breaking becauase the same vector appears in both the tensor and its complement: after recovering $\mathscr{R}_{(f)}$ we know some of the vectors in $\mathscr{R}_{(d-f)}$. The tensor $\mathscr{R}_{(d-f)}$ can be reconstructed from those known factors together with the constraints of the row space $\mathbf{T}^{(f)}$, see Theorem 6.2 and Figure 3.

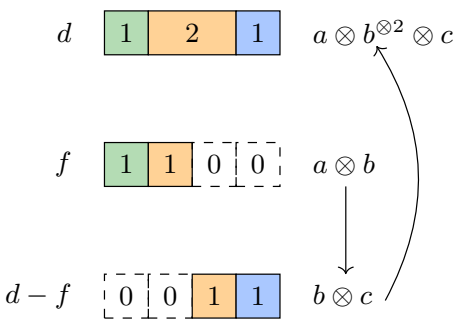


FIGURE 3. Symmetry–breaking case $d = (1,2,1)$, $f = (1,1,0)$: use shared vectors and the row space of $\mathbf{T}^{(f)}$ to recover the full summand.

When f is not symmetry breaking (for example $d = (1,1,1,1)$, $f = (1,1,0,0)$), one flattening is insufficient. In this case we use two tuples f, f' satisfying

$$f' \neq d - f, \qquad \min(f, f') \neq 0, \qquad f \not\geq f', \tag{20}$$

where $\min(f, f') \neq 0$ means $\nexists i$ such that $f_i = f'_i = 0$ and proceed as follows (see Figure 4)

(1) Recover a rank-one tensor in $\operatorname{colspan}(\mathbf{T}^{(f)})$.
(2) Use the shared factors of f and f' to recover a rank-one tensor in $\operatorname{colspan}(\mathbf{T}^{(f')})$.
(3) If $\min(f, f') > 0$, recovery is complete. Otherwise, use the shared factors between $d - f$ and f' to find the remaining factors.

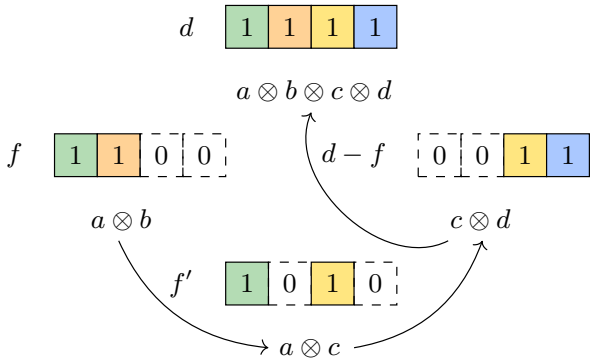


FIGURE 4. Steps to recover the rank-one tensor in the non-symmetry breaking case where $d = (1,1,1,1), f = (1,1,0,0), f' = (1,0,1,0)$.

We summarize the procedure for recovering $\mathscr{R}_{(d-f)}$ in Algorithm 3. We next show how to recover a rank-one tensor in a linear subspace when some of its factors are known.

THEOREM 6.2. *Let* $\mathcal{A} = \operatorname{span}\Big\{ \bigotimes_{j=1}^{\ell} (\mathbf{v}_i^{(j)})^{\otimes d_j} \ : \ 1 \leq i \leq r \Big\}$, *where the vectors* $\mathbf{v}_i^{(j)}$ *are generic with norm one, and assume* $\dim \mathcal{A} = r$. *Stack an orthonormal basis of* $\mathcal{A}$ *to form the tensor* $\mathscr{T}^{(f)} \in \bigotimes_{i=1}^{\ell} S^{d_i}(\mathbb{R}^{m_i}) \otimes \mathbb{R}^r$. *For* $1 \leq i \leq r$ *and* $a = (a_1, \ldots, a_\ell)$ *define*

$$\mathscr{R}^{(i)}_{(a)} := \bigotimes_{j=1}^{\ell} (\mathbf{v}_i^{(j)})^{\otimes a_j}.$$

Algorithm 3 Recovery of complementary factors based on the type of flattening f

Input: Flattening tuple f and partially symmetric tensor $\mathscr{T} = \sum_{i=1}^{r} \lambda_i \bigotimes_{j=1}^{\ell} (\mathbf{v}_i^{(j)})^{\otimes d_j}$. Rank-one tensor $\mathscr{R}_{(f)} = \bigotimes_{i=1}^{\ell} (\mathbf{v}_1^{(i)})^{\otimes f_i}$.

Output: Its complement $\mathscr{R}_{(d-f)} = \bigotimes_{i=1}^{\ell} (\mathbf{v}_1^{(i)})^{\otimes (d_i - f_i)}$.

if $f_i > 0$ for all i **then**

All factors in $\mathscr{R}_{(d-f)}$ also appear in $\mathscr{R}_{(f)}$, so it can be found directly.

else if f is **symmetry breaking then**

Some factors of $\mathscr{R}_{(d-f)}$ are known.

Reconstruct the remaining factors using the row span of $\mathbf{T}^{(f)}$ (see Figure 3).

else

Choose tuple f' such that

$$f \neq d - f', \qquad \min(f, f') \neq 0, \qquad f \not> f'.$$

Use the factors shared by f and f' to find a rank-one tensor in $\operatorname{colspan}(\mathbf{T}^{(f')})$.

if $\min(f, f') > 0$ **then**

Recovery is complete.

else

Use the factors shared by f' and $d-f$ to recover the remaining factors from $\mathbf{T}^{(d-f)}$.

For a tuple $f \notin \{0, (d_1, \ldots, d_\ell)\}$, let $\mathbf{M}$ flatten $\mathscr{T}^{(f)} \cdot \mathscr{R}_{(f)}^{(i)}$ into a matrix with r columns. Then $\operatorname{vect}(\mathscr{R}_{(d-f)}^{(i)})$ is the unique left singular vector of $\mathbf{M}$ with singular value 1 that is a vectorization of a rank-one tensor.

Proof. Let $(\mathbf{u}, \mathbf{v})$ be a singular pair of $\mathbf{M}$ with singular value σ. Let $\mathbf{M}_{\mathcal{A}}$ be the flattening of $\mathscr{T}^{(f)}$ via tuple $(d_1, \ldots, d_\ell, 0)$ with r orthonormal columns. By definition of $\mathbf{M}$, we have

$$\mathbf{M}_{\mathcal{A}}^{\mathsf{T}} \operatorname{vect}\big(\mathbf{u} \otimes \mathscr{R}_{(f)}^{(i)}\big) \;=\; \sigma \mathbf{v},$$

where we regard $\mathbf{u} \otimes \mathscr{R}_{(f)}^{(i)}$ as an element of $\bigotimes_{j=1}^{\ell} S^{d_j}(\mathbb{R}^{m_j})$ by appropriately reordering the tensor factors. We have

$$\sigma = \big\| \mathbf{M}_{\mathcal{A}}^{\mathsf{T}} \operatorname{vect}(\mathbf{u} \otimes \mathscr{R}_{(f)}^{(i)}) \big\|_2 \leq \|\mathbf{u} \otimes \mathscr{R}_{(f)}^{(i)}\|_2 = 1,$$

with equality when $\mathbf{u} \otimes \mathscr{R}_{(f)}^{(i)} \in \mathcal{A}$, since the last slices of $\mathscr{T}^{(f)}$ are orthonormal. Conversely, we assume $\mathbf{u} \otimes \mathscr{R}_{(f)}^{(i)} \in \mathcal{A}$ and define $\mathbf{v} = \mathbf{M}_{\mathcal{A}}^{\mathsf{T}} \operatorname{vect}(\mathbf{u} \otimes \mathscr{R}_{(f)}^{(i)})$. It follows that $(\operatorname{vect}(\mathbf{u} \otimes \mathscr{R}_{(f)}^{(i)}), \mathbf{v})$ is a pSVT of $\mathbf{M}_{\mathcal{A}}$ with singular value one. The tuple $(\mathbf{u}, \mathbf{v})$ is a pSVT of $\mathbf{M}$ with singular value one, by definition of $\mathbf{M}$.

Let $\mathcal{B}$ be the span of the left singular vectors of $\mathbf{M}$ with singular value one. Suppose that $\operatorname{vect}(\mathscr{T}') \in \mathcal{B}$ is another rank-one tensor. Then $\mathscr{T}' \otimes \mathscr{R}_{(f)}^{(i)} \in \mathcal{A}$. Moreover, the r rank-one tensors $\{\mathscr{R}_{(f)}^{(i)}\}_{i=1}^{r}$ are the only rank-one elements in $\mathcal{A}$ (up to scale), since $r \leq r(f)$. It remains to exclude the possibility that $\mathscr{R}_{(f)}^{(i)}$ and $\mathscr{R}_{(f)}^{(j)}$ are collinear for some $i \neq j$. This cannot occur under the genericity assumption on the tensors $\{\mathscr{R}_{(d)}^{(i)}\}_{i=1}^{r}$, which ensures that their corresponding factors are pairwise linearly independent. □

COROLLARY 6.3. *The largest rank for which one can uniquely recover a partially symmetric decomposition of a generic tensor* $\mathscr{T} \in \bigotimes_{i=1}^{\ell} S^{d_i}(\mathbb{R}^{m_i})$ *via finding pSVTs and matching up factors is*

(1) with a flattening f *with all* $f_i > 0$*:*

$$r_{\max}(f) = r(f);$$

(2) with a symmetry–breaking flattening f*:*

$$r_{\max}(f) = \min\{\, r(f),\ r(d-f)\,\} = \min\{\, n_{\text{codim}}^{f},\ n_{\text{codim}}^{d-f}\,\};$$

(3) with a pair f, f' *that are not symmetry breaking and satisfy* (20)*:*

$$r_{\max}(f,f') = \begin{cases} \min\{\, r(f),\ r(f')\,\} & \text{if } \min\{f,f'\} > 0, \\ \min\{\, r(f),\ r(f'),\ r(d-f),\,\}, & \text{elsewhere.} \end{cases}$$

PROOF. For unique decomposition of $\mathscr{T}$, the largest rank r need to satisfy $r \le r(g)$ for any flattening tuple g used either to extract rank-one tensors via pSVTs as in Theorem 3.5 or to recover complementary factors as in Algorithm 3. The formulae then follow directly from Theorem 4.2 and Theorem 6.2. □

DEFINITION 6.4. A tuple f or a pair of tuples f, f' (when f is not symmetry-breaking) is called *optimal* if it attains the maximal recovery rank, i.e.,

$$r_{\max}(f) \quad \text{or} \quad r_{\max}(f,f')$$

is maximal over all flattenings.

We obtain formulae for the optimal flattening for $(\ell,1)$–symmetric tensors with $\ell = 2,3,4$.

COROLLARY 6.5. *Consider a partially symmetric tensor in* $S^{\ell}(\mathbb{R}^n) \otimes \mathbb{R}^k$*. The optimal flattening tuple depends on* (n,k) *as follows. For* $\ell = 2$ *and* 3*, the optimal tuple is* $f = (1,1)$*. For* $\ell = 4$*, the optimal tuple is*

$$f = \begin{cases} (2,1), & k \le \dfrac{n^2+3n-2}{2n-2}, \\ (1,1), & k > \dfrac{n^2+3n-2}{2n-2}. \end{cases}$$

PROOF. We restrict to flattenings $(k,1)$, since all other tuples need to pair with some $(k,1)$ tuple, which yields smaller $r_{\max}$, by Corollary 6.3. For $\ell = 2$, the only tuple with format $(*,1)$ is $(1,1)$, which is completable. Thus, $(1,1)$ is optimal. For $\ell = 3$, we have

- $r(1,1) = \begin{cases} \min\{\binom{n+1}{2}, (n-1)(k-1)\}, & \min\{n,k\} > 2, \\ \min\{3,k\}, & n = 2, \\ n, & k = 2. \end{cases}$
- $r(2,1) = \min\left\{n, \binom{n+1}{2}k - n - k + 1\right\} = n.$

Since $r(1,1) \ge r(2,1)$ in all cases, the tuple $(1,1)$ is optimal.

For $\ell = 4$, the three candidates are $(1,1), (2,1)$ and $(3,1)$. They yield

- $r(1,1) = \begin{cases} \min\left\{\binom{n+2}{3},\ (n-1)(k-1)\right\}, & n > 2, \\ \min\{4,\, k\}, & n = 2, \\ \min\{4,\, n\}, & k = 2. \end{cases}$

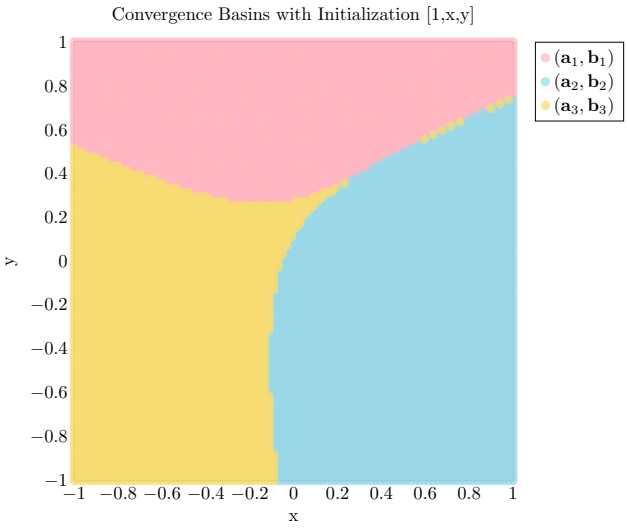


FIGURE 5. Basin of attraction for pSVTs of $\mathscr{T}^{(f)}$ under random initialization. Each point is an initialization vector $\mathbf{c} = [1, x, y]^\mathsf{T}$. It is colored according to which rank-one component $(\mathbf{a}_i, \mathbf{b}_i)$ PS-HOPM converges to. The three pSVTs are all attractors, converged to with comparable frequency.

- $r(2,1) = \min\left\{\binom{n+1}{2}, \binom{n+1}{2}k + 1 - n - k\right\} = \binom{n+1}{2}$
- $r(3,1) = \min\left\{n, \binom{n+2}{3}k + 1 - n - k\right\} = n$

We have $r(3,1) \le r(2,1)$, so $(3,1)$ cannot be optimal. When $n = 2$ or $k = 2$, $(2,1)$ dominates. For $n > 2, k > 2$, $(1,1)$ becomes optimal once $(n-1)(k-1) > \binom{n+1}{2}$, i.e.

$$k > \frac{n^2 + 3n - 2}{2n - 2}.$$

□

We have seen how to find the rest of a rank-one summand in a decomposition. We now consider how to find the remaining terms of the decomposition. In principle, all rank-one summands may be converged to from some initialization. In practice, one would have to initialize many times to obtain all r components. This is sped up via deflation: once a rank-one tensor is found, subtract its summand from $\mathscr{T}$ to decrease $\dim \mathcal{A}$ by one.

EXAMPLE 6.6 (Basins of attraction of the rank-one summands). Let

$$\mathscr{T} = \sum_{i=1}^{3} \mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}_i \in \mathbb{R}^{3\times3\times3}, \qquad \|\mathbf{a}_i\| = \|\mathbf{b}_i\| = \|\mathbf{c}_i\| = 1.$$

Fix $f = (1,1,0)$. Let $\mathscr{T}^{(f)} \in \mathbb{R}^{3\times3\times3}$ be the stack of an orthonormal basis of $\mathcal{A} = \operatorname{colspan}(\mathbf{T}^{(f)})$. Each $(\mathbf{a}_i, \mathbf{b}_i)$ is part of a pSVT of $\mathscr{T}^{(f)}$ with singular value 1, by Theorem 3.5.

We run PS-HOPM (Algorithm 1) on $\mathscr{T}^{(f)}$ with $\mathbf{c}$ initialized to $[1, x, y]^\mathsf{T}/\|(1, x, y)\|$ and $\mathbf{a}, \mathbf{b}$ to the top singular vector pair of $\mathscr{T}^{(f)}(*, *, \mathbf{c})$. For 10000 pairs $(x, y) \in [-1, 1]^2$, every run converges to a pSVT containing some $(\mathbf{a}_i, \mathbf{b}_i)$. The basins of attraction are in Figure 5.

If $\mathscr{T}$ has rank r, then $\mathscr{T}_\mathcal{A}$ has r pSVTs with singular value one. Assuming basin of attraction of comparable size, the classical coupon collection problem [**16**, pp. 224] implies that the expected number of random initializations required to recover all r pSVTs is about $r\left(1 + \frac{1}{2} + \cdots + \frac{1}{r}\right) \approx r \log r$.

Let $\mathscr{T} = \sum_{i=1}^{r} \lambda_i \bigotimes_{j=1}^{\ell} (\mathbf{v}_i^{(j)})^{\otimes d_j}$, where $\lambda_i \in \mathbb{R}$, $\mathbf{v}_i^{(j)} \in \mathbb{S}^{m_j - 1}$, and fix a tuple $f = (f_1, \ldots, f_\ell)$. Assume $r \le r_{\max}(f)$, so $\operatorname{colspan}(\mathbf{T}^{(f)})$ contains exactly the r rank-one tensors $\bigotimes_{j=1}^{\ell} (\mathbf{v}_i^{(j)})^{\otimes f_j}$. Suppose we have identified a rank-one tensor $\mathscr{R}_{(f)} := \bigotimes_{i=1}^{\ell} (\mathbf{v}_1^{(i)})^{\otimes f_i}$ and its complement $\mathscr{R}_{(d-f)} := \bigotimes_{i=1}^{\ell} (\mathbf{v}_1^{(i)})^{\otimes (d_i - f_i)}$. There is a unique λ_1 such that the rank of

$$\mathbf{T}^{(f)} - \lambda_1 \operatorname{vect}(\mathscr{R}_{(f)}) \otimes \operatorname{vect}(\mathscr{R}_{(d-f)})$$

drops by one, by the Wedderburn rank reduction formula [**56**]. This λ_1 is the coefficient of $\mathscr{R}^{(1)} := \bigotimes_{i=1}^{\ell} (\mathbf{v}_1^{(i)})^{\otimes d_i}$ in $\mathscr{T}$. Thus deflation removes $\mathscr{T}^{(1)}$ and reduces the problem size by one. The following gives scalar λ_1 and an orthonormal basis of the deflated linear space. The proof follows the same argument as [**29**, Proposition 3.5].

Proposition 6.7. *Fix*

$$\mathscr{T} = \sum_{i=1}^{r} \lambda_i \bigotimes_{j=1}^{\ell} (\mathbf{v}_i^{(j)})^{\otimes d_j}, \quad f = (f_1, \ldots, f_\ell),\ 0 \le f_i \le d_i.$$

Let $\mathbf{T}^{(f)}$ denote the f-flattening of $\mathscr{T}$ with

$$\mathbf{T}^{(f)} = \mathbf{U}\mathbf{C}^{-1}\mathbf{V}^{\mathsf{T}},$$

where $\mathbf{U} \in \mathbb{R}^{(\prod_{i=1}^{\ell} m_i^{f_i}) \times r}, \mathbf{V} \in \mathbb{R}^{(\prod_{i=1}^{\ell} m_i^{(d_i - f_i)}) \times r}$ have orthonormal columns and $\mathbf{C} \in \mathbb{R}^{r \times r}$ is nonsingular. Define $\mathscr{T}_{(f)} := \bigotimes_{i=1}^{\ell} (\mathbf{v}_1^{(i)})^{\otimes f_i}$ and $\mathscr{T}_{(d-f)} := \bigotimes_{i=1}^{\ell} (\mathbf{v}_1^{(i)})^{\otimes (d_i - f_i)}$. Then,

(1) The coefficient of $\mathscr{R}^{(1)} = \bigotimes_{i=1}^{\ell} (\mathbf{v}_1^{(i)})^{\otimes d_i}$ in $\mathscr{T}$ is

$$\lambda_1 = \frac{1}{\mathbf{a}_v^{\mathsf{T}} \mathbf{C}\, \mathbf{a}_u}, \qquad \mathbf{a}_u := \mathbf{U}^{\mathsf{T}} \operatorname{vect}(\mathscr{R}_{(f)}), \quad \mathbf{a}_v := \mathbf{V}^{\mathsf{T}} \operatorname{vect}(\mathscr{R}_{(d-f)}).$$

(2) Let $\mathbf{O}_u, \mathbf{O}_v \in \mathbb{R}^{r \times (r-1)}$ have orthonormal columns spanning $\operatorname{span}\{\mathbf{C}\mathbf{a}_u\}^{\perp}$ and $\operatorname{span}\{\mathbf{C}\mathbf{a}_v\}^{\perp}$, respectively. Define

$$\hat{\mathbf{U}} := \mathbf{U}\mathbf{O}_v, \qquad \hat{\mathbf{C}} := \mathbf{O}_u^{\mathsf{T}} \mathbf{C} \mathbf{O}_v, \qquad \hat{\mathbf{V}} := \mathbf{V}\mathbf{O}_u.$$

Then

(a) $\hat{\mathbf{U}}, \hat{\mathbf{V}}$ have orthonormal columns and $\hat{\mathbf{C}}$ is nonsingular;

(b) $\hat{\mathbf{U}}\hat{\mathbf{C}}^{-1}\hat{\mathbf{V}}^{\mathsf{T}} = \mathbf{T}^{(f)} - \lambda_1 \operatorname{vect}(\mathscr{R}_{(f)}) \otimes \operatorname{vect}(\mathscr{R}_{(d-f)})$;

(c) The columns of $\hat{\mathbf{U}}$ form an orthonormal basis of $\{\operatorname{vect}\Big(\bigotimes_{j=1}^{\ell} (\mathbf{v}_i^{(j)})^{\otimes f_j}\Big) : i \ge 2\}$

7. Numerical experiments

Our Multi-Subspace Power Method algorithm (Algorithm 4, below) for CP decomposition of partially symmetric tensors combines subspace extraction, power iteration, and deflation. We flatten the input tensor along a well-chosen tuple f, extract the column span of the flattening, and iteratively recover rank-one components via PS-HOPM (Algorithm 1). After identifying each component, we find its complementary factors (Algorithm 3). We remove the recovered term using a deflation step. In this section, we compare our algorithm to state-of-the-art approaches to see that it is accurate and fast.

Algorithm 4 Multi-Subspace Power Method (MSPM)

Input: Tensor $\mathscr{T} \in \bigotimes_{j=1}^{\ell} S^{d_j}(\mathbb{R}^{m_j})$ with rank $r \le r_{\max}(f)$ (or $r \le r_{\max}(f, f')$) and flattening tuple $f = (f_1, \dots, f_\ell)$ and f' (if f is not symmetry-breaking). Shift parameters $\gamma_1, \dots, \gamma_\ell \ge 0$, tolerances $\mathtt{f}_{\text{tol}}$, $\mathtt{grad}_{\text{tol}} > 0$.

Output: Decomposition $\mathscr{T} = \sum_{i=1}^{r} \lambda_i \bigotimes_{j=1}^{\ell} (\mathbf{v}_i^{(j)})^{\otimes d_j}$

Compute flattening $\mathbf{T}^{(f)} \leftarrow \text{flatten}(\mathscr{T}, f)$ — SUBSPACE EXTRACTION
$(\mathbf{U}, \mathbf{D}, \mathbf{V}) \leftarrow \mathtt{svd}(\mathbf{T}^{(f)})$
Update $(\mathbf{U}, \mathbf{D}, \mathbf{V})$ by truncating to top r components
for $i = 1$ to r **do**

$\mathscr{T}^{(f)} \leftarrow$ reshape columns of $\mathbf{U}$ into tensors and stack — PS-HOPM
Use Algorithm 1 to find pSVT $(\mathbf{v}_i^{(1)}, \dots, \mathbf{v}_i^{(\ell)})$ of $\mathscr{T}^{(f)}$ with singular value one
$\mathscr{R}_{(i)}^{(f)} \leftarrow \bigotimes_{j=1}^{\ell} (\mathbf{v}_i^{(j)})^{\otimes f_j}$

if $f_i > 0$ for all $i = 1, \dots, \ell$ **then** — SUMMAND COMPLETION
Immediately obtain $\mathscr{R}_{(i)}^{(d-f)} = \bigotimes_{j=1}^{\ell} (\mathbf{v}_i^{(j)})^{\otimes (d_j - f_j)}$
else if f is symmetry breaking **then**
Use partially known factors and row span of $\mathbf{T}^{(f)}$ to obtain $\mathscr{R}_{(i)}^{(d-f)}$ (see Fig. 3)
else
Use flattenings f', $d - f$ to compute $\mathscr{R}_{(i)}^{(d-f)}$ (see Fig. 4)

$\mathbf{a}_u \leftarrow \mathbf{U}^{\mathsf{T}} \operatorname{vect}(\mathscr{R}_{(i)}^{(f)})$, $\quad \mathbf{a}_v \leftarrow \mathbf{V}^{\mathsf{T}} \operatorname{vect}(\mathscr{R}_{(i)}^{(d-f)})$ — DEFLATION
$\lambda_i \leftarrow (\mathbf{a}_v^{\mathsf{T}} \mathbf{C} \mathbf{a}_u)^{-1}$
if $i < r$ **then**
$\mathbf{O}_u \leftarrow$ orthonormal basis of $\operatorname{span}\{\mathbf{C}\alpha_u\}^{\perp}$
$\mathbf{O}_v \leftarrow$ orthonormal basis of $\operatorname{span}\{\mathbf{C}\alpha_v\}^{\perp}$
Update $(\mathbf{U}, \mathbf{C}, \mathbf{V}) \leftarrow (\mathbf{U}\mathbf{O}_v,\ \mathbf{O}_u^{\mathsf{T}}\mathbf{C}\mathbf{O}_v,\ \mathbf{V}\mathbf{O}_u)$

return $\{(\lambda_i, \mathbf{v}_i^{(1)}, \dots, \mathbf{v}_i^{(\ell)})\}_{i=1}^{r}$

7.1. (2,1) Symmetry. We generate the ground-truth tensor as

$$\mathscr{T}_{\text{true}} = \sum_{i=1}^{r} \lambda_i \mathbf{a}_i^{\otimes 2} \otimes \mathbf{b}_i \in S^2(\mathbb{R}^{100}) \otimes \mathbb{R}^{50}, \quad r = 80.$$

The vectors $\mathbf{b}_i$ are sampled with i.i.d. Gaussian entries and normalized to unit length. Tensors of this format appear in temporal networks [**19**]. We consider two settings for $\mathbf{a}_i$:

- Orthogonal case: $\mathbf{a}_i$ are taken as the orthogonalized columns of a random Gaussian matrix in $\mathbb{R}^{100\times 80}$, i.e. we sample $\mathbf{a}_i$ via the Harr measure of the Stiefel manifold.
- Generic case: $\mathbf{a}_i$ are sampled with i.i.d. Gaussian entries and normalized.

The weights are chosen as $\lambda_i = \exp(2u_i - 1)$ with $u_i \sim \mathcal{U}[0, 1]$. We then add Gaussian noise

$$\mathscr{T} = \mathscr{T}_{\text{true}} + \eta \cdot \mathscr{Z},$$

scaled so that $\|\eta \cdot \mathscr{Z}\| = 0.01 \|\mathscr{T}_{\text{true}}\|$, where $\mathscr{Z}$ has i.i.d. standard Gaussian entries.

7.1.1. *Orthogonal case.* We compare MSPM with nonlinear optimization algorithms (nonlinear least squares (NLS), unconstrained nonlinear optimization (MINF), alternating least squares (ALS)) implemented in tensorlab [**54**], simultaneous diagonalization methods (FFDIAG [**58**], Jacobi [**7**], Jennrich), and the generalized SVD approach HOGSVD [**42**]. Due to the orthogonality, we also test SVD applied to $\mathscr{T}$.reshape$(100, 5000)$, both directly and as initialization for NLS.

We compare the recovered $\mathbf{a}_i'$ with the true $\mathbf{a}_i$, since diagonalization methods only recover the $\mathbf{a}_i$. We match the recovered $\mathbf{a}_i'$ and the true $\mathbf{a}_i$ via a greedy approach: we match $\mathbf{a}_1$ with the vector among the $\mathbf{a}_i'$s with the largest absolute cosine similarity, flip its sign if necessary, then proceed to $\mathbf{a}_2$ with the remaining columns, and so on until all are matched. The *Ascore* is the mean cosine similarity between the matched pairs

$$\text{Ascore}(\{\mathbf{a}_i\}_{i=1}^r, \{\mathbf{a}_i'\}_{i=1}^r) = \frac{1}{r}\sum_{i=1}^r |\langle \mathbf{a}_i, \mathbf{a}_i' \rangle|.$$

We plot log-runtime against log(1-Ascore) across the 10 algorithms in Figure 6 for 40 randomly generated noisy tensors. Our MSPM method achieves both high accuracy and fast runtime. It appears in dark blue in the bottom left of the plot.

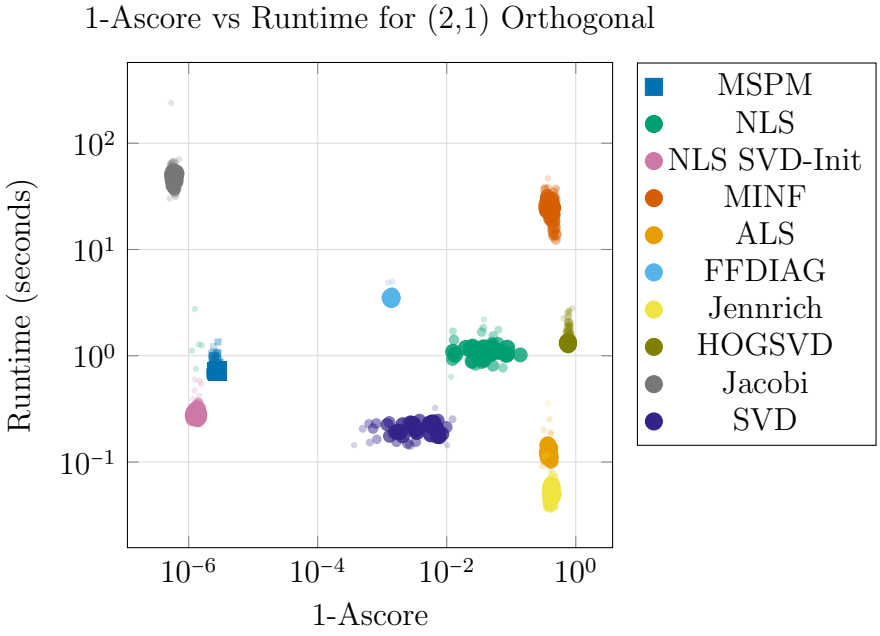


FIGURE 6. Comparison of methods to decompose a $(2,1)$-partially symmetric tensor in terms of accuracy and runtime. The ground truth tensor is $\mathscr{T}_{\text{true}} = \sum_{i=1}^{80} \lambda_i \mathbf{a}_i^{\otimes 2} \otimes \mathbf{b}_i \in S^2(\mathbb{R}^{100}) \times \mathbb{R}^{50}$ where $\mathbf{a}_i$ are orthogonal. Larger, more opaque markers indicate runs that occurred with higher frequency across 40 trials.

7.1.2. *Generic Case.* When $\mathbf{a}_i$ are drawn from normalized Gaussian vectors without orthogonalization, the problem is more challenging. We compare MSPM with nonlinear optimization algorithms (NLS, MINF, ALS) in tensorlab, simultaneous diagonalization methods (FFDIAG, QRJ1D [**4**], Jennrich), and the generalized SVD approach HOGSVD.

Accuracy is measured by Ascore, as above. Figure 7 reports log-runtime versus log(1-Ascore) across methods for 40 randomly generated noisy tensors. MSPM is the fastest

accurate method. The only exceptions are few runs of NLS (green): the left cluster of green points outperforms MSPM in accuracy, while the right cluster is less accurate.

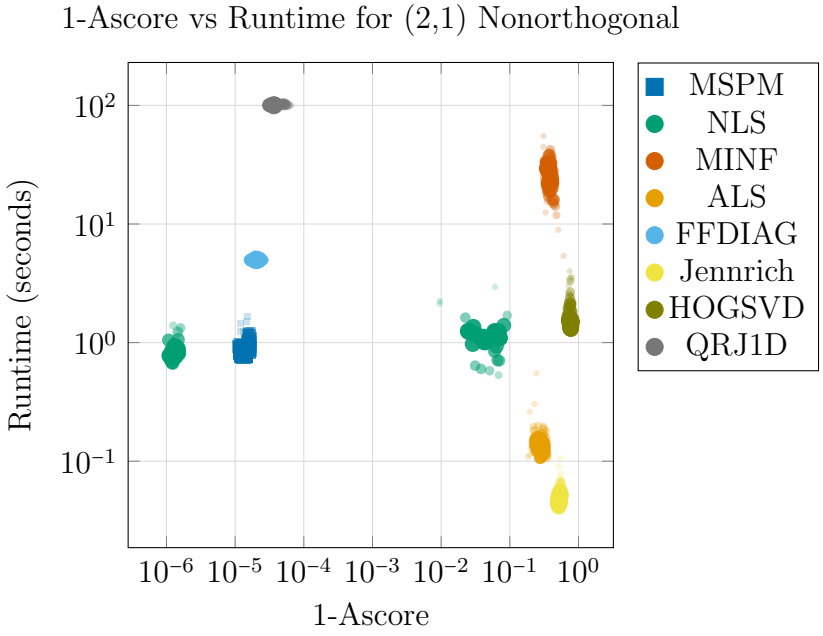


FIGURE 7. Comparison of methods for decomposing a $(2,1)$-partially symmetric tensor in terms of accuracy and runtime. The ground truth tensor is $\mathscr{T}_{\text{true}} = \sum_{i=1}^{80} \lambda_i \mathbf{a}_i^{\otimes 2} \otimes \mathbf{b}_i \in S^2(\mathbb{R}^{100}) \times \mathbb{R}^{50}$ where $\mathbf{a}_i$ are generic Gaussian vectors normalized to unit length.

7.2. (4,1) Symmetry. We generate the ground-truth low-rank tensor as

$$\mathscr{T}_{\text{true}} = \sum_{i=1}^{r} \lambda_i \mathbf{a}_i^{\otimes 4} \otimes \mathbf{b}_i \in S^4(\mathbb{R}^{25}) \otimes \mathbb{R}^{10}, \quad r = 50,$$

where $\mathbf{a}_i$ and $\mathbf{b}_i$ are sampled with i.i.d. Gaussian entries and normalized to unit norm. These format of tensors appear in computational algebraic geometry, data analysis and machine learning [**28, 55, 52**]. The weights are chosen as $\lambda_i = \exp(2u_i - 1)$ for $u_i \sim \mathcal{U}[0,1]$. We add Gaussian noise

$$\mathscr{T} = \mathscr{T}_{\text{true}} + \eta \cdot \mathscr{Z},$$

scaled so that $\|\eta \cdot \mathscr{Z}\| = 0.01 \|\mathscr{T}_{\text{true}}\|$, where $\mathscr{Z}$ has i.i.d. standard Gaussian entries subject to $(4,1)$ symmetry. For PS-HOPM, we use the $(2,1)$ flattening.

We compare MSPM with optimization-based algorithms NLS, MINF, and ALS. In addition, we apply MSPM, NLS and MINF ignoring the partial symmetry, then restore it by averaging the learned factors. We refer to these variants as asym-avg. We evaluate the accuracy of each algorithm by the reconstruction error

$$\|\mathscr{T} - \sum_{i=1}^{80} \lambda_i' \mathbf{a}_i'^{\otimes 2} \otimes \mathbf{b}_i'\|,$$

where $\mathbf{a}_i', \mathbf{b}_i', \lambda_i'$ are the recovered factors and their coefficients.

Figure 8 plots the runtime and reconstruction error in log-log scale for all seven methods over 40 noisy synthetic tensors. Across all trials in the (4,1) case, MSPM is the most accurate and the fastest. The asym-avg variants reach similar accuracy but require longer runtimes. Enforcing symmetry within MSPM provides a 1.3× speedup.

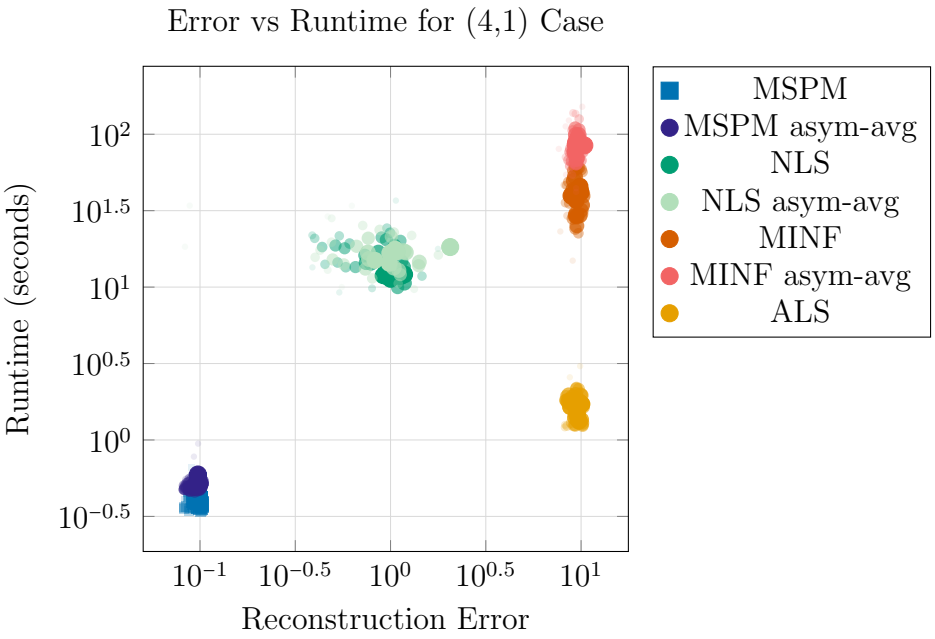


FIGURE 8. Reconstruction error versus runtime for seven decomposition algorithms on $(4,1)$ partially symmetric tensors. Each point is one run on a synthetic noisy tensor of fixed size and rank.

7.3. Order three CP decomposition. Three-way tensors arise in a range of scientific and engineering domains [**23, 31, 40**]. We generate a ground-truth tensor $\mathscr{T}_{\text{true}} \in (\mathbb{R}^{100})^{\otimes 3}$ with rank $r = 80$:

$$\mathscr{T}_{\text{true}} = \sum_{i=1}^{r} \lambda_i \, \mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}_i, \quad r = 80$$

where the $\mathbf{a}_i, \mathbf{b}_i, \mathbf{c}_i$ are independently sampled from standard Gaussian distributions and normalized to unit norm. The weights are sampled as $\lambda_i = \exp(2u_i - 1)$ with $u_i \sim \mathcal{U}[0,1]$. We add Gaussian noise

$$\mathscr{T} = \mathscr{T}_{\text{true}} + \eta \cdot \mathscr{Z},$$

scaled so that $\|\eta \cdot \mathscr{Z}\| = 0.01\|\mathscr{T}_{\text{true}}\|$, where $\mathscr{Z}$ has i.i.d. standard Gaussian entries.

We compare MSPM with six alternatives: NLS, MINF, ALS, Simulteneous diagonalization (SD) [**10**], Simultaneous generalized Schur decomposition (SGSD) [**12**] and Jennrich's Algorithm. We plot the runtime and reconstruction error of the seven algorithms in log-log scale for 40 synthetic random noisy tensors generated as above in Figure 9. MSPM achieves the fastest runtime among accurate methods.

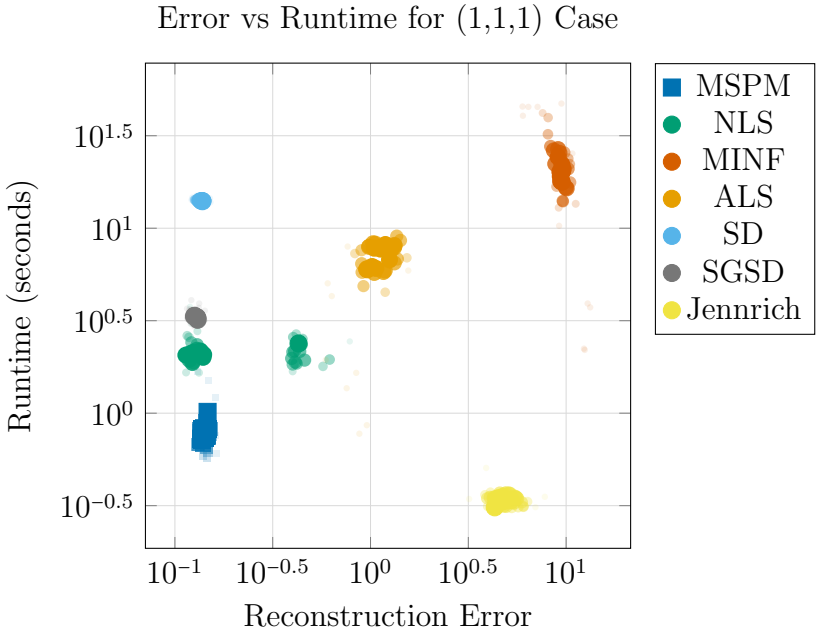


FIGURE 9. Comparison of decomposition methods for asymmetric tensors in $\mathbb{R}^{100\times100\times100}$ with rank 80. Each point is a single run.

8. Conclusion

We introduced the *Multi-Subspace Power Method (MSPM)*, an algorithm for decomposing tensors with arbitrary partial symmetries. Our algorithm transforms an input tensor into one with orthonormal slices. We showed that such tensors have rank decompositions whose summands are in one-to-one correspondence with their partially symmetric singular vector tuples (pSVTs) of singular value one. Further, we developed a *Partially Symmetric Higher-Order Power Method (PS-HOPM)* to compute pSVTs and proved its global convergence to first-order critical points as well as local linear convergence in the two-block case to second-order critical points. Finally, we demonstrated that MSPM is a competitive algorithm compared to prior methods, through numerical experiments on noisy tensors with symmetry types appearing in applications.

Several open questions remain. One direction is to determine whether the local linear convergence result in Proposition 5.6 can be extended to partially symmetric tensors with more than two blocks. Another question is whether the relationship between pSVTs and rank-one summands extends to tensors of higher rank, potentially revealing deeper algebraic or geometric structure in partially symmetric tensor decompositions.

Acknowledgments. A.S. was supported in part by the Sloan Foundation. J.P. was supported in part by a start-up grant from the University of Georgia. J.K. was supported in part by NSF DMS 2309782, NSF DMS 2436499, NSF CISE-IIS 2312746, DE SC0025312, and the Sloan Foundation.

References

[1] Abubakar Abid, Martin J Zhang, Vivek K Bagaria, and James Zou. Exploring patterns enriched in a dataset with contrastive principal component analysis. *Nature Communications*, 9(1):2134, 2018.

[2] Hirotachi Abo. On non-defectivity of certain Segre–Veronese varieties. *Journal of Symbolic Computation*, 45(12):1254–1269, 2010.
[3] Hirotachi Abo, Maria Chiara Brambilla, Francesco Galuppi, and Alessandro Oneto. Non-defectivity of Segre–Veronese varieties. *Proceedings of the American Mathematical Society, Series B*, 11(51):589–602, 2024.
[4] Bijan Afsari. Simple LU and QR based non-orthogonal matrix joint diagonalization. In *International Conference on Independent Component Analysis and Signal Separation*, pages 1–7. Springer, 2006.
[5] Animashree Anandkumar, Rong Ge, Daniel J Hsu, Sham M Kakade, Matus Telgarsky, et al. Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, 15(1):2773–2832, 2014.
[6] Paul Breiding and Sascha Timme. Homotopycontinuation.jl: A package for homotopy continuation in julia. In *International Congress on Mathematical Software*, pages 458–465. Springer, 2018.
[7] Jean-François Cardoso and Antoine Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 17(1):161–164, 1996.
[8] Luca Chiantini and Ciro Ciliberto. Weakly defective varieties. *Transactions of the American Mathematical Society*, 354(1):151–178, 2002.
[9] Pierre Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.
[10] Lieven De Lathauwer. A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 28(3):642–666, 2006.
[11] Lieven De Lathauwer, Pierre Comon, Bart De Moor, and Joos Vandewalle. Higher-order power method. *Nonlinear Theory and its Applications, NOLTA95*, 1(4), 1995.
[12] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition. *SIAM Journal on Matrix Analysis and Applications*, 26(2):295–327, 2004.
[13] Vin De Silva and Lek-Heng Lim. Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1084–1127, 2008.
[14] Andrew C Doherty, Pablo A Parrilo, and Federico M Spedalieri. Distinguishing separable and entangled states. *Physical Review Letters*, 88(18):187904, 2002.
[15] Jan Draisma, Giorgio Ottaviani, and Alicia Tocino. Best rank-k approximations for tensors: generalizing Eckart–Young. *Research in the Mathematical Sciences*, 5(2):27, 2018.
[16] William Feller. *An Introduction to Probability Theory and its Applications*, volume 2. John Wiley & Sons, 1991.
[17] Gerald B Folland. *Quantum Field Theory: A Tourist Guide for Mathematicians*, volume 149. American Mathematical Soc., 2021.
[18] Shmuel Friedland and Giorgio Ottaviani. The number of singular vector tuples and uniqueness of best rank-one approximation of tensors. *Foundations of Computational Mathematics*, 14(6):1209–1242, 2014.
[19] Laetitia Gauvin, André Panisson, and Ciro Cattuto. Detecting the community structure and activity patterns of temporal networks: A non-negative tensor factorization approach. *PloS One*, 9(1):e86028, 2014.
[20] Rong Ge and Tengyu Ma. On the optimization landscape of tensor decompositions. *Advances in Neural Information Processing Systems*, 30, 2017.
[21] Deren Han, Hui-Hui Dai, and Liqun Qi. Conditions for strong ellipticity of anisotropic elastic materials. *Journal of Elasticity*, 97(1):1–13, 2009.
[22] Joe Harris. *Algebraic Geometry: A First Course*, volume 133. Springer Science & Business Media, 2013.
[23] Richard A Harshman et al. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working papers in Phonetics*, 16(1):84, 1970.
[24] Johan Håstad. Tensor rank is NP-complete. *Journal of Algorithms*, 11(4):644–654, 1990.
[25] Christopher J Hillar and Lek-Heng Lim. Most tensor problems are NP-hard. *Journal of the ACM (JACM)*, 60(6):1–39, 2013.
[26] Roger A Horn and Charles R Johnson. *Matrix Analysis*. Cambridge University Press, 2012.
[27] Emil Horobeţ and Ettore Teixeira Turatti. When does subtracting a rank-one approximation decrease tensor rank? *Linear Algebra and its Applications*, 709:397–415, 2025.

[28] Nathaniel Johnston, Benjamin Lovitz, and Aravindan Vijayaraghavan. Computing linear sections of varieties: quantum entanglement, tensor decompositions and beyond. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1316–1336. IEEE, 2023.
[29] Joe Kileel and João M Pereira. Subspace power method for symmetric tensor decomposition. *Numerical Algorithms*, pages 1–38, 2025.
[30] Eleftherios Kofidis and Phillip A Regalia. On the best rank-1 approximation of higher-order supersymmetric tensors. *SIAM Journal on Matrix Analysis and Applications*, 23(3):863–884, 2002.
[31] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.
[32] Tamara G Kolda and Jackson R Mayo. Shifted power method for computing tensor eigenpairs. *SIAM Journal on Matrix Analysis and Applications*, 32(4):1095–1124, 2011.
[33] Joseph B Kruskal. Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear algebra and its applications*, 18(2):95–138, 1977.
[34] Joseph M Landsberg. *Geometry and Complexity Theory*, volume 169. Cambridge University Press, 2017.
[35] Suhua Li, Chaoqian Li, and Yaotang Li. M-eigenvalue inclusion intervals for a fourth-order partially symmetric tensor. *Journal of Computational and Applied Mathematics*, 356:391–401, 2019.
[36] Lek-Heng Lim. Singular values and eigenvalues of tensors: a variational approach. In *1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2005.*, pages 129–132. IEEE, 2005.
[37] Stanislaw Łojasiewicz. Ensembles semi-analytiques. *IHES Notes*, page 220, 1965.
[38] Daniel Miao, Gilad Lerman, and Joe Kileel. Tensor-based synchronization and the low-rankness of the block trifocal tensor. *Advances in Neural Information Processing Systems*, 37:69505–69532, 2024.
[39] Jie Ni, Yuhong Dai, and Zheng Peng. An alternating algorithm for structure preserving CP-decompositions of partially symmetric tensors. *Computational Optimization and Applications*, pages 1–33, 2025.
[40] Maximilian Nickel, Volker Tresp, Hans-Peter Kriegel, et al. A three-way model for collective learning on multi-relational data. In *International Conference on Machine Learning*, volume 11, pages 3104482–3104584, 2011.
[41] João M Pereira, Joe Kileel, and Tamara G Kolda. Tensor moments of Gaussian mixture models: Theory and applications. *arXiv preprint arXiv:2202.06930*, 2022.
[42] Sri Priya Ponnapalli, Michael A Saunders, Charles F Van Loan, and Orly Alter. A higher-order generalized singular value decomposition for comparison of global mRNA expression from multiple organisms. *PloS One*, 6(12):e28072, 2011.
[43] Kristian Ranestad, Anna Seigal, and Kexin Wang. A real generalized trisecant trichotomy. *arXiv preprint arXiv:2409.01356*, 2024.
[44] Phillip A Regalia and Eleftherios Kofidis. Monotonic convergence of fixed-point algorithms for ICA. *IEEE Transactions on Neural Networks*, 14(4):943–949, 2003.
[45] Werner C Rheinboldt. *Methods for Solving Systems of Nonlinear Equations*. SIAM, 1998.
[46] Alvaro Ribot, Emil Horobet, Anna Seigal, and Ettore Teixeira Turatti. Decomposing tensors via rank-one approximations. *arXiv:2411.15935*, 2024.
[47] Elina Robeva. Orthogonal decomposition of symmetric tensors. *SIAM Journal on Matrix Analysis and Applications*, 37(1):86–102, 2016.
[48] Elina Robeva and Anna Seigal. Singular vectors of orthogonally decomposable tensors. *Linear and Multilinear Algebra*, 65(12):2457–2471, 2017.
[49] Reinhold Schneider and André Uschmajew. Convergence results for projected line-search methods on varieties of low-rank matrices via Łojasiewicz inequality. *SIAM Journal on Optimization*, 25(1):622–646, 2015.
[50] Alwin Stegeman and Pierre Comon. Subtracting a best rank-1 approximation may increase tensor rank. *Linear Algebra and its Applications*, 433(7):1276–1300, 2010.
[51] André Uschmajew. A new convergence proof for the higher-order power method and generalizations. *arXiv preprint arXiv:1407.4586*, 2014.
[52] Konstantin Usevich, Clara Dérand, Ricardo Borsoi, and Marianne Clausel. Identifiability of deep polynomial neural networks. *arXiv preprint arXiv:2506.17093*, 2025.

[53] Nick Vannieuwenhoven, Johannes Nicaise, Raf Vandebril, and Karl Meerbergen. On generic nonexistence of the Schmidt–Eckart–Young decomposition for complex tensors. *SIAM Journal on Matrix Analysis and Applications*, 35(3):886–903, 2014.
[54] Nico Vervliet, Otto Debals, Laurent Sorber, Marc Van Barel, and Lieven De Lathauwer. Tensorlab 3.0, Mar. 2016. Available online.
[55] Kexin Wang, Aida Maraj, and Anna Seigal. Contrastive independent component analysis. *arXiv preprint arXiv:2407.02357*, 2024.
[56] Joseph Henry Maclagan Wedderburn. *Lectures on Matrices*, volume 17. American Mathematical Soc., 1934.
[57] Yifan Zhang and Joe Kileel. Moment estimation for nonparametric mixture models through implicit tensor decomposition. *SIAM Journal on Mathematics of Data Science*, 5(4):1130–1159, 2023.
[58] Andreas Ziehe, Pavel Laskov, Guido Nolte, and Klaus-Robert Müller. A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation. *Journal of Machine Learning Research*, 5(Jul):777–800, 2004.

Appendix A. Proofs in Section 5

LEMMA A.1. *Fix a tensor $\mathscr{T}^{(f)} \in \bigotimes_{i=1}^{\ell} S^{f_i}(\mathbb{R}^{m_i}) \otimes \mathbb{R}^r$ whose last slices $\mathscr{S}_1, \dots, \mathscr{S}_r$ are orthonormal with span $\mathcal{A}$.*

(1) The tuple $(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)}) \in \mathbb{S}^{m_1-1} \times \cdots \times \mathbb{S}^{m_\ell-1}$ is a pSVT of $\widetilde{\mathscr{T}}^{(f)}$ with singular value σ^2 if and only if $(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)}, \mathbf{w})$ is a pSVT of $\mathscr{T}^{(f)}$ with singular value σ, where $\mathbf{w}$ is the vector $\mathscr{T}^{(f)} \cdot \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes f_i}$, rescaled to be norm one.
(2) The critical points of $F_{\mathcal{A}}$ are the pSVTs of $\widetilde{\mathscr{T}}^{(f)}$. The function $F_{\mathcal{A}}$ takes values in $[0, 1]$ with 1 attained when $\bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes f_i} \in \mathcal{A}$.

PROOF. (1) We establish the correspondence between the pSVTs of $\widetilde{\mathscr{T}}^{(f)}$ and $\mathscr{T}^{(f)}$. Fix $k \in \{1, \dots, \ell\}$. For a tuple $(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)})$, we compute

$$\begin{aligned}
\widetilde{\mathscr{T}}^{(f)} \cdot \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes (2f_i - \delta_{i,k})} &= \sum_{j=1}^{r} \left\langle \mathscr{S}_j, \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes f_i} \right\rangle \cdot \left(\mathscr{S}_j \cdot \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes (f_i - \delta_{i,k})} \right) \\
&= \left\| \mathscr{T}^{(f)} \cdot \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes f_i} \right\| \sum_{j=1}^{r} \mathbf{w}_j \cdot \left(\mathscr{S}_j \cdot \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes (f_i - \delta_{i,k})} \right) \\
&= \left\| \mathscr{T}^{(f)} \cdot \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes f_i} \right\| \mathscr{T}^{(f)} \cdot \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes (f_i - \delta_{i,k})} \otimes \mathbf{w}.
\end{aligned}$$

If $(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)})$ is a pSVT of $\widetilde{\mathscr{T}}^{(f)}$ with singular value σ^2, then $\widetilde{\mathscr{T}}^{(f)} \cdot \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes (2f_i - \delta_{i,k})} = \sigma^2 \mathbf{v}^{(k)}$. Equivalently,

$$\mathscr{T}^{(f)} \cdot \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes (f_i - \delta_{i,k})} \otimes \mathbf{w} = \sigma \mathbf{v}^{(k)},$$

since $\left\| \mathscr{T}^{(f)} \cdot \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes f_i} \right\| = \left(\widetilde{\mathscr{T}}^{(f)} \cdot \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes 2f_i} \right)^{1/2} = \sigma$. The definition of $\mathbf{w}$ also says that $\mathscr{T}^{(f)} \cdot \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes f_i} = \sigma \mathbf{w}$. Hence $(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)}, \mathbf{w})$ satisfies the pSVT equations for $\mathscr{T}^{(f)}$ with singular value σ.

(2) The contraction $\widetilde{\mathscr{T}}^{(f)} \cdot \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes 2f_i}$ can be rewritten as

$$\begin{aligned}\sum_{j=1}^{r} \left\langle P_{\mathrm{sym}}(\mathscr{S}_j \otimes \mathscr{S}_j), \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes 2f_i} \right\rangle &= \sum_{j=1}^{r} \left\langle \mathscr{S}_j \otimes \mathscr{S}_j, P_{\mathrm{sym}}\left(\bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes 2f_i} \right) \right\rangle \\ &= \sum_{j=1}^{r} \left\langle \mathscr{S}_j, \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes f_i} \right\rangle^2 \\ &= \left\| P_{\mathcal{A}} \left(\bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes f_i} \right) \right\|^2 = F_{\mathcal{A}}(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(\ell)}).\end{aligned}$$

The critical points of the function

$$(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(\ell)}) \;\mapsto\; \widetilde{\mathscr{T}}^{(f)} \cdot \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes 2f_i} = F_{\mathcal{A}}(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(\ell)})$$

are the pSVTs of $\widetilde{\mathscr{T}}^{(f)}$, by [**36**]. We have $F_{\mathcal{A}}(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(\ell)}) \in [0,1]$, by definition of $F_{\mathcal{A}}$ and $\| \bigotimes_{j=1}^{\ell} (\mathbf{v}^{(j)})^{\otimes f_j} \| = 1$. It follows from (1) that the global maxima of $F_{\mathcal{A}}$ are the pSVTs of $\mathscr{T}^{(f)}$ with singular value one. The tensor $\bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes f_i}$ lies in $\mathcal{A}$, by Theorem 3.5. □

Proof of Proposition 5.4. We show that the shifts make

$$F(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(\ell)}) := \left\langle \widetilde{\mathscr{T}}^{(f)}, \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes 2f_i} \right\rangle + \sum_{i=1}^{\ell} \gamma_i \, \|\mathbf{v}^{(i)}\|^{2f_i}$$

convex in each block. Global convergence of Algorithm 1 then follows from Theorem 5.1. Consider block i. If $f_i = 1$, then the i-th block of F is a quadratic form hence convex, so $\gamma_i = 0$ suffices. This proves (a). When $f_i > 1$, Lemma A.1 gives the identity

$$F_{\mathcal{A}}(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(\ell)}) = \left\langle \widetilde{\mathscr{T}}^{(f)}, \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes 2f_i} \right\rangle = \sum_{i=1}^{r} \langle \mathscr{S}_i, \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes f_i} \rangle^2.$$

Its Hessian is

$$\nabla_i^2 F_{\mathcal{A}} = 2f_i^2 \sum_{i=1}^{r} \Big(\mathscr{S}_i \cdot \bigotimes_{j=1}^{\ell} (\mathbf{v}^{(j)})^{\otimes (f_i - \delta_{ij})} \Big)^{\otimes 2} + 2f_i(f_i - 1) \sum_{i=1}^{r} \langle \mathscr{S}_i, \bigotimes_{i=1}^{\ell} (\mathbf{v}^{(i)})^{\otimes f_i} \rangle \mathscr{S}_i \cdot \bigotimes_{j=1}^{\ell} (\mathbf{v}^{(j)})^{\otimes (f_i - 2\delta_{ij})}.$$

It follows that

$$\nabla_i^2 F = \nabla_i^2 F_{\mathcal{A}} + \nabla_i^2 \sum_{i=1}^{\ell} \gamma_i \|\mathbf{v}^{(i)}\|^{2f_i} = \nabla_i^2 F_{\mathcal{A}} + 4f_i(f_i - 1)\gamma_i \|\mathbf{v}^{(i)}\|^{2f_i - 4} (\mathbf{v}^{(i)})^{\otimes 2} + 2f_i \gamma_i \|\mathbf{v}^{(i)}\|^{2f_i - 2} I.$$

Let $\mathscr{S}_{j\setminus i}(\mathbf{v})$ denote $\mathscr{S}_j \cdot \bigotimes_{j\neq i}(\mathbf{v}^{(j)})^{\otimes f_j}$. For any $\mathbf{y} \in \mathbb{S}^{m_i-1}$, we obtain

$$\begin{aligned}
\frac{1}{2f_i}\,\mathbf{y}^\mathsf{T}\nabla_i^2 F(\mathbf{v}^{(1)},\dots,\mathbf{v}^{(\ell)})\,\mathbf{y} &= f_i \sum_{j=1}^{r} \Big\langle \mathscr{S}_{j\setminus i}(\mathbf{v}),\ (\mathbf{v}^{(i)})^{\otimes f_i-1}\otimes \mathbf{y}\Big\rangle^2 \\
&\quad + (f_i-1)\sum_{j=1}^{r}\langle \mathscr{S}_i, \bigotimes_{i=1}^{\ell}(\mathbf{v}^{(i)})^{\otimes f_i}\rangle\langle \mathscr{S}_{j\setminus i}(\mathbf{v}),\ (\mathbf{v}^{(i)})^{\otimes f_i-2}\otimes \mathbf{y}^{\otimes 2}\rangle \\
&\quad + 2(f_i-1)\gamma_i\langle \mathbf{v}^{(i)},\mathbf{y}\rangle^2 + \gamma_i \\
&\geq (f_i-1)\sum_{j=1}^{r}\langle \mathscr{S}_i, \bigotimes_{i=1}^{\ell}(\mathbf{v}^{(i)})^{\otimes f_i}\rangle\langle \mathscr{S}_{j\setminus i}(\mathbf{v}),\ (\mathbf{v}^{(i)})^{\otimes f_i-2}\otimes \mathbf{y}^{\otimes 2}\rangle + \gamma_i \\
&= (f_i-1)\langle P_{\mathcal{A}}(\bigotimes_{i=1}^{\ell}(\mathbf{v}^{(i)})^{\otimes f_i})\cdot \bigotimes_{j\neq i}(\mathbf{v}^{(j)})^{\otimes f_j}, (\mathbf{v}^{(i)})^{\otimes f_i-2}\otimes \mathbf{y}^{\otimes 2}\rangle + \gamma_i.
\end{aligned}$$

We note that

$$\langle P_{\mathcal{A}}(\bigotimes_{i=1}^{\ell}(\mathbf{v}^{(i)})^{\otimes f_i})\cdot \bigotimes_{j\neq i}(\mathbf{v}^{(j)})^{\otimes f_j}, (\mathbf{v}^{(i)})^{\otimes f_i}\rangle = \|P_{\mathcal{A}}(\bigotimes_{i=1}^{\ell}(\mathbf{v}^{(i)})^{\otimes f_i})\|^2.$$

We write the vector $\mathbf{y}$ as $\mathbf{y} = \alpha\mathbf{x} + \beta\bar{\mathbf{y}}$, where $\bar{\mathbf{y}} \perp \mathbf{x}$, and use the same argument as for [**29**, Lemma 4.4] to obtain

$$\frac{1}{2f_i}\,\mathbf{y}^\mathsf{T}\nabla_i^2 F(\mathbf{v})\,\mathbf{y} \ \geq\ -\sqrt{\tfrac{f_i-1}{f_i}}\,h(\nu) + \gamma_i,$$

where $h(\nu)$ is as defined in the statement. □