

SOLVING MULTILINEAR SYSTEMS VIA TENSOR INVERSION*

M. BRAZELL[†], N. LI[‡], C. NAVASCA[§], AND C. TAMON[¶]

Abstract. Higher order tensor inversion is possible for even order. This is due to the fact that a tensor group endowed with the contracted product is isomorphic to the general linear group of degree n . With these isomorphic group structures, we derive a tensor SVD which we have shown to be equivalent to well-known canonical polyadic decomposition and multilinear SVD provided that some constraints are satisfied. Moreover, within this group structure framework, multilinear systems are derived and solved for problems of high-dimensional PDEs and large discrete quantum models. We also address multilinear systems which do not fit the framework in the least-squares sense. These are cases when there is an odd number of modes or when each mode has distinct dimension. Numerically we solve multilinear systems using iterative techniques, namely, biconjugate gradient and Jacobi methods.

Key words. tensor and matrix inversions, multilinear system, tensor decomposition, least-squares method polynomial

AMS subject classifications. 15A69, 15A18, 65F15, 65F99

DOI. 10.1137/100804577

1. Introduction. Tensor decompositions have been successfully applied across many fields which include, among others, chemometrics [54, 48, 6], signal processing [10, 15], and computer vision [53]. More recent applications are in large-scale PDEs through a reduced rank representation of operators with applications to quantum chemistry [34] and aerospace engineering [21]. State-of-the-art tensor methods have been applied to problems in quantum chemistry: Khoromskij, Khoromskaia, and Flad [35] solved the fundamental Hartree–Fock equation and Beylkin and Mohlenkamp [3, 4] worked on multidimensional operators in quantum models. Hackbusch and Khoromskij [26] and Hackbusch, Khoromskij, and Tyrtysnikov [27] have solved multidimensional boundary and eigenvalue problems using a reduced low-dimensional tensor-product space through separated representation and hierarchical Kronecker tensor from the underlying high spatial dimensions. See the survey papers [15, 34, 36] and the references therein for more applications and tensor-based methods. Extensive studies (e.g., [11, 14, 16, 37]) have exposed many aspects of the differences between tensors and matrices despite the fact that tensors are multidimensional generalizations of matrices.

In this paper, we continue to investigate the relationship between matrices and tensors. Here we address the following question: when is it possible to matricize (*tensorize*) and apply matrix-(tensor) based methods to high-dimensional problems and data with inherent tensor (matrix) structure. Specifically, we address tensor inversion through group theoretic structures and by providing numerical methods for specific multilinear systems in quantum mechanical models and high-dimensional

*Received by the editors August 5, 2010; accepted for publication (in revised form) by B. Hendrickson February 20, 2013; published electronically May 14, 2013. The second and third authors were both supported in part by National Science Foundation grant DMS-0915100.

<http://www.siam.org/journals/simax/34-2/80457.html>

[†]Department of Mechanical Engineering, University of Wyoming, Laramie, WY 82071 (michaeljbrazell@gmail.com).

[‡]Department of Mathematics, Clarkson University, Potsdam, NY 13699 (nali@clarkson.edu).

[§]Corresponding author. Department of Mathematics, University of Alabama at Birmingham, Birmingham, AL 35294 (cnavasca@uab.edu).

[¶]Department of Computer Science, Clarkson University, Potsdam, NY 13699 (tino@clarkson.edu).

PDEs. Since the inversion of tensor impinges upon a tensor-tensor multiplication definition, the contracted product for tensor multiplication was chosen since it provides a natural setting for multilinear systems and high-dimensional eigenvalue problems considered here. It is also an intrinsic extension of the matrix product rule. Still other choices of multiplication rules should be considered for particular application in hand. For example, in the matrix case, there is the alternative multiplication of Strassen [50] which improves the computational complexity by using block structure format and the optimized matrix multiplication based on blocking for improving cache performance by Demmel [20]. In a recent work of Ragnarsson and Van Loan [44, 45], the idea of blocking is extended to tensors. In the work of Braman [5], an alternative tensor multiplication is used in image processing applications. Our choice of the standard canonical tensor-tensor multiplication provides a useful setting for algorithms for decompositions, inversions, and multilinear iterative solvers.

Associated with tensors are multilinear systems. Multilinear systems model many phenomena in engineering and science. For example, in continuum physics and engineering, isotropic and anisotropic elasticity are modelled [41] as multilinear systems. Approximating solutions to PDEs in high dimensions amounts to solving multilinear systems. Current tensor-based methods for solving PDEs require a reduction of the spatial dimensions and some applications of tensor decomposition techniques; here we focus on tensor iterative methods for solving high-dimensional Poisson problems in the multilinear system framework.

Tensor representations are also common in large discrete quantum models like the discrete Schrödinger and Anderson models. The Anderson model¹ [1] is the most studied model for spectral and transport properties of an electron in a disordered medium. The study of spectral theory of the Anderson model is a very active research topic, but there are still many open problems and conjectures for high-dimensional $d \geq 3$ cases; see [31, 38, 49] and the references therein.

Powerful computer simulations and mathematical modeling are increasingly used as visualization tools for understanding crystal structure and evolution. In addition, numerical (multi)linear algebra techniques are becoming useful tools in understanding complicated models and difficult problems in quantum statistical mechanics. For example, Bai et al. [2] have developed numerical linear algebra methods for the many-electrons Hubbard model and quantum Monte Carlo simulations.

We develop a tensor-based visualization tool for localization and for verifying some conjectures in high dimension ($d \geq 3$) for all disorder λ and at various distributions. The Hamiltonians of the discrete Schrödinger and Anderson models are even order tensors which satisfy the symmetry requirement in the tensor SVD that we describe in section 3. Moreover, the numerical results provide some validation that these localizations exist for large disorder for dimension $d > 1$ for a sufficient amount of atoms; see the conjectures in [31, 38, 49].

The contributions of this paper are three-fold. First, we define the tensor group which provides the framework for formulating multilinear systems and tensor inversion. Second, we discuss tensor decompositions derived from the isomorphic group structure and show that they are special cases of the well-known canonical polyadic (CP) decomposition [8, 28] and multilinear SVD [51, 52, 16] provided that some conditions are satisfied. These decompositions appear in many signal processing applications, e.g., see [10, 18] and the references therein. Moreover, we describe multilinear

¹P.W. Anderson received the Nobel Prize in Physics in 1977 for his work on the spectral and transport properties of an electron in a disordered medium.

systems in PDEs and quantum models while providing numerical methods for solving multilinear systems. Multilinear systems which do not fit in the framework are addressed by pseudoinversion methods.

2. Preliminaries. We denote the scalars in \mathbb{R} with lowercase letters (a, b, \dots) and the vectors with bold lowercase letters $(\mathbf{a}, \mathbf{b}, \dots)$. The matrices are written as bold uppercase letters $(\mathbf{A}, \mathbf{B}, \dots)$ and the symbol for tensors are calligraphic letters $(\mathcal{A}, \mathcal{B}, \dots)$. The subscripts represent the scalars $(\mathcal{A})_{ijk} = a_{ijk}$, $(\mathbf{A})_{ij} = a_{ij}$, $(\mathbf{a})_i = a_i$ unless noted otherwise. The superscripts indicate the length of the vector or the size of the matrices. For example, \mathbf{b}^K is a vector with length K and $\mathbf{B}^{N \times K}$ is a $N \times K$ matrix.

The order of a tensor refers to the cardinality of the index set. A matrix is a second order tensor and a vector is a first order tensor.

DEFINITION 2.1 (even and odd tensors). *Given an N th tensor $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$. If N is even (odd), then \mathcal{T} is an even (odd) N th order tensor.*

DEFINITION 2.2 (Einstein product [22]). *For any N , the Einstein product is defined by the operation $*_N$ via*

$$(2.1) \quad (\mathcal{A} *_N \mathcal{B})_{i_1 \dots i_N k_{N+1} \dots k_M} = \sum_{k_1 \dots k_N} a_{i_1 i_2 \dots i_N k_1 \dots k_N} b_{k_1 \dots k_N k_{N+1} k_{N+2} \dots k_M},$$

where $\mathcal{A} \in \mathbb{R}^{I_1 \times \dots \times I_N \times K_1 \times \dots \times K_N}$ and $\mathcal{B} \in \mathbb{R}^{K_1 \times \dots \times K_N \times K_{N+1} \times \dots \times K_M}$.

For example, if $\mathcal{T}, \mathcal{S} \in \mathbb{R}^{I \times J \times I \times J}$, the operation $*_2$ is defined by the following:

$$(2.2) \quad (\mathcal{T} *_2 \mathcal{S})_{ij\hat{i}\hat{j}} = \sum_{u=1}^I \sum_{v=1}^J t_{ijuv} s_{uv\hat{i}\hat{j}}.$$

The Einstein product is a contracted product that it is widely used in the area of continuum mechanics [41] and in the study of the theory of relativity [22]. Notice that the Einstein product $*_1$ is the standard matrix multiplication since

$$(2.3) \quad (\mathbf{M} *_1 \mathbf{N})_{ij} = \sum_{k=1}^K m_{ik} n_{kj} = (\mathbf{MN})_{ij}$$

for $\mathbf{M} \in \mathbb{R}^{I \times K}$, $\mathbf{N} \in \mathbb{R}^{K \times J}$.

DEFINITION 2.3 (Tucker mode- n product). *Given a tensor $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$ and some matrices $\mathbf{A} \in \mathbb{R}^{\hat{I} \times I}$, $\mathbf{B} \in \mathbb{R}^{\hat{J} \times J}$, and $\mathbf{C} \in \mathbb{R}^{\hat{K} \times K}$, the Tucker mode- n products are the following:*

$$\begin{aligned} (\mathcal{T} \bullet_1 \mathbf{A})_{\hat{i},j,k} &= \sum_{i=1}^I t_{ijk} a_{\hat{i}i} \quad \forall \hat{i}, j, k \quad (\text{mode-1 product}), \\ (\mathcal{T} \bullet_2 \mathbf{B})_{\hat{j},i,k} &= \sum_{j=1}^J t_{ijk} b_{\hat{j}j} \quad \forall \hat{j}, i, k \quad (\text{mode-2 product}), \\ (\mathcal{T} \bullet_3 \mathbf{C})_{\hat{k},i,j} &= \sum_{k=1}^K t_{ijk} c_{\hat{k}k} \quad \forall \hat{k}, i, j \quad (\text{mode-3 product}). \end{aligned}$$

Notice that the Tucker product \bullet_n is the Einstein product $*_1$ when the mode summation is specified.

Below we define matrix and tensor blocks. These blockings have been described in the papers of Ragnarsson and Van Loan [44, 45].

DEFINITION 2.4 (partition of indices). *Let I be the index set with n partitions and $|I|$ cardinality. Then there exist I_k subindex sets for $k = 1, \dots, n$ such that $\sum_k |I_k| = |I|$. Collectively, (n_1, n_2, \dots, n_m) denotes the number of partitions on multiple index sets I_1, I_2, \dots, I_m .*

DEFINITION 2.5 (matrix blocks). *Given a matrix \mathbf{A} of size $I \times J$ with partition (k, l) on I and J , respectively, a matrix $\mathbf{A} \in \mathbb{R}^{I \times J}$ is partitioned into blocks of kl matrices of size $k \times l$. If the partition is $(I, 1)$, then matrix $\mathbf{A} \in \mathbb{R}^{I \times J}$ is partitioned into blocks of $1 \times J$ row vectors of size J , denoted by $\mathbf{a}_k^{(1)} = \mathbf{A}(k, :) \in \mathbb{R}^J$ for $k = 1, \dots, I$. Similarly, if the partition is $(1, J)$, then matrix $\mathbf{A} \in \mathbb{R}^{I \times J}$ is partitioned into blocks of $I \times 1$ column vectors of size I , denoted by $\mathbf{a}_l^{(2)} = \mathbf{A}(:, l) \in \mathbb{R}^I$ for $l = 1, \dots, J$.*

DEFINITION 2.6 (tensor blocks of third order tensor). *Given a tensor \mathcal{A} of size $I \times J \times K$ with partition (l, m, n) on I, J , and K , a third order tensor $\mathcal{A} \in \mathbb{R}^{I \times J \times K}$ is partitioned into $l \cdot m \cdot n$ number of tensor blocks of size $l \times m \times n$.*

Here are some special cases. If the partition is $(I, 1, 1)$, then $\mathcal{A} \in \mathbb{R}^{I \times J \times K}$ is partitioned into matrix blocks of size $J \times K$. Each matrix block (top-bottom matrix slice) is denoted by $\mathbf{A}_i^{(1)} = \mathcal{A}(i, :, :) \in \mathbb{R}^{J \times K}$ for $i = 1, \dots, I$. If the partition is $(1, J, 1)$, then $\mathcal{A} \in \mathbb{R}^{I \times J \times K}$ is partitioned into matrix blocks of size $I \times K$. Each matrix block (left-right matrix slice) is denoted by $\mathbf{A}_j^{(2)} = \mathcal{A}(:, j, :) \in \mathbb{R}^{I \times K}$ for $j = 1, \dots, J$. If the partition is $(1, 1, K)$, then $\mathcal{A} \in \mathbb{R}^{I \times J \times K}$ is partitioned into matrix blocks of size $I \times K$. Each matrix block (front-back matrix slice) is denoted by $\mathbf{A}_k^{(3)} = \mathcal{S}(:, :, k) \in \mathbb{R}^{I \times J}$ for $k = 1, \dots, K$. Moreover, the standard *flattenings* of third order tensor are concatenations of the matrix slices forming the mode-one, mode-two, and mode-three matricizations [16]: $\mathbf{A}^{(1)} = [(\mathbf{A}_1^{(1)})^T \ (\mathbf{A}_2^{(1)})^T \ \dots \ (\mathbf{A}_I^{(1)})^T] \in \mathbb{R}^{K \times I \cdot J}$, $\mathbf{A}^{(2)} = [\mathbf{A}_1^{(2)} \ \mathbf{A}_2^{(2)} \ \dots \ \mathbf{A}_J^{(2)}] \in \mathbb{R}^{I \times J \cdot K}$, and $\mathbf{A}^{(3)} = [(\mathbf{A}_1^{(3)})^T \ (\mathbf{A}_2^{(3)})^T \ \dots \ (\mathbf{A}_K^{(3)})^T] \in \mathbb{R}^{J \times K \cdot I}$.

DEFINITION 2.7 (tensor blocks of N th order tensor). *Given a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ with $N \geq 4$ with partition (J_1, J_2, \dots, J_N) on the index sets I_1, I_2, \dots, I_N , a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is partitioned into $j_1 j_2 \dots j_N$ number of tensor blocks of size $J_1 \times J_2 \times \dots \times J_N$.*

Here we discuss some relevant tensor blocks in this paper. Given a fourth order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$ with partition $(1, 1, I_3, I_4)$, $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$ is partitioned into matrix blocks of size $I_1 \times I_2$. Each block is denoted by $\mathbf{A}_{i_3, i_4}^{(3,4)} = \mathcal{A}(:, :, i_3, i_4) \in \mathbb{R}^{I_1 \times I_2}$ with $i_3 = 1, \dots, I_3$ and $i_4 = 1, \dots, I_4$. Similarly, given a fourth order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$ with partition $(I_1, I_2, 1, 1)$, $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$ is partitioned into matrix blocks $\mathbf{A}_{i_1, i_2}^{(1,2)}$ of size $I_3 \times I_4$. See Figure 2.1. Now given a sixth order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4 \times I_5 \times I_6}$ with partition $(1, I_2, 1, I_4, 1, I_6)$, $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4 \times I_5 \times I_6}$ is partitioned into tensor blocks $\mathcal{A}_{i_1 i_3, i_5}^{(2,4,6)}$ of size $I_1 \times I_3 \times I_5$. See Figure 2.2.

2.1. Tensor group set-up. Recall that the group of invertible $N \times N$ matrices is called the general linear group of degree n , denoted $GL(n, \mathbb{R})$, where the binary operation is the matrix multiplication. Let's denote this group $\mathbb{M}_{N,N}(\mathbb{R})$ consisting of all invertible $N \times N$ matrices with the matrix multiplication.

DEFINITION 2.8 (transformation). *Define the transformation $f : \mathbb{T}_{N,M,N,M}(\mathbb{R}) \rightarrow \mathbb{M}_{N \cdot M, N \cdot M}(\mathbb{R})$ with $f(\mathcal{A}) = \mathbf{A}$ defined component-wise as*

$$(2.4) \quad (\mathcal{A})_{nmnm} \xrightarrow{f} (\mathbf{A})_{[n+(m-1)N][n+(m-1)N]},$$

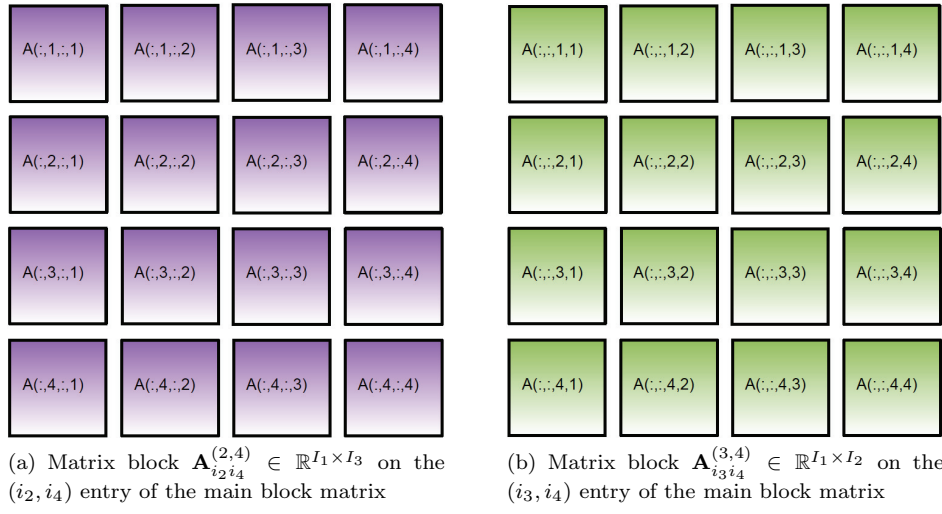


FIG. 2.1. Fourth order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$ with $I_1 = I_2 = I_3 = I_4 = 4$ with index partitioning: $(1, I_2, 1, I_4)$ (left) and $(1, 1, I_3, I_4)$ (right).

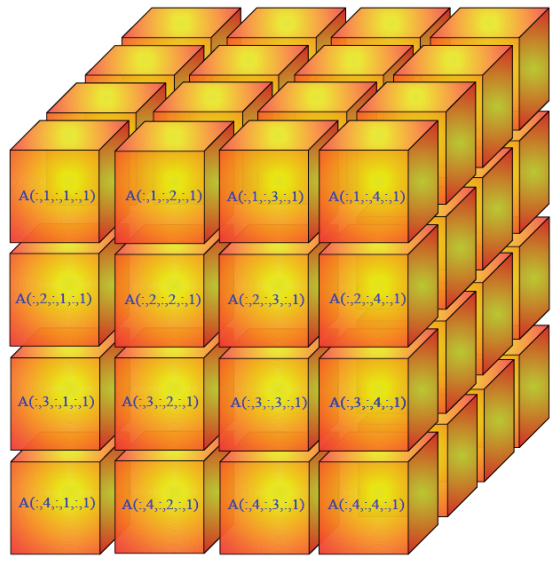


FIG. 2.2. Sixth order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4 \times I_5 \times I_6}$ with index partitioning $(1, I_2, 1, I_4, 1, I_6)$ into tensor blocks $\mathcal{A}_{i_1 i_3 i_5}^{(2,4,6)}$ of size $I_1 \times I_3 \times I_5$. The tensor blocks $\mathcal{A}_{i_2 i_4 i_6}^{(2,4,6)}$ are arranged in a third order tensor structure on the (i_2, i_4, i_6) entry of the main block tensor.

where $\mathbb{T}_{N,M,N,M}(\mathbb{R}) = \{\mathcal{A} \in \mathbb{R}^{N \times M \times N \times M} : \det(f(\mathcal{A})) \neq 0\}$. In general, the transformation is defined as

$$(2.5) \quad (\mathcal{A})_{i_1 i_2 \dots i_n j_1 j_2 \dots j_n} \xrightarrow{f} (\mathbf{A})_{[i_1 + \sum_{k=2}^n (i_k - 1) \prod_{l=1}^{k-1} I_l] [j_1 + \sum_{k=2}^n (j_k - 1) \prod_{l=1}^{k-1} J_l]}$$

when $\mathcal{A} \in \mathbb{T}_{I_1, \dots, I_n, J_1, \dots, J_n}(\mathbb{R})$ and $\mathbf{A} \in \mathbb{M}_{I_1 \cdot I_2 \dots I_{n-1} \cdot I_n, J_1 \cdot J_2 \dots J_{n-1} \cdot J_n}(\mathbb{R})$.

Remark 2.9. These transformations are known as column (row) major format in several computer languages which are typically used to enhance efficiency in accessing

arrays. This mapping is also the matrix unfolding of fourth order tensors in signal processing applications, e.g., see [18].

Next in section 3, we will show that $\mathbb{T}_{N,M,N,M}$ with the Einstein product (2.2) is a group and the transformation f is an isomorphism between $\mathbb{T}_{N,M,N,M}$ and $\mathbb{M}_{N \cdot M, N \cdot M}$. So even order tensor inverses exist through the mapping to the general linear group. Consequently, we will show that tensors with different mode dimensions have no inverses under the contraction product. These are tensor analogues to rectangular matrices. In addition, we also discuss the notion of pseudoinverses in multilinear systems for odd and even orders with distinct mode lengths.

2.2. Standard tensor decompositions. In 1927, Hitchcock [29, 30] introduced the idea that a tensor is decomposable into a sum of a finite number of rank-one tensors. Today, we refer to this decomposition as CP tensor decomposition (also known as CANDECOMP [8] or PARAFAC [28]). CP is a linear combination of rank-one tensors, i.e.,

$$(2.6) \quad \mathcal{T} = \sum_{r=1}^R a_r \circ b_r \circ c_r \circ d_r,$$

where $\mathcal{T} \in \mathbb{R}^{I \times J \times K \times L}$, $a_r \in \mathbb{R}^I$, $b_r \in \mathbb{R}^J$, $c_r \in \mathbb{R}^K$, and $d_r \in \mathbb{R}^L$. The column vectors a_r, b_r, c_r , and d_r form the so-called factor matrices \mathbf{A} , \mathbf{B} , \mathbf{C} , and \mathbf{D} , respectively. The tensorial rank [30] is the minimum $R \in \mathbb{N}$ such that \mathcal{T} can be expressed as a sum of R rank-one tensors. Moreover, in 1977 Kruskal [39] proved that for third order tensor,

$$2R + 2 \leq k_A(\mathbf{A}) + k_B(\mathbf{B}) + k_C(\mathbf{C})$$

is the sufficient condition for uniqueness of $\mathcal{T} = \sum_{r=1}^R a_r \circ b_r \circ c_r$ up to permutation and scalings where Kruskal's rank k_A is the maximum number r such that any set of r columns of \mathbf{A} is linearly independent. Kruskal's uniqueness condition was then generalized for $n \geq 3$ by Sidiropoulos and Bro [47]:

$$(2.7) \quad 2R + (n - 1) \leq \sum_{j=1}^n k_{A^{(j)}}(\mathbf{A}^{(j)})$$

for $\mathcal{T} = \sum_{r=1}^R a_r^{(1)} \circ \dots \circ a_r^{(n)}$.

Another decomposition called higher order SVD (also known as Tucker and multilinear SVD) was introduced by Tucker [51, 52] in which a tensor is decomposable into a core tensor *multiplied* (Definition 2.3) by a matrix along each mode, i.e.,

$$(2.8) \quad \mathcal{T} = \mathcal{S} \bullet_1 \mathbf{A} \bullet_2 \mathbf{B} \bullet_3 \mathbf{C} \bullet_4 \mathbf{D},$$

where $\mathcal{T}, \mathcal{S} \in \mathbb{R}^{I \times J \times K \times L}$ are fourth order tensors with four orthogonal factors $\mathbf{A} \in \mathbb{R}^{I \times I}$, $\mathbf{B} \in \mathbb{R}^{J \times J}$, $\mathbf{C} \in \mathbb{R}^{K \times K}$, and $\mathbf{D} \in \mathbb{R}^{L \times L}$. The Tucker decomposition is not unique, i.e., if there exist invertible matrices $\mathbf{R} \in \mathbb{R}^{I \times I}$, $\mathbf{S} \in \mathbb{R}^{J \times J}$, $\mathbf{T} \in \mathbb{R}^{K \times K}$, and $\mathbf{U} \in \mathbb{R}^{L \times L}$, then

$$\mathcal{T} = \mathcal{S} \bullet_1 \mathbf{A} \bullet_2 \mathbf{B} \bullet_3 \mathbf{C} \bullet_4 \mathbf{D} = \bar{\mathcal{S}} \bullet_1 \bar{\mathbf{A}} \bullet_2 \bar{\mathbf{B}} \bullet_3 \bar{\mathbf{C}} \bullet_4 \bar{\mathbf{D}},$$

where $\bar{\mathcal{S}} = \bullet_1 \mathbf{R}^{-1} \bullet_2 \mathbf{S}^{-1} \bullet_3 \mathbf{T}^{-1} \bullet_4 \mathbf{U}^{-1}$, $\bar{\mathbf{A}} = \mathbf{R}\mathbf{A}$, $\bar{\mathbf{B}} = \mathbf{S}\mathbf{B}$, $\bar{\mathbf{C}} = \mathbf{T}\mathbf{C}$, and $\bar{\mathbf{D}} = \mathbf{U}\mathbf{D}$.

The CP decomposition is a special case of the Tucker decomposition framework, i.e.,

$$(2.9) \quad \mathcal{T} = \sum_{r=1}^R a_r \circ b_r \circ c_r \circ d_r = \mathcal{S}_{CP} \bullet_1 \mathbf{A} \bullet_2 \mathbf{B} \bullet_3 \mathbf{C} \bullet_4 \mathbf{D},$$

where its core tensor $\mathcal{S}_{CP} \in \mathbb{R}^{I \times J \times K \times L}$ is *diagonal*, that is, the nonzero entries are located at $(\mathcal{S})_{iii}$ for $i = 1, \dots, R$ with $R = \min\{I, J, K, L\}$. Next, we discuss the connections of CP and Tucker to the tensor decompositions built from the isomorphic map.

3. Tensor group, multilinear systems and decompositions. For the sake of clarity, the main discussion in this section is limited to fourth order tensors, although the definitions and theorems presented are easily extended to the general case of even ordered tensors. Here we define a group structure on a set of fourth order tensor through a push-forward map on the general linear group. Also several consequential results from the group structure will be discussed.

LEMMA 3.1. *Let f be the map defined in (2.4). Then the following properties hold:*

1. *The map f is a bijection. Moreover, there exists a bijective inverse map $f^{-1} : \mathbb{M}_{I_1 I_2, I_1 I_2}(\mathbb{R}) \rightarrow \mathbb{T}_{I_1, I_2, I_1, I_2}(\mathbb{R})$.*
2. *The map satisfies $f(\mathcal{A} *_2 \mathcal{B}) = f(\mathcal{A}) \cdot f(\mathcal{B})$, where \cdot refers to the usual matrix multiplication.*

Proof. Here we denote $[n] = \{1, 2, \dots, n\}$ and its cardinality as $|n|$.

(1) According to the definition of f , we can define a map $h : [I_1] \times [I_2] \rightarrow [I_1 I_2]$ by $h(i_1, i_2) = i_1 + (i_2 - 1)I_1$. Clearly, the map h is a bijection since f is a bijection.

(2) Since f is a bijection, for some $1 \leq i, j \leq I_1 I_2$, there exist some unique indices i_1, i_2, j_1, j_2 for $1 \leq i_1, j_1 \leq I_1, 1 \leq i_2, j_2 \leq I_2$ such that $(i_2 - 1)I_1 + i_1 = i$ and $(j_2 - 1)I_1 + j_1 = j$. So,

$$[f(\mathcal{A} *_2 \mathcal{B})]_{ij} = (\mathcal{A} *_2 \mathcal{B})_{i_1 i_2 j_1 j_2} = \sum_{u,v} a_{i_1 i_2 u v} b_{u v j_1 j_2},$$

$$[f(\mathcal{A}) \cdot f(\mathcal{B})]_{ij} = \sum_{r=1}^{|I_1 I_2|} [f(\mathcal{A})]_{ir} [f(\mathcal{B})]_{rj}.$$

For every $1 \leq r \leq I_1 I_2$, there exist some unique u, v such that $(u - 1)I_1 + v = r$. So,

$$\sum_{u,v} a_{i_1 i_2 u v} b_{u v j_1 j_2} = \sum_{r=1}^{|I_1 I_2|} [f(\mathcal{A})]_{ir} [f(\mathcal{B})]_{rj}. \quad \square$$

It follows from the properties of f that the Einstein product (2.2) can be defined through the transformation:

$$(3.1) \quad \mathcal{A} *_2 \mathcal{B} = f^{-1}[f(\mathcal{A} *_2 \mathcal{B})] = f^{-1}[f(\mathcal{A}) \cdot f(\mathcal{B})].$$

Consequently, the inverse map f^{-1} satisfies

$$(3.2) \quad f^{-1}(\mathbf{A} \cdot \mathbf{B}) = f^{-1}(\mathbf{A}) *_2 f^{-1}(\mathbf{B}).$$

THEOREM 3.2. *Suppose (\mathbb{M}, \cdot) is a group. Let $f : \mathbb{T} \rightarrow \mathbb{M}$ be any bijection. Then we can define a group structure on \mathbb{T} by defining*

$$\mathcal{A} *_2 \mathcal{B} = f^{-1}[f(\mathcal{A}) \cdot f(\mathcal{B})]$$

for all $\mathcal{A}, \mathcal{B} \in \mathbb{T}$. In other words, the binary operation $*_2$ satisfies the group axioms. Moreover, the mapping f is an isomorphism.

The proof is straightforward; see the details in the appendix. As a matter of fact, the group structure can be formulated as a ring isomorphism.

COROLLARY 3.3. *Let $I_m = J_m$ for $m = 1, \dots, N$. Then the ordered pair $(\mathbb{T}_{I_1, \dots, I_N, J_1, \dots, J_N}(\mathbb{R}), *_N)$ is a group where the operation $*_N$ is the generalized Einstein product in (2.1).*

Proof. The generalization of the transformation f (2.5) on the set $\mathbb{T}_{I_1, \dots, I_N, J_1, \dots, J_N}(\mathbb{R})$ with the binary operation $*_N$ easily provides the extension for this case. \square

THEOREM 3.4. *The ordered pair $(\mathbb{T}_{I_1, I_2, \dots, I_{2N-1}}(\mathbb{R}), *_N)$ is not a group under the operation $*_N$.*

Proof. Take $N = 2$. Then $\mathcal{T} *_2 \mathcal{S} \notin \mathbb{T}_{I_1, I_2, \dots, I_{2N-1}}(\mathbb{R})$, where $\mathcal{T}, \mathcal{S} \in \mathbb{T}_{I_1, I_2, \dots, I_{2N-1}}(\mathbb{R})$. It follows that $\mathbb{T}_{I_1, I_2, \dots, I_{2N-1}}(\mathbb{R})$ is not closed under $*_2$. Thus the ordered pair $(\mathbb{T}_{I_1, I_2, \dots, I_{2N-1}}(\mathbb{R}), *_2)$ is not a group. \square

Theorem 3.4 implies that odd order tensors have no inverses with respect to the operation $*_N$, although such binary operation may exist in which the set of odd order tensors exhibits a group structure. Lemma 3.1 and Theorem 3.2 show that the transformation f (2.4) is an isomorphism between groups \mathbb{T} and \mathbb{M} . From Corollary 3.3, it follows that these structural properties are preserved for any ordered pair $(\mathbb{T}_{I_1, \dots, I_N, J_1, \dots, J_N}(\mathbb{R}), *_N)$ for any N .

3.1. Multilinear systems. A linear system is conveniently expressed as $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A} \in \mathbb{R}^{M \times N}$, $\mathbf{x} \in \mathbb{R}^N$, and $\mathbf{b} \in \mathbb{R}^M$. Recall that $\mathbf{A} \in \mathbb{R}^{M \times N}$ defines a linear transformation $\mathcal{L} : \mathbb{R}^N \rightarrow \mathbb{R}^M$ such that $\mathcal{L}(\mathbf{x}) = \mathbf{Ax}$. A bilinear system is defined through $\mathfrak{B} : \mathbb{R}^M \times \mathbb{R}^N \rightarrow \mathbb{R}$ with $\mathfrak{B}(\mathbf{x}, \mathbf{y}) = \mathbf{y}^T \mathbf{Bx} = \mathbf{B} \bullet_1 \mathbf{x} \bullet_2 \mathbf{y}$, where $\mathbf{B} \in \mathbb{R}^{M \times N}$. The bilinear map has the linearity properties

$$\begin{aligned} \mathfrak{B}(c\mathbf{x}_1 + d\mathbf{x}_2, \mathbf{y}) &= c\mathfrak{B}(\mathbf{x}_1, \mathbf{y}) + d\mathfrak{B}(\mathbf{x}_2, \mathbf{y}) \\ &\text{and} \\ \mathfrak{B}(\mathbf{x}, \bar{c}\mathbf{y}_1 + \bar{d}\mathbf{y}_2) &= \bar{c}\mathfrak{B}(\mathbf{x}, \mathbf{y}_1) + \bar{d}\mathfrak{B}(\mathbf{x}, \mathbf{y}_2) \end{aligned}$$

for some scalars c, \bar{c}, d, \bar{d} and vectors $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N, \mathbf{y}, \mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^M$. In general, we define multilinear transformations for the following systems:

1. $\mathcal{B} \bullet_1 \mathbf{x} \bullet_2 \mathbf{y} = b$, where $\mathcal{B} \in \mathbb{R}^{I \times J \times K}$, $\mathbf{x} \in \mathbb{R}^I$, $\mathbf{y} \in \mathbb{R}^J$, and $b \in \mathbb{R}$
2. $\mathcal{M} *_2 \mathbf{X} *_2 \mathbf{Y} = b$, where $\mathcal{M} \in \mathbb{R}^{I \times J \times K \times L}$, $\mathbf{X} \in \mathbb{R}^{K \times L}$, $\mathbf{Y} \in \mathbb{R}^{I \times J}$, and $b \in \mathbb{R}$
3. $\mathcal{M} *_2 \mathcal{X} *_3 \mathcal{Y} = \mathbf{B}$, where $\mathcal{M} \in \mathbb{R}^{I \times J \times K \times L \times M \times N}$, $\mathcal{X} \in \mathbb{R}^{M \times N \times O}$, $\mathcal{Y} \in \mathbb{R}^{K \times L \times O}$, and $\mathbf{B} \in \mathbb{R}^{I \times J}$.

See [25] for more discussions on multilinear transformation.

Multilinear systems model many phenomena in engineering and sciences. For example, in continuum physics and engineering, isotropic and anisotropic elastic models are multilinear systems [41] like

$$\mathcal{C} *_2 \mathbf{E} = \mathbf{T},$$

where \mathbf{T} and \mathbf{E} are second order tensors modeling stress and strain, respectively, and the fourth order tensor \mathcal{C} refers to the elasticity tensor. Multilinear systems are also prevalent in solving PDEs numerically. In section 4, we derive and solve multilinear systems in discretized Poisson problems and eigenvalue problems in quantum models by using tensor-based iterative methods. There are modern applications in emerging fields like system biology, e.g., sparse multilinear systems model interactions among genes and proteins in living cells [43].

| Higher-Order Biconjugate Gradient Method |
|---|
| Given $\mathcal{A} \in \mathbb{R}^{N \times N \times N \times N}, \mathcal{B}^{N \times N}, tol$ Initial guess $\mathcal{X}_0 \in \mathbb{R}^{N \times N}$ $\mathcal{A} \leftarrow \mathcal{A}^T *_2 \mathcal{A}$ $\mathcal{B} \leftarrow \mathcal{A}^T *_2 \mathcal{B}$ $\mathcal{X} = \mathcal{X}_0$ $\mathcal{R} = \mathcal{B} - \mathcal{A} *_2 \mathcal{X}$ %Residual $\mathcal{P} = \mathcal{R}$ %Search Direction while $\ \mathcal{R}\ > tol$ $r = \langle \mathcal{R}, \mathcal{R} \rangle$ $\alpha = \frac{r}{\langle \mathcal{P}, \mathcal{A} *_2 \mathcal{P} \rangle}$ %Compute Search Parameter $\mathcal{X} \leftarrow \mathcal{X} + \alpha \mathcal{P}$ %Update Solution $\mathcal{R} \leftarrow \mathcal{R} - \alpha \mathcal{A} *_2 \mathcal{P}$ %Update Residual $\beta = \frac{\langle \mathcal{R}, \mathcal{R} \rangle}{r}$ $\mathcal{P} \leftarrow \mathcal{R} + \beta \mathcal{P}$ %Compute Search Direction end |

(a) Higher Order Biconjugate Gradient

| Higher-Order Jacobi Method |
|---|
| Given $\mathcal{A} \in \mathbb{R}^{N \times N \times N \times N}, \mathcal{B}^{N \times N}, MAX$ Initial guess $\mathcal{X}^0 \in \mathbb{R}^{N \times N}$ for $k = 1$ to MAX for $i = 1$ to N for $j = 1$ to N $(\mathcal{X}^{(k+1)})_{ij} = \left[\mathcal{B}_{ij} - \sum_{\substack{i,j \\ i_j \neq i_j}} (\mathcal{A})_{ij i_j i_j} (\mathcal{X}^{(k)})_{i_j i_j} \right] / (\mathcal{A})_{ij ij}$ end end end |

(b) Higher Order Jacobi

FIG. 3.1. Pseudocodes for iterative solvers.

3.1.1. Solving multilinear systems. Our first approach for solving multilinear systems numerically is to use the Gauss–Newton algorithm for approximating \mathcal{A}^{-1} through the function

$$g(\mathcal{X}) = \mathcal{A} *_2 \mathcal{X} - \mathcal{I} = 0,$$

where \mathcal{I} is the identity fourth order tensor defined in (3.10) and tensor \mathcal{X} is unknown. This method is highly inefficient because calculating the inversion of a Jacobian is very expensive. To save memory and operational costs, we consider iterative methods for solving multilinear systems. The pseudocodes in Figure 3.1 describe the biconjugate gradient (BiCG) and the higher order Jacobi methods for solving multilinear system $\mathcal{A} *_2 \mathcal{X} = \mathcal{B}$. Note that the algorithms solely use tensor computations, i.e., no matricizations are involved. Recall that the BiCG method requires symmetric and positive definite matrix so that the multilinear system is premultiplied by its transpose \mathcal{A}^T which is later defined in Definition 3.5. The BiCG method solves the multilinear system by searching along $\mathcal{X}_k = \mathcal{X}_{k-1} + \alpha_{k-1} \mathcal{P}_{k-1}$ with a line parameter α_{k-1} and a search direction \mathcal{P}_{k-1} while minimizing the objective function $\phi(\mathcal{X}_k + \alpha_{k-1} \mathcal{P}_k)$, where $\phi(\mathcal{X}) = \frac{1}{2} \mathcal{X}^T *_2 \mathcal{A} *_2 \mathcal{X} - \mathcal{X}^T *_2 \mathcal{B}$. It follows that $\phi(\hat{\mathcal{X}})$ attains a minimum iteratively and precisely at an optimizer $\hat{\mathcal{X}}$, where $\mathcal{A} *_2 \hat{\mathcal{X}} = \mathcal{B}$. The higher order Jacobi method is also implemented for comparison. The Jacobi method for tensors is based on *split-*

ting the tensor \mathcal{A} into its diagonal entries from the lower and upper diagonal entries of the tensor.

We approximate the solution to the multilinear systems considered in section 4 using these two multilinear iterative methods; see Figure 4.2 for the pseudocodes of the algorithms. We also discuss the advantages of these methods.

3.1.2. Overdetermined multilinear systems. According to Theorem 3.4, odd order tensors have no inverses with respect to the operation $*_N$. *Nonrectangular* tensors also do not have inverses. Multilinear systems with tensors of odd order or that are nonrectangular (distinct mode sizes) are referred to as overdetermined multilinear systems, for example,

1. $\mathcal{A} \bullet_3 \mathbf{x} = \mathbf{B}$, where $\mathcal{A} \in \mathbb{R}^{I \times J \times K}$, $\mathbf{x} \in \mathbb{R}^K$, and $\mathbf{B} \in \mathbb{R}^{I \times J}$,
2. $\mathcal{A} * \mathcal{X} = \mathcal{B}$, where $\mathcal{A} \in \mathbb{R}^{I \times J \times S \times T}$, $\mathcal{X} \in \mathbb{R}^{S \times T \times K \times L}$, and $\mathcal{B} \in \mathbb{R}^{I \times J \times K \times L}$.

\mathcal{A}^{-1} does not exist for either case. We extend the concepts of pseudoinversion for odd order tensors and nonrectangular tensors. The optimization formulations,

$$(3.3) \quad \min_{\mathbf{x}} \|\mathcal{A} \bullet_3 \mathbf{x} - \mathbf{B}\|_F \quad \text{and} \quad \min_{\mathcal{X}} \|\mathcal{A} * \mathcal{X} - \mathcal{B}\|_F,$$

are considered to find *multilinear least-squares* solutions of the systems. Note that the Frobenius norm, $\|\cdot\|_F$, is defined as $\|\mathcal{A}\|_F^2 = \sum_{i_1 i_2 \dots i_N} |a_{i_1 i_2 \dots i_N}|^2$ for $\mathcal{A}^{I_1 \times I_2 \times \dots \times I_N}$.

The linear least-squares (LLS) method is a well-known method for overdetermined linear systems. Often the number of observations \mathbf{b} exceed the number of unknown parameters \mathbf{x} in LLS, forming an overdetermined system, e.g.,

$$(3.4) \quad \mathbf{A}\mathbf{x} = \mathbf{b},$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{b} \in \mathbb{R}^m$ with $m > n$. Through minimization of the residual,

$$\min_{\mathbf{x}} \|\mathbf{r}\|_{\ell_2},$$

where $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$, the overdetermined system (3.4) is solved. This is called the LLS method. The minimizer, the vector \mathbf{x}^* , is called the least-squares solution of the linear system (3.4).

We now describe tensor transposition to facilitate the discussions on the normal equations for multilinear systems.

DEFINITION 3.5 (transpose). *A transpose of $\mathcal{S} \in \mathbb{R}^{I \times J \times I \times J}$ is a tensor \mathcal{T} which has entries $t_{ijkl} = s_{klij}$. We denote the transpose of \mathcal{S} as $\mathcal{T} = \mathcal{S}^T$. If $\mathcal{S} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N \times J_1 \times \dots \times J_N}$, then $(\mathcal{T})_{i_1, i_2, \dots, i_n, j_1, j_2, \dots, j_n} = (\mathcal{S}^T)_{j_1, j_2, \dots, j_n, i_1, i_2, \dots, i_n}$ is the transpose of \mathcal{S} .*

More generally, tensor transposition can be described through the permutation of indices. Let permutation σ be defined as $\sigma(i_1 i_2 \dots i_n) = i_{m(1)} i_{m(2)} \dots i_{m(n)}$, where $m(j) \in \{1, 2, \dots, n\}$. Then the transpose of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ with respect to σ is

$$\mathcal{A}_\sigma^T = \mathcal{A}_{i_{m(1)} i_{m(2)} \dots i_{m(n)}}.$$

In Definition 3.5, $\mathcal{S}_\sigma^T = \mathcal{S}_{klij}$, where $\sigma(ijkl) = klij$. A transpose of a third order tensor $\mathcal{A} \in \mathbb{R}^{I \times J \times K}$ is $\mathcal{A}_\sigma^T = \mathcal{A}_{kij}$ with $\sigma(ijk) = kij$.

Recall that $(\mathbf{A}^T)^T = \mathbf{A}$ in the matrix case. In the following lemmas, we describe some similar properties for higher order tensors.

LEMMA 3.6. *Let $\mathcal{A} \in \mathbb{R}^{I \times J \times K}$ and $\rho = (\rho_1, \rho_2, \rho_3)$ be a cyclic permutation of length three on the index set $\{ijk\}$. Then $((\mathcal{A}_{\rho_1}^T)_{\rho_2}^T)_{\rho_3}^T = \mathcal{A}$.*

Proof. For the index set $\{ijk\}$, there are two cyclic permutations: $\rho_1(ijk) = jki$, $\rho_2(jki) = kij$, $\rho_3(kij) = ijk$ and $\bar{\rho}_1(ikj) = kji$, $\bar{\rho}_2(kji) = jik$, $\bar{\rho}_3(jik) = ikj$. It follows that $((\mathcal{A}_{\rho_1}^T)_{\rho_2}^T)_{\rho_3}^T = (\mathcal{A}_{kij})_{\rho_3}^T = \mathcal{A}_{ijk}$ since $\mathcal{A}_{\rho_1}^T = \mathcal{A}_{jki}$ and $(\mathcal{A}_{\rho_1}^T)_{\rho_2}^T = (\mathcal{A}_{jki})_{\rho_2}^T$. \square

Although there are six permutations on the index set $\{ijk\}$, there are only two set of cyclic permutations. For the fourth order tensor, there are twenty-four permutations on the index set $\{ijst\}$ and six cyclic permutations of length four.

LEMMA 3.7 (property of fourth order tensor transpose). *Let $\mathcal{A} \in \mathbb{R}^{I \times J \times S \times T}$ and ρ be a cyclic permutation $\rho = (\rho_1, \rho_2, \rho_3, \rho_4)$ of length four on the index set $\{ijst\}$. Then $((\mathcal{A}_{\rho_1}^T)_{\rho_2}^T)_{\rho_3}^T)_{\rho_4}^T = \mathcal{A}$. For N th order tensors, the number of tensor transposes is $N!$ with $(N - 1)!$ cyclic permutations on the index set $\{i_1 i_2 \dots i_N\}$.*

Henceforth, we drop the subscript σ in the transposes.

DEFINITION 3.8 (critical point). *Let $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuously differentiable function. A critical point of ϕ is a point $\bar{\mathbf{x}} \in \mathbb{R}^n$ such that*

$$\nabla \phi(\bar{\mathbf{x}}) = \mathbf{0}.$$

Consider the multilinear system

$$(3.5) \quad \mathcal{A} \bullet_3 \mathbf{x} = \mathbf{B},$$

where $\mathcal{A} \in \mathbb{R}^{I \times J \times K}$, $\mathbf{x} \in \mathbb{R}^K$, and $\mathbf{B} \in \mathbb{R}^{I \times J}$ and define

$$(3.6) \quad \phi_1(\mathbf{x}) = \|\mathcal{A} \bullet_3 \mathbf{x} - \mathbf{B}\|_F^2.$$

LEMMA 3.9. *Any minimizer $\bar{\mathbf{x}} \in \mathbb{R}^K$ of ϕ_1 satisfies the following system:*

$$(3.7) \quad \mathcal{A}^T *_2 \mathcal{A} \bullet_3 \mathbf{x} = \mathcal{A}^T *_2 \mathbf{B},$$

where $\mathcal{A}^T \in \mathbb{R}^{K \times I \times J}$ with entries $\mathcal{A}_{kij}^T = (\mathcal{A})_{ijk}$.

Proof. We expand the objective function,

$$\phi_1(\mathbf{x}) = \langle \mathcal{A} \bullet_3 \mathbf{x} - \mathbf{B}, \mathcal{A} \bullet_3 \mathbf{x} - \mathbf{B} \rangle = (\mathcal{A} \bullet_3 \mathbf{x})^T (\mathcal{A} \bullet_3 \mathbf{x}) - 2\mathbf{B}^T (\mathcal{A} \bullet_3 \mathbf{x}) + \mathbf{B}^T \mathbf{B}$$

and calculate $\frac{\partial \phi_1}{\partial \mathbf{x}}(\mathbf{x})$. It follows that

$$\frac{\partial \phi_1}{\partial \mathbf{x}}(\mathbf{x}) = 2\mathcal{A}^T *_2 \mathcal{A} \bullet_3 \mathbf{x} - 2\mathcal{A}^T *_2 \mathbf{B}$$

since

$$\begin{aligned} \frac{\partial}{\partial \mathbf{x}} [(\mathcal{A} \bullet_3 \mathbf{x})^T (\mathcal{A} \bullet_3 \mathbf{x})] &= \frac{\partial}{\partial \mathbf{x}} \left[\sum_{ij} \left(\sum_{kl} \mathbf{x}_k \mathcal{A}_{kij} \mathcal{A}_{ijl} \mathbf{x}_l \right) \right] \\ &= \frac{\partial}{\partial \mathbf{x}} \left[\sum_{kl} \left(\sum_{ij} \mathbf{x}_k \mathcal{A}_{kij} \mathcal{A}_{ijl} \mathbf{x}_l \right) \right] \\ &= \frac{\partial}{\partial \mathbf{x}} \left[\sum_{kl} \mathbf{x}_k (\mathcal{A}^T *_2 \mathcal{A})_{kl} \mathbf{x}_l \right] = 2(\mathcal{A}^T *_2 \mathcal{A}) \mathbf{x} \\ &= 2\mathcal{A}^T *_2 \mathcal{A} \bullet_3 \mathbf{x} \end{aligned}$$

and

$$\begin{aligned} 2\frac{\partial}{\partial \mathbf{x}} [(\mathcal{A} \bullet_3 \mathbf{x})^T \mathbf{B}] &= 2\frac{\partial}{\partial \mathbf{x}} \left[\sum_{ij} \left(\sum_k \mathbf{x}_k \mathcal{A}_{kij} \mathbf{B}_{ij} \right) \right] = 2\frac{\partial}{\partial \mathbf{x}} \left[\sum_k \left(\sum_{ij} \mathbf{x}_k \mathcal{A}_{kij} \mathbf{B}_{ij} \right) \right] \\ &= 2\frac{\partial}{\partial \mathbf{x}} \left[\sum_k \mathbf{x}_k (\mathcal{A}^T *_2 \mathbf{B})_k \right] \\ &= 2\mathcal{A}^T *_2 \mathbf{B}. \end{aligned}$$

Clearly, the minimizer $\bar{\mathbf{x}}$ of ϕ_1 satisfies

$$\mathcal{A}^T *_2 \mathcal{A} \bullet_3 \mathbf{x} = \mathcal{A}^T *_2 \mathbf{B}.$$

Furthermore, the critical point is $\bar{\mathbf{x}} = (\mathcal{A}^T *_2 \mathcal{A})^{-1} *_2 \mathcal{A}^T *_2 \mathbf{B}$ provided that $(\mathcal{A}^T *_2 \mathcal{A})^{-1}$ exists. \square

For the problem

$$\mathcal{A} *_2 \mathcal{X} = \mathcal{B},$$

where $\mathcal{A} \in \mathbb{R}^{I \times J \times S \times T}$, $\mathcal{X} \in \mathbb{R}^{S \times T \times K \times L}$, and $\mathcal{B} \in \mathbb{R}^{I \times J \times K \times L}$ and the objective function

$$(3.8) \quad \phi_2(\mathcal{X}) = \|\mathcal{A} *_2 \mathcal{X} - \mathcal{B}\|_F^2,$$

we have the following lemma.

LEMMA 3.10. Any minimizer $\bar{\mathcal{X}} \in \mathbb{R}^{S \times T \times K \times L}$ of ϕ_2 satisfies the system

$$(3.9) \quad \mathcal{A}^T *_2 \mathcal{A} *_2 \mathcal{X} = \mathcal{A}^T *_2 \mathcal{B},$$

where $\mathcal{A}^T \in \mathbb{R}^{S \times T \times I \times J}$ denotes the transpose of $\mathcal{A} \in \mathbb{R}^{I \times J \times S \times T}$. Moreover, the critical point of ϕ_2 is $\bar{\mathcal{X}} = (\mathcal{A}^T *_2 \mathcal{A})^{-1} *_2 \mathcal{A}^T *_2 \mathcal{B}$ provided that $(\mathcal{A}^T *_2 \mathcal{A})$ is invertible.

Remark 3.11. We omit the proof for Lemma 3.10 since it is similar to that of Lemma 3.9. The critical points, $\bar{\mathbf{x}} = (\mathcal{A}^T *_2 \mathcal{A})^{-1} *_2 \mathcal{A}^T *_2 \mathbf{B}$ and $\bar{\mathcal{X}} = (\mathcal{A}^T *_2 \mathcal{A})^{-1} *_2 \mathcal{A}^T *_2 \mathcal{B}$, are unique minimizers for (3.6) and (3.8), respectively, since ϕ_1 and ϕ_2 are quadratic functions. Equations (3.7) and (3.9) are called the high order normal equations. See Tables 3.1–3.2 for specific tensor transposes in multilinear systems.

TABLE 3.1
Third order tensor transposes in multilinear systems.

| \mathcal{A} | \mathbf{x} | \mathbf{B} | \mathcal{A}^T |
|------------------------------------|----------------|---------------------------|------------------------------------|
| $\mathbb{R}^{I \times J \times K}$ | \mathbb{R}^K | $\mathbb{R}^{I \times J}$ | $\mathbb{R}^{K \times I \times J}$ |
| $\mathbb{R}^{J \times K \times I}$ | \mathbb{R}^I | $\mathbb{R}^{J \times K}$ | $\mathbb{R}^{I \times J \times K}$ |
| $\mathbb{R}^{K \times I \times J}$ | \mathbb{R}^J | $\mathbb{R}^{K \times I}$ | $\mathbb{R}^{J \times K \times I}$ |
| $\mathbb{R}^{I \times K \times J}$ | \mathbb{R}^J | $\mathbb{R}^{K \times I}$ | $\mathbb{R}^{J \times I \times K}$ |
| $\mathbb{R}^{K \times J \times I}$ | \mathbb{R}^I | $\mathbb{R}^{K \times J}$ | $\mathbb{R}^{I \times K \times J}$ |
| $\mathbb{R}^{J \times I \times K}$ | \mathbb{R}^K | $\mathbb{R}^{J \times I}$ | $\mathbb{R}^{K \times J \times I}$ |

TABLE 3.2
Fourth order tensor transposes in multilinear systems.

| \mathcal{A} | \mathcal{X} | \mathcal{B} | \mathcal{A}^T |
|---|---|---|---|
| $\mathbb{R}^{I \times J \times S \times T}$ | $\mathbb{R}^{S \times T \times K \times L}$ | $\mathbb{R}^{S \times T \times K \times L}$ | $\mathbb{R}^{S \times T \times I \times J}$ |
| $\mathbb{R}^{J \times S \times T \times I}$ | $\mathbb{R}^{T \times I \times K \times L}$ | $\mathbb{R}^{J \times S \times K \times L}$ | $\mathbb{R}^{T \times I \times J \times S}$ |
| $\mathbb{R}^{S \times T \times I \times J}$ | $\mathbb{R}^{I \times J \times K \times L}$ | $\mathbb{R}^{S \times T \times K \times L}$ | $\mathbb{R}^{I \times J \times S \times T}$ |
| $\mathbb{R}^{T \times I \times J \times S}$ | $\mathbb{R}^{J \times S \times K \times L}$ | $\mathbb{R}^{T \times I \times K \times L}$ | $\mathbb{R}^{J \times S \times T \times I}$ |

3.2. Decompositions via isomorphic group structures. Theorem 3.2 implies that $(\mathbb{T}, *_2)$ is structurally similar to (\mathbb{M}, \cdot) . Thus we endow $(\mathbb{T}, *_2)$ with the group structure such that $(\mathbb{T}, *_2)$ and (\mathbb{M}, \cdot) are isomorphic as groups. Here we discuss some of the definitions, theorems, and decompositions preserved by the transformation.

DEFINITION 3.12 (diagonal tensor). A tensor $\mathcal{D} \in \mathbb{R}^{I \times J \times I \times J}$ is called diagonal if $d_{ijkl} = 0$ when $i \neq k$ and $j \neq l$.

DEFINITION 3.13 (identity tensor). The identity tensor \mathcal{I} is a diagonal tensor with entries

$$\mathcal{I}_{i_1 i_2 j_1 j_2} = \delta_{i_1 j_1} \delta_{i_2 j_2},$$

where

$$\delta_{lk} = \begin{cases} 1, & l = k, \\ 0, & l \neq k. \end{cases}$$

It generalizes to

$$(3.10) \quad (\mathcal{I})_{i_1 i_2 \dots i_N j_1 j_2 \dots j_N} = \prod_{k=1}^N \delta_{i_k j_k}$$

for the $2N$ th order identity tensor.

Remark 3.14. The diagonal core tensor $\mathcal{D} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ of CP (2.6) has nonzero entries d_{i_1, i_2, \dots, i_N} when $i_1 = \dots = i_N$.

DEFINITION 3.15 (orthogonal tensor). A tensor $\mathcal{U} \in \mathbb{R}^{I \times J \times I \times J}$ is orthogonal if $\mathcal{U}^T *_2 \mathcal{U} = \mathcal{I}$, where \mathcal{I} is the identity tensor under the binary operation $*_2$.

DEFINITION 3.16 (symmetric tensor). A tensor $\mathcal{S} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N \times J_1 \times \dots \times J_N}$ is symmetric if $\mathcal{S} = \mathcal{S}^T$, that is, $s_{i_1, i_2, \dots, i_n, j_1, j_2, \dots, j_n} = s_{j_1, j_2, \dots, j_n, i_1, i_2, \dots, i_n}$.

THEOREM 3.17 (SVD). Let $\mathcal{A} \in \mathbb{R}^{I \times J \times I \times J}$ with $R = \text{rank}(f(\mathcal{A}))$, where f is the transformation in (2.4). The SVD for tensor \mathcal{A} has the form

$$(3.11) \quad \mathcal{A} = \mathcal{U} *_2 \mathcal{D} *_2 \mathcal{V}^T,$$

where $\mathcal{U} \in \mathbb{R}^{I \times J \times I \times J}$ and $\mathcal{V} \in \mathbb{R}^{I \times J \times I \times J}$ are orthogonal tensors and $\mathcal{D} \in \mathbb{R}^{I \times J \times I \times J}$ is a diagonal tensor with entries σ_{ijij} called the singular values. Moreover, (3.11) can be written as

$$(3.12) \quad \mathcal{A} = \sum_{kl} \sum_{i_j, \hat{i}_j} \sigma_{klkl} (\mathbf{U}_{kl}^{(3,4)})_{ij} \circ (\mathbf{V}_{kl}^{(3,4)})_{\hat{i}_j},$$

a sum of fourth order tensors where the left and right singular matrices, $\mathbf{U}_{ij}^{(3,4)}$ and $\mathbf{V}_{ij}^{(3,4)}$, are matrix blocks of \mathcal{U} and \mathcal{V} , respectively.

The symbol \circ denotes the outer product where $\mathcal{A}_{ijkl} = \mathbf{B}_{ij} \circ \mathbf{C}_{kl} = \mathbf{B}_{ij} \mathbf{C}_{kl}$.

Proof. Let $\mathbf{A} = f(\mathcal{A})$. From the isomorphic property (3.2) and Theorem 3.2, we have

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{D} \cdot \mathbf{V}^T \xrightarrow{f^{-1}} \mathcal{A} = \mathcal{U} *_2 \mathcal{D} *_2 \mathcal{V}^T,$$

where \mathbf{U} and \mathbf{V} are orthogonal matrices and \mathbf{D} is a diagonal matrix. In addition, $\mathbf{U} \cdot \mathbf{U}^T = \mathbf{I}$ and $\mathbf{V} \cdot \mathbf{V}^T = \mathbf{I} \xrightarrow{f^{-1}} \mathcal{U}^T *_2 \mathcal{U} = \mathcal{I}$ and $\mathcal{V}^T *_2 \mathcal{V} = \mathcal{I}$. \square

THEOREM 3.18 (eigenvalue decomposition (EVD) for symmetric tensor). *Let $\bar{\mathcal{A}} \in \mathbb{R}^{I \times J \times I \times J}$ and $R = \text{rank}(f(\mathcal{A}))$, where f is the transformation in (2.4). $\bar{\mathcal{A}}$ is a real symmetric tensor if and only if there is a real orthogonal tensor $\mathcal{P} \in \mathbb{R}^{I \times J \times I \times J}$ and a real diagonal tensor $\bar{\mathcal{D}} \in \mathbb{R}^{I \times J \times I \times J}$ such that*

$$(3.13) \quad \bar{\mathcal{A}} = \mathcal{P} *_2 \bar{\mathcal{D}} *_2 \mathcal{P}^T,$$

where $\mathcal{P} \in \mathbb{R}^{I \times J \times I \times J}$ is an orthogonal tensor and $\bar{\mathcal{D}} \in \mathbb{R}^{I \times J \times I \times J}$ is a diagonal tensor with entries $\bar{\sigma}_{klkl}$ called the eigenvalues. Moreover, (3.13) can be written as

$$(3.14) \quad \bar{\mathcal{A}} = \sum_{kl} \sum_{ij\hat{i}\hat{j}} \bar{\sigma}_{klkl} (\mathbf{P}_{kl}^{(3,4)})_{ij} \circ (\mathbf{P}_{kl}^{(3,4)})_{\hat{i}\hat{j}},$$

a sum of fourth order tensors where the eigenmatrices $\mathbf{P}_{kl}^{(3,4)} \in \mathbb{R}^{I \times J}$ are the matrix blocks of \mathcal{P} .

Proof. Through the mapping provided through the isomorphic property (3.2) and Theorem 3.2, there exist an orthogonal tensor \mathcal{P} and a diagonal tensor $\bar{\mathcal{D}}$ such that $\bar{\mathcal{A}} = \mathcal{P} *_2 \bar{\mathcal{D}} *_2 \mathcal{P}^T$. Moreover, the fourth order tensor $\hat{\mathcal{P}}_{ij\hat{i}\hat{j}} = (\mathbf{P}_{kl}^{(3,4)})_{ij} \circ (\mathbf{P}_{kl}^{(3,4)})_{\hat{i}\hat{j}}$ is symmetric since $\hat{\mathcal{P}}_{ij\hat{i}\hat{j}} = \sum_{ij,\hat{i}\hat{j}} (\mathbf{P}_{kl}^{(3,4)})_{ij} \circ (\mathbf{P}_{kl}^{(3,4)})_{\hat{i}\hat{j}} = \sum_{ij,\hat{i}\hat{j}} \mathcal{P}_{ijkl} \mathcal{P}_{\hat{i}\hat{j}kl}^T = \sum_s \mathbf{P}_{rs} \cdot \mathbf{P}_{\hat{r}\hat{s}}^T = \sum_s \mathbf{P}_{\hat{r}\hat{s}} \cdot \mathbf{P}_{rs}^T = \sum_{ij,\hat{i}\hat{j}} \mathcal{P}_{\hat{i}\hat{j}kl} \mathcal{P}_{ijkl}^T = (\mathbf{P}_{kl}^{(3,4)})_{\hat{i}\hat{j}} \circ (\mathbf{P}_{kl}^{(3,4)})_{ij} = \hat{\mathcal{P}}_{\hat{i}\hat{j}ij}$. \square

Remark 3.19. If the eigenmatrix $(\mathbf{P}_{kl}^{(3,4)})$ is symmetric, that is, $(\mathbf{P}_{kl}^{(3,4)})_{ij} = (\mathbf{P}_{kl}^{(3,4)})_{ji}$, then the entries of $\bar{\mathcal{A}}$ have the following symmetry: $\bar{a}_{jilk} = \bar{a}_{ijkl}$. If $\bar{a}_{jilk} = \bar{a}_{ijkl}$ and $\bar{a}_{ijkl} = \bar{a}_{klij}$, then (3.14) is exactly the tensor eigendecomposition found in the paper of De Lathauwer, Castaing, and Cardoso [18] when $I = J$. The fourth order tensor in [18] is a quadricovariance tensor in the blind identification of underdetermined mixtures problems in signal processing. Also, see Figure 3.2 for the comparison of the core tensors from CP decomposition (2.6) and tensor SVD (3.18).

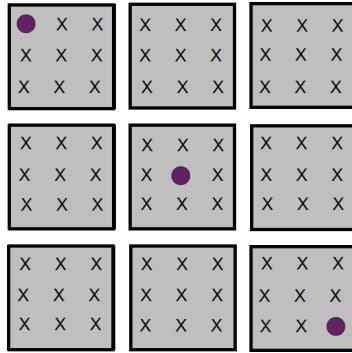
3.3. Connections to standard tensor decompositions. Here we relate tensor SVD (3.11) to both CP (2.6) and Tucker (2.8) decompositions.

LEMMA 3.20. *Let $\mathcal{T} \in \mathbb{R}^{I \times J \times I \times J}$ and $R = \text{rank}(f(\mathcal{T}))$, where f is the transformation in (2.4). The tensor SVD (3.11) in Theorem 3.17 is equivalent to CP (2.6) if there exist $\mathcal{A} \in \mathbb{R}^{I \times I \times J}$, $\mathcal{B} \in \mathbb{R}^{J \times I \times J}$, $\mathcal{C} \in \mathbb{R}^{I \times I \times J}$, and $\mathcal{D} \in \mathbb{R}^{J \times I \times J}$ such that $a_{ikl} b_{jkl} = u_{ijkl}$ and $c_{ikl} d_{jkl} = v_{\hat{i}\hat{j}kl}$.*

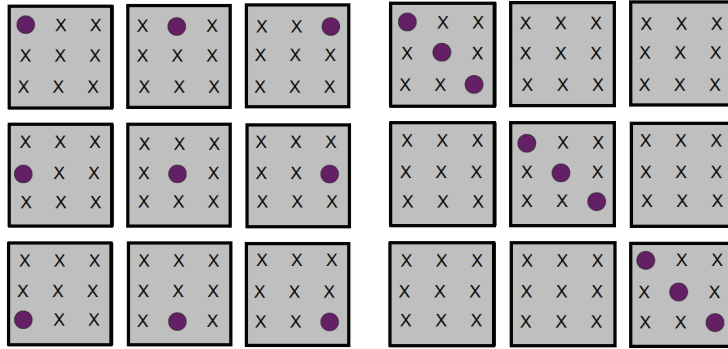
Proof. Define $r = k + (l - 1)I$. Then

$$\mathcal{T}_{ij\hat{i}\hat{j}} = \sum_{kl} \sigma_{klkl} (\mathbf{U}_{kl}^{(3,4)})_{ij} \circ (\mathbf{V}_{kl}^{(3,4)})_{\hat{i}\hat{j}} = \sum_{r=1}^R \bar{\sigma}_{rr} (\mathbf{U}_r^{(3,4)})_{ij} \circ (\mathbf{V}_r^{(3,4)})_{\hat{i}\hat{j}},$$

where $\sigma_{klkl} = \hat{\sigma}_{rr}$. Since $u_{ijkl} = (\mathbf{U}_{kl}^{(3,4)})_{ij} = (\mathbf{U}_r^{(3,4)})_{ij} = a_{ir} b_{jr}$ and $v_{\hat{i}\hat{j}kl} =$



(a) Matrix block $\mathbf{S}_{i_3 i_4}^{(3,4)}$ of fourth order core tensor from the CP decomposition



(b) Matrix block $\mathbf{S}_{i_3 i_4}^{(3,4)}$ of fourth order core tensor from tensor SVD

(c) Matrix block $\mathbf{S}_{i_2 i_4}^{(2,4)}$ of fourth order core tensor from tensor SVD

FIG. 3.2. Matrix blocks of the core tensors $\mathcal{A} \in \mathbb{R}^{3 \times 3 \times 3 \times 3}$ with nonzero diagonal entries (purple dots).

$(\mathbf{V}_{kl}^{(3,4)})_{i\hat{j}} = (\mathbf{V}_r^{(3,4)})_{i\hat{j}} = c_{i_r} d_{j_r}$, it follows that

$$\begin{aligned} \mathcal{T}_{i_j i_{\hat{j}}} &= \sum_{r=1}^R \hat{\sigma}_{rr} (\mathbf{U}_r^{(3,4)})_{ij} \circ (\mathbf{V}_r^{(3,4)})_{i\hat{j}} \\ &= \sum_{r=1}^R \hat{\sigma}_{rr} (\mathbf{U}_r^{(3,4)})_{ij} \circ (\mathbf{V}_r^{(3,4)})_{i\hat{j}} = \sum_{r=1}^R \hat{\sigma}_{rr} a_{i_r} b_{j_r} c_{i_r} d_{j_r}. \end{aligned}$$

Then,

$$(3.15) \quad \mathcal{T} = \sum_{r=1}^R \bar{\sigma}_{rr} a_r \circ b_r \circ c_r \circ d_r.$$

Moreover, the factor matrices $\mathbf{A} \in \mathbb{R}^{I \times IJ}$, $\mathbf{B} \in \mathbb{R}^{J \times IJ}$, $\mathbf{C} \in \mathbb{R}^{I \times IJ}$, and $\mathbf{D} \in \mathbb{R}^{J \times IJ}$ are built from concatenating the vectors a_r , b_r , c_r , and d_r , respectively. \square

Remark 3.21. To satisfy existence and uniqueness of the CP decomposition, the inequality (2.7) must hold. If $R = IJ = \text{rank}(f(\mathcal{T}))$, then (3.15) does not satisfy (2.7). However, if $f(\mathcal{T})$ is sufficiently low rank, that is, $R = \text{rank}(f(\mathcal{T})) < IJ$ for some I and J , then (2.7) holds.

LEMMA 3.22. Let $\mathcal{T} \in \mathbb{R}^{I \times J \times I \times J}$ and $R = \text{rank}(f(\mathcal{T}))$, where f is the transformation in (2.4). The tensor SVD (3.11) in Theorem 3.17 is equivalent to multilinear SVD (2.8) if there exist $\mathbf{A} \in \mathbb{R}^{I \times I}$, $\mathbf{B} \in \mathbb{R}^{J \times J}$, $\mathbf{C} \in \mathbb{R}^{I \times I}$, and $\mathbf{D} \in \mathbb{R}^{J \times J}$ such that $a_{ik}b_{jl} = u_{ijkl}$ and $c_{ik}d_{jl} = v_{ijkl}$.

Proof. From (3.12), we have $\mathcal{T}_{ij\hat{i}\hat{j}} = \sum_{kl} \sigma_{klkl} (\mathbf{U}_{kl}^{(3,4)})_{ij} \circ (\mathbf{V}_{kl}^{(3,4)})_{\hat{i}\hat{j}}$ which implies $\mathcal{T}_{ij\hat{i}\hat{j}} = \sum_{kl} \sigma_{klkl} a_{ik} b_{jl} c_{ik} d_{jl}$. \square

Remark 3.23. Typically, the core tensor of a multilinear SVD (2.8) is dense. However, the core tensor resulting from Lemma 3.22 is not dense (possibly sparse), i.e., there are IJ nonzeros elements from the total I^2J^2 entries in the fourth order core tensor of size $I \times J \times I \times J$. Similarly, the existence of the decomposition impinges upon the existence of the factors $\mathbf{A} \in \mathbb{R}^{I \times I}$, $\mathbf{B} \in \mathbb{R}^{J \times J}$, $\mathbf{C} \in \mathbb{R}^{I \times I}$, and $\mathbf{D} \in \mathbb{R}^{J \times J}$ such that $\mathcal{U} = \mathbf{A} \circ \mathbf{B}$ and $\mathcal{V} = \mathbf{C} \circ \mathbf{D}$.

COROLLARY 3.24. Let $\mathcal{T} \in \mathbb{R}^{I \times J \times I \times J}$ be symmetric and $R = \text{rank}(f(\mathcal{T}))$, where f is the transformation in (2.4). The tensor EVD (3.14) in Theorem 3.18 is equivalent to CP (2.6) if there exist $\mathcal{A} \in \mathbb{R}^{I \times I \times J}$, $\mathcal{B} \in \mathbb{R}^{J \times I \times J}$ such that $a_{ikl}b_{jkl} = p_{ijkl}$.

COROLLARY 3.25. Let $\mathcal{T} \in \mathbb{R}^{I \times J \times I \times J}$ with symmetries $t_{ijkl} = t_{klij}$ and $t_{jikl} = t_{ijkl}$ with $R = \text{rank}(f(\mathcal{T}))$. The tensor EVD (3.14) in Theorem 3.18 is equivalent to CP (2.6) if there exist $\mathcal{A} \in \mathbb{R}^{I \times I \times J}$, $\mathcal{B} \in \mathbb{R}^{J \times I \times J}$ such that $a_{ikl}b_{jkl} = p_{ijkl}$.

Remark 3.26. From Corollary 3.24, we have $\mathcal{T}_{ij\hat{i}\hat{j}} = \sum_{r=1}^R \bar{\sigma}_{rr} (\mathbf{P}_r^{(3,4)})_{ij} \circ (\mathbf{P}_r^{(3,4)})_{\hat{i}\hat{j}} = \sum_{r=1}^R \bar{\sigma}_{rr} a_{ir} b_{jr} a_{ir} b_{jr}$ with identical factors, $\mathbf{A} = \mathbf{C}$ and $\mathbf{B} = \mathbf{D}$. As in Remark 3.17, the existence of the factors \mathbf{A} and \mathbf{B} requires the matrix blocks $\mathbf{P}_{kl}^{(3,4)}$ to be rank-one matrices. In Corollary 3.25, the partial symmetry $t_{jikl} = t_{ijkl}$ implies that $\mathbf{P}_{kl}^{(3,4)}$ is symmetric (as well as rank-one). Thus, $(\mathbf{P}_{kl}^{(3,4)})_{ij} = a_{ikl}a_{jkl} \implies \mathcal{T} = \sum_{r=1}^R \bar{\sigma}_{rr} a_{ir} a_{jr} a_{ir} a_{jr}$. This decomposition is known as symmetric CP decomposition [11].

COROLLARY 3.27. Let $\mathcal{T} \in \mathbb{R}^{I \times J \times I \times J}$ be symmetric and $R = \text{rank}(f(\mathcal{T}))$, where f is the transformation in (2.4). The tensor EVD (3.14) in Theorem 3.18 is equivalent to multilinear SVD (2.8) if there exist $\mathbf{A} \in \mathbb{R}^{I \times I}$ and $\mathbf{B} \in \mathbb{R}^{J \times J}$ such that $a_{ik}b_{jl} = p_{ijkl}$.

Remark 3.28. The multilinear SVD from Corollary 3.27 is $\mathcal{T}_{ij\hat{i}\hat{j}} = \sum_{kl} \sigma_{klkl} (\mathbf{P}_{kl}^{(3,4)})_{ij} \circ (\mathbf{P}_{kl}^{(3,4)})_{\hat{i}\hat{j}} = \sum_{kl} \sigma_{klkl} a_{ik} b_{jl} a_{ik} b_{jl}$ following Lemma 3.22.

4. Numerical examples involving multilinear systems.

4.1. Multilinear systems in Poisson problems. Consider the two-dimensional (2D) Poisson problem

$$(4.1) \quad \begin{aligned} -\nabla^2 v &= f && \text{in } \Omega, \\ u &= 0 && \text{on } \Gamma, \end{aligned}$$

where $\Omega = \{(x, y) : 0 < x, y < 1\}$ with boundary Γ , f is a given function, and

$$\nabla^2 v = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}.$$

We compute an approximation of the unknown function $v(x, y)$ in (4.1). Several problems in physics and mechanics are modeled by (4.1), where the solution v represents, for example, temperature, electromagnetic potential, or displacement of an elastic membrane fixed at the boundary.

The mesh points are obtained by discretizing the unit square domain with step sizes, Δx in the x -direction and Δy in the y -direction. From the standard central

difference approximations, the difference formula,

$$(4.2) \quad \frac{v_{l-1,m} - 2v_{l,m} + v_{l+1,m}}{\Delta x^2} + \frac{v_{l,m-1} - 2v_{l,m} + v_{l,m+1}}{\Delta y^2} = f(x_l, y_m),$$

is obtained. If we assume $\Delta x = \Delta y$, then the difference equation (4.2) is equivalent to

$$(4.3) \quad \mathbf{A}_N \mathbf{V} + \mathbf{V} \mathbf{A}_N = (\Delta x)^2 \mathbf{F},$$

where

$$(4.4) \quad \mathbf{A}_N = \begin{bmatrix} 2 & -1 & & 0 \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ 0 & & -1 & 2 \end{bmatrix},$$

$$(4.5) \quad \mathbf{V} = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1N} \\ v_{21} & v_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & v_{N-1N} \\ v_{N1} & \dots & v_{NN-1} & v_{NN} \end{bmatrix}, \quad \text{and}$$

$$\mathbf{F} = \begin{bmatrix} f_{11} & f_{12} & \dots & f_{1N} \\ f_{21} & f_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & f_{N-1N} \\ f_{N1} & \dots & f_{NN-1} & f_{NN} \end{bmatrix}.$$

The entries of \mathbf{V} and \mathbf{F} are the values on the mesh on the unit square where, $(x_i, y_j) = (i\Delta x, j\Delta x) \in [0, 1] \times [0, 1]$. Here the Dirichlet boundary conditions are imposed so the values of \mathbf{V} are zero at the boundary of the unit square, i.e., $v_{i0} = v_{iN+1} = v_{0,j} = v_{N+1,j} = 0$ for $0 < i, j < N + 1$.

Typically, \mathbf{V} and \mathbf{F} are vectorized which gives the following linear system:

$$(4.6) \quad \mathbf{A}_{N \times N} \cdot \mathbf{v} = \begin{bmatrix} \mathbf{A}_N + 2\mathbf{I}_N & -\mathbf{I}_N & & 0 \\ -\mathbf{I}_N & \mathbf{A}_N + 2\mathbf{I}_N & \ddots & \\ & \ddots & \ddots & -\mathbf{I}_N \\ 0 & & -\mathbf{I}_N & \mathbf{A}_N + 2\mathbf{I}_N \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{12} \\ \vdots \\ v_{NN} \end{bmatrix} \\ = (\Delta x)^2 \begin{bmatrix} f_{11} \\ f_{12} \\ \vdots \\ f_{NN} \end{bmatrix}.$$

Poisson's equation in two dimensions is expressed as a sum of Kronecker products [19], i.e.,

$$(4.7) \quad \mathbf{A}_{N \times N} = \mathbf{I}_N \otimes \mathbf{A}_N + \mathbf{A}_N \otimes \mathbf{I}_N.$$

Moreover, the discretized problem in three dimensions is

$$(4.8) \quad (\mathbf{A}_N \otimes \mathbf{I}_N \otimes \mathbf{I}_N + \mathbf{I}_N \otimes \mathbf{A}_N \otimes \mathbf{I}_N + \mathbf{I}_N \otimes \mathbf{I}_N \otimes \mathbf{A}_N) \cdot \text{vec}(\mathbf{V}) = (\Delta x)^3 \text{vec}(\mathbf{F}).$$

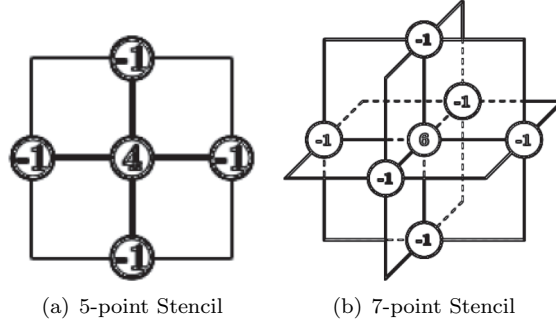


FIG. 4.1. Stencils for higher order tensors. The main node of the five-point and seven-point stencils sits on the diagonal entries of fourth order and sixth order Laplacian tensors, respectively.

The higher order tensor representation of the 2D discretized Poisson problem (4.1) is

$$(4.9) \quad \mathcal{A}_N *_2 \mathbf{V} = \mathbf{F},$$

where $\mathcal{A}_N \in \mathbb{R}^{N \times N \times N \times N}$ and matrices $\mathbf{V} \in \mathbb{R}^{N \times N}$ and $\mathbf{F} \in \mathbb{R}^{N \times N}$ are the discretized functions v and f on a unit square mesh defined in (4.5). The nonzeros entries of the matrix block $\mathbf{A}_N^{(2,4)}_{k=\alpha, l=\beta} \in \mathbb{R}^{N \times N}$ are the following:

$$(4.10) \quad \begin{cases} (\mathbf{A}_N^{(2,4)})_{\alpha, \beta} = \frac{4}{(\Delta x)^2}, \\ (\mathbf{A}_N^{(2,4)})_{\alpha-1, \beta} = \frac{-1}{(\Delta x)^2}, \\ (\mathbf{A}_N^{(2,4)})_{\alpha+1, \beta} = \frac{-1}{(\Delta x)^2}, \\ (\mathbf{A}_N^{(2,4)})_{\alpha, \beta-1} = \frac{-1}{(\Delta x)^2}, \\ (\mathbf{A}_N^{(2,4)})_{\alpha, \beta+1} = \frac{-1}{(\Delta x)^2} \end{cases}$$

for $\alpha, \beta = 2, \dots, N-1$. These entries form a five-point stencil; see Figure 4.1. The discretized three-dimensional Poisson equation is

$$(4.11) \quad \bar{\mathcal{A}}_N *_3 \mathcal{V} = \mathcal{F},$$

where $\bar{\mathcal{A}}_N \in \mathbb{R}^{N \times N \times N \times N \times N \times N}$ and $\mathcal{V}, \mathcal{F} \in \mathbb{R}^{N \times N \times N}$. Both \mathcal{V} and \mathcal{F} are discretized on the unit cube. The entries on the tensor block $(\bar{\mathcal{A}}_N)_{l, m, n}^{(2,4,6)} \in \mathbb{R}^{N \times N \times N}$ of $\bar{\mathcal{A}}_N$ would follow a seven-point stencil, i.e.,

$$(4.12) \quad \begin{cases} ((\bar{\mathcal{A}}_N)_{\alpha, \beta, \gamma}^{(2,4,6)})_{\alpha, \beta, \gamma} = \frac{6}{(\Delta x)^3}, \\ ((\bar{\mathcal{A}}_N)_{\alpha, \beta, \gamma}^{(2,4,6)})_{\alpha-1, \beta, \gamma} = \frac{-1}{(\Delta x)^3}, \\ ((\bar{\mathcal{A}}_N)_{\alpha, \beta, \gamma}^{(2,4,6)})_{\alpha+1, \beta, \gamma} = \frac{-1}{(\Delta x)^3}, \\ ((\bar{\mathcal{A}}_N)_{\alpha, \beta, \gamma}^{(2,4,6)})_{\alpha, \beta-1, \gamma} = \frac{-1}{(\Delta x)^3}, \\ ((\bar{\mathcal{A}}_N)_{\alpha, \beta, \gamma}^{(2,4,6)})_{\alpha, \beta+1, \gamma} = \frac{-1}{(\Delta x)^3}, \\ ((\bar{\mathcal{A}}_N)_{\alpha, \beta, \gamma}^{(2,4,6)})_{\alpha, \beta, \gamma-1} = \frac{-1}{(\Delta x)^3}, \\ ((\bar{\mathcal{A}}_N)_{\alpha, \beta, \gamma}^{(2,4,6)})_{\alpha, \beta, \gamma+1} = \frac{-1}{(\Delta x)^3} \end{cases}$$

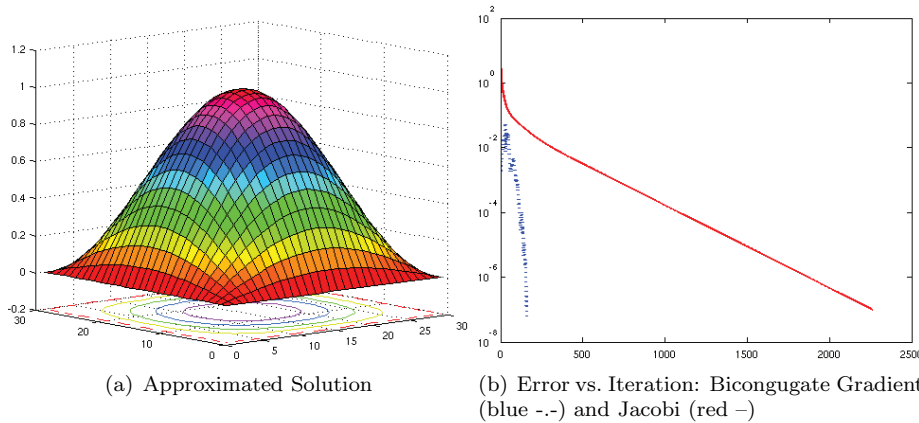


FIG. 4.2. A solution to the Poisson equation in two dimensions with Dirichlet boundary conditions.

for $\alpha, \beta, \gamma = 2, \dots, N - 1$ since v_{ijk} satisfies

$$6v_{ijk} - v_{i-1jk} - v_{i+1jk} - v_{ij-1k} - v_{ij+1k} - v_{ijk-1} - v_{ijk+1} = (\Delta x)^3 f_{ijk}.$$

The tensor representation of Poisson problems consists of the multilinear systems (4.9) and (4.11). The iterative methods used to solve (4.9) and (4.11) are described in Figure 3.1; see Figure 4.2 for the numerical results. The convergence of Jacobi is slow since the spectral radius with respect to the Poisson's equation is near one [19]. Also, the approximation in Figure 4.2 is first order accurate.

There are some advantages in computing in a tensor structured domain. First, without vectorization, the solution and the computational grid have a one-to-one correspondence so that sophisticated boundary conditions are easily integrated in the multilinear system. Second, the Laplacian tensor preserves the low bandwidth since the main nodal points sit on the tensor diagonal entries and the rest of the stencil points lie on the off-diagonal positions. Although the Laplacian matrices in (4.4) and (4.6) are banded, the Laplacian matrices in higher dimensions have larger bandwidths. The Laplacian tensor has a lower bandwidth than the Laplacian matrix. In fact, reducing the bandwidth of these sparse matrices directly and substantially improves the number of operations and storage locations; see the reordering methods of Cuthill and McKee [13] and George and Liu [24].

4.2. Multilinear systems in eigenvalue problems: The Anderson model and localization properties. The Anderson model is the most studied model for understanding spectral and transport properties of an electron in a disordered medium. In 1958, Anderson [1] described the behavior of electrons in a crystal with impurities, that is, when electrons can deviate from their sites by hopping from atom to atom and are constrained to an external random potential modeling the random environment. He argued heuristically that electrons in such systems result in a loss of the conductivity properties of the crystal, transforming it from conductors to insulators.

The Anderson model is a discrete random Schrödinger operator defined on a lattice \mathbb{Z}^d . More specifically, the Anderson model is a random Hamiltonian H_ω on $\ell^2(\mathbb{Z}^d)$, $d \geq 1$, defined by

$$(4.13) \quad H_\omega = -\Delta + \lambda V_\omega,$$

where $\Delta(x, y) = 1$ if $|x - y| = 1$ and zero otherwise (the discrete Laplacian) with

spectrum $[-2d, 2d]$ and the random potential $V_\omega = \{V_\omega(x), x \in \mathbb{Z}^d\}$ consisting of i.i.d. random variables on $[-1, 1]$ which we assume to have bounded and compactly supported density ρ . The disorder parameter is a nonnegative $\lambda > 0$. The spectrum of H_ω can be explicitly described by

$$\sigma(H_\omega) = \sigma(-\Delta) + \lambda \operatorname{supp}(\rho) = [-2d, 2d] + \lambda \operatorname{supp}(\rho).$$

Remark 4.1. The random potential V_ω is a multiplication operator on $\ell_2(\mathbb{Z}^d)$ with matrix elements $V_\omega(x) = v_x(\omega)$, where $(v_x(\omega))_{x \in \mathbb{Z}^d}$ is a collection of i.i.d. random variables with distribution ρ indexed by \mathbb{Z}^d .

The random Schrödinger operator models disordered solids. The atoms or nuclei of ideal crystals are distributed in a lattice in a *regular* way. Since most solids are not ideal crystals, the positions of the atoms may deviate away from the lattice positions. This phenomena can be attributed to imperfections in the crystallization, glassy materials, or a mixture of alloys or doped semiconductors. To model disorder, a random potential V_ω perturbs the pure laplacian Hamiltonian $(-\Delta)$ of a perfect metal. The time evolution of a quantum particle ψ is determined by the Hamiltonian H_ω , i.e.,

$$\psi(t) = e^{itH_\omega} \psi_0.$$

Thus the spectral properties of H_ω are studied to extract valuable information. In this case, the localization properties of the Anderson model are of interest. For instance, the localization properties are characterized by the spectral properties of the Hamiltonian H_ω ; see the references [31, 38, 49]. The Hamiltonian H_ω exhibits spectral localization if H_ω has almost surely pure point spectrum with exponentially decaying eigenfunctions.

Remark 4.2. Recall that from [46] for any self-adjoint operator H , the spectral decomposition is

$$\sigma(H) = \sigma_p(H) \cup \sigma_{ac}(H) \cup \sigma_{sc}(H)$$

corresponding to the invariant subspaces H_p of point spectrum, to H_{ac} of absolutely continuous spectrum, and to H_{sc} of singular continuous spectrum.

The localization properties of the Anderson model can be described by spectral or dynamical properties. Let $I \subset \mathbb{R}$.

DEFINITION 4.3. *We say that H_ω exhibits spectral localization in I if H_ω almost surely has pure point spectrum in I (with probability one), that is,*

$$\sigma(H_\omega) \cap I \subset \sigma_p(H_\omega) \text{ with probability one.}$$

Moreover, the random Schrödinger operator H_ω has exponential spectral localization in I , and the eigenfunctions corresponding to eigenvalues in I decay exponentially.

Thus if for almost all ω , the random Hamiltonian H_ω has a complete set of eigenvectors $(\psi_{\omega,n})_{n \in \mathbb{N}}$ in the energy interval I satisfying

$$|\psi_{\omega,n}(x)| \leq C_{\omega,n} e^{-\mu|x-x_{\omega,n}|}$$

with localization center $x_{\omega,n}$ for $\mu > 0$ and $C_{\omega,n} < \infty$, then the exponential spectral localization hold on I .

Remark 4.4. Let $V : \ell_2(\mathbb{Z}) \rightarrow \ell_2(\mathbb{Z})$ be a multiplication operator and suppose $v : \mathbb{Z} \rightarrow \mathbb{R}$ is a function. Then, $Vf(x) = v(x)f(x)$, and thus $\sigma(V) = \operatorname{range}(v)$.

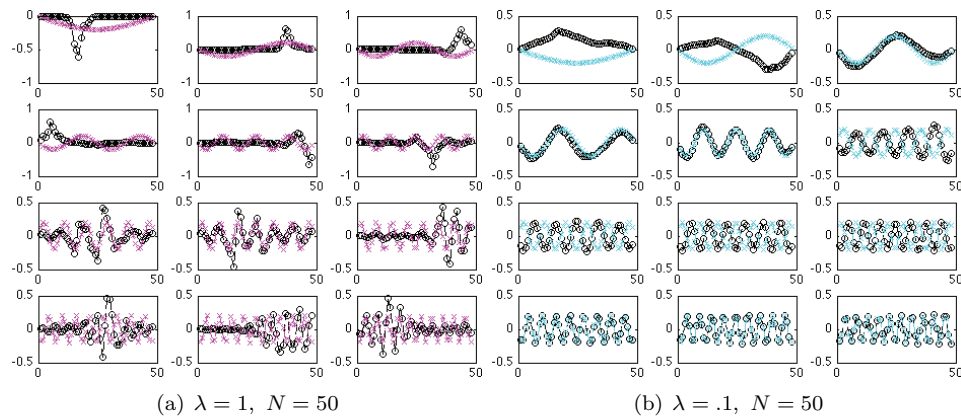


FIG. 4.3. One-dimensional eigenvectors of the discrete Schrödinger operator (—) and the Anderson model (—, —o) for various modes.

Suppose $f(x)$ is the Dirac delta function, i.e.,

$$f(x) = \delta(x - x_0) = \begin{cases} 1 & x = x_0, \\ 0 & x \neq x_0. \end{cases}$$

Then $Vf(x) = v(x_0)f(x)$ which implies that $\sigma(V) = \sigma_p(V)$, i.e., V has a pure point spectrum.

DEFINITION 4.5. A random Schrödinger operator has strong dynamical localization in an interval I if for all $q > 0$ and all $\phi \in \ell_2(\mathbb{Z}^d)$ with compact support,

$$\mathbb{E} \left[\sup_t \| |X|^q e^{-itH_\omega} \chi_I(H_\omega) \psi \|^2 < \infty \right],$$

where χ_I is an indicator function and X is a multiplicative operator from $\ell_2(\mathbb{Z}^d) \rightarrow \ell_2(\mathbb{Z}^d)$ defined as $|X|\psi = |x|\psi(x)$.

Dynamical localization in this form implies that all moments of the position operator are bounded in time.

The Anderson model is a well-studied subject. So there are numerous results in both physics and mathematics literature; see [31] and the references therein. There are several theoretical results on the existence of localization for $d = 1$ for all energies and arbitrary disorder λ ; see Kunz and Souillard [40] and Carmona, Klein, and Martinelli [7]. For any d , it is known that the localized states are present for all energies for sufficiently large disorder ($\lambda \gg 1$). For $d = 2$ and with a Gaussian distribution, it is conjectured that there is a localization for any amount of disorder λ as in the case for $d = 1$. There are many more open problems for higher dimension like the extended state conjecture [23].

In material science, many techniques like X-ray beams and X-ray micro diffraction (e.g., see [32, 33]) have been used to study elemental composition to understand how the atoms are arranged and to identify some defects. Powerful computer simulations and mathematical modeling are being applied increasingly as visualization tools for understanding crystal structure and evolution. Here we develop a tensor-based visualization tool for localization and for verifying some conjectures in high dimension ($d \geq 3$) for all disorder λ and at various distributions.

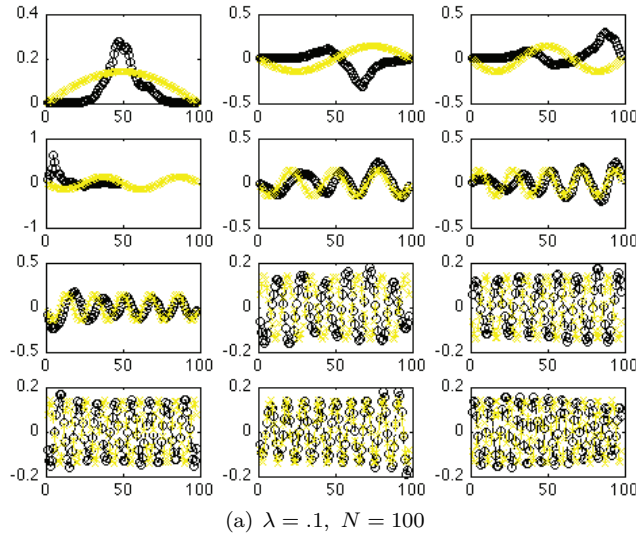


FIG. 4.4. One-dimensional eigenvectors of the discrete Schrödinger operator (-x-) and the Anderson model (black, -o-) for various modes.

In Figures 4.3–4.7, the eigenvectors of the Anderson model are compared against the eigenvectors of the discrete Schrödinger operator model for dimension $d = 1, 2, 3$. Recall that the discrete Schrödinger operator models *perfect* crystals without disorder, that is, where $\lambda = 0$. In Figure 4.5, the 2D eigenvectors of the Anderson and the discrete Schrödinger are calculated through the tensor SVD (3.18) based on the isomorphic map and the standard multilinear SVD (2.8).

4.2.1. Approximation of eigenvectors. To approximate the eigenvectors of the multidimensional Anderson model, the eigenvalue decomposition in Theorem 3.18 is applied to the Hamiltonian H_ω . The Hamiltonian H_ω in two and three dimensions are formed into fourth and sixth order tensors using the same stencils in Figure 4.1 with entries in (4.10) and (4.12), respectively. Note that the center nodes are around zero and have random entries, i.e.,

$$(4.14) \quad (H_{\omega_{k=\alpha, l=\beta}}^{(3,4)})_{\alpha, \beta} = \frac{\sigma}{(\Delta x)^2}$$

$$(4.15) \quad \text{and} \quad (H_{\omega_{l=\alpha, m=\beta, n=\gamma}}^{(4,5,6)})_{\alpha, \beta, \gamma} = \frac{\tau}{(\Delta x)^3},$$

where σ and τ are random numbers with uniform distribution on $[-1, 1]$ accounting for the random diagonal potential V_ω . Higher order tensor representation easily preserved the uniform distribution on $[-1, 1]$ on the random potential, but this is not necessarily true for Hamiltonians in (4.7) and (4.8). To compute numerically the higher-dimensional eigenvector, the tensor representation of the Hamiltonian is necessary before the appropriate Einstein product rules and mappings are applied.

In Figures 4.3, 4.4, 4.5, 4.6, and 4.7, the eigenfunctions are approximated by the eigenvectors from the both discrete Schrödinger and random Schrödinger (Anderson) models. In Figures 4.3 and 4.4, the eigenvectors of the Anderson model in one dimension are definitely more localized than the eigenvectors of the discrete random

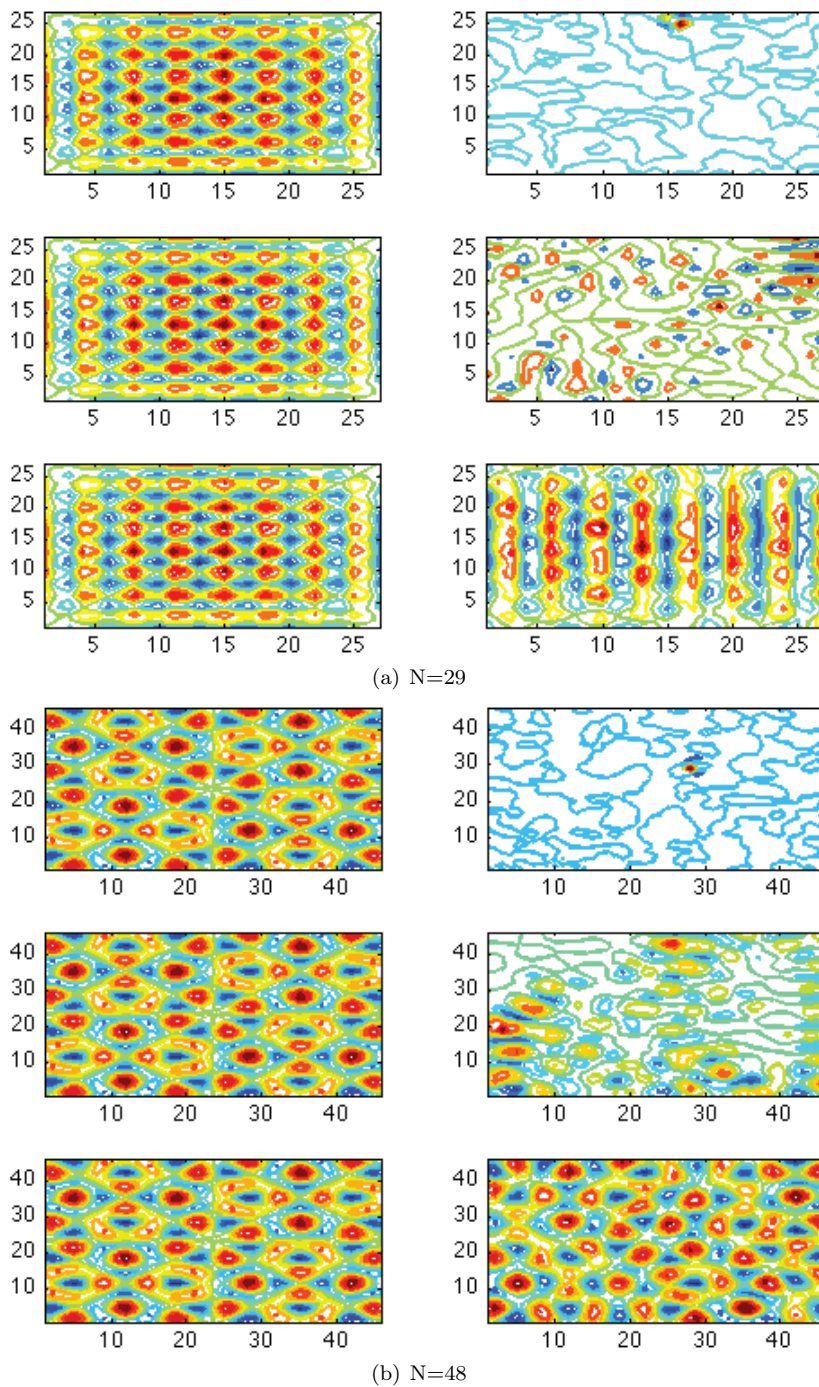


FIG. 4.5. 2D eigenvectors of the discrete Schrödinger operator (left column) and the Anderson model (right column) for varying disorder ($\lambda = 10$ (top), $\lambda = 1$ (middle), and $\lambda = .1$ (bottom)).

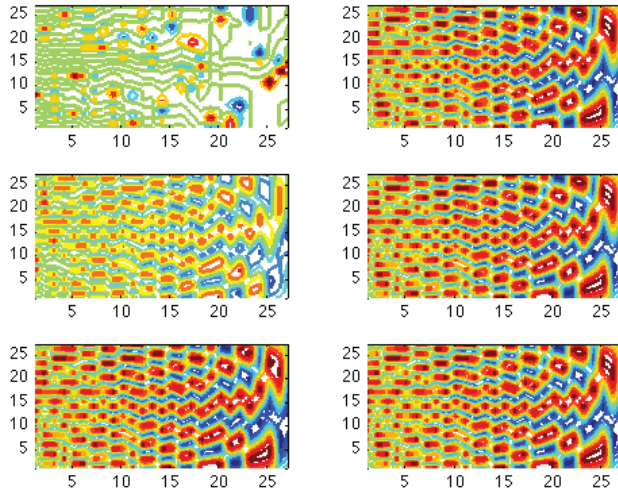


FIG. 4.6. Factors of the multilinear SVD decomposition [16] of the 2D discrete Schrödinger operator (right column) and the Anderson model (left column) for varying disorder ($\lambda = 10$ (top), $\lambda = 1$ (middle), and $\lambda = .1$ (bottom)).

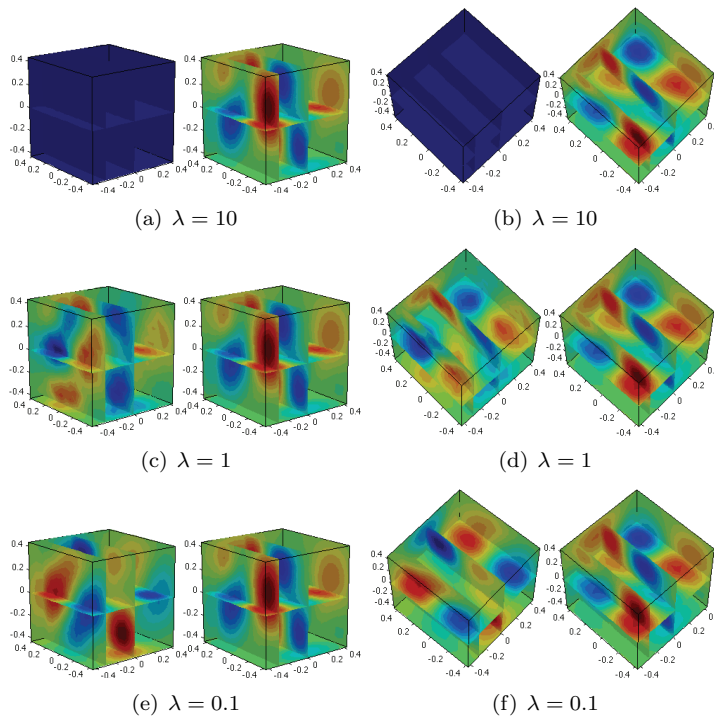


FIG. 4.7. Two views (first column and second column) of the three-dimensional eigenvectors of the Anderson model (left) and the discrete Schrödinger operator (right) for varying disorder.

Schrödinger model in one dimension which are consistent with the results in [31] for the Anderson model in one dimension. Observe that for large amount of disorder (e.g., $\lambda = 1$), the localized states are apparent. However this is not true for a smaller amount of disorder (e.g., $\lambda = .1$). The localization is not so apparent for $\lambda = .1$

$N = 50$, but when the number of atoms is increased, that is, setting $N = 100$, the localized eigenvectors are present as in the case when $\lambda = 1$; see Figures 4.3(b) and 4.4.

In the contour plots of Figures 4.5 and 4.7, the eigenvectors in two and three dimensions of the Anderson model are more *peaked* than those of the nonrandomized Schrödinger for large disorder $\lambda \geq 1$. As in the case for one dimension, localization is not apparent for small disorder ($\lambda = .1$) as seen in Figure 4.5. Moreover, as N increases, the eigenstates of both discrete Schrödinger and Anderson models seem to coincide for small disorder. This does not necessarily mean that the localization is absent for this regime, but rather the localized states are harder to find for a small amount of disorder. A larger amount of atoms has to be considered for this case. In Figure 4.5, localization is not clearly visible for even $\lambda = 1$ in the factors calculated via the multilinear SVD decomposition (2.8) while localization is detected in the plots in Figure 4.5 when $\lambda = 1$. The plots in Figure 4.6 are generated by applying the higher order orthogonal iteration algorithm [17] to the Hamiltonian tensors (4.14) and (4.15).

The numerical results provide some validations that these localizations exist for large disorder for dimensions $d > 1$ for a sufficient amount of atoms.

5. Conclusion. We have shown that even order tensors with a rectangular form $I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_N$ are invertible by a transformation to the general linear group equipped with the Einstein contracted product. While odd order tensors and even order tensors with distinct modes are not invertible, we have extended the notion of pseudoinversion for these cases. Alternative multilinear SVD and EVD decompositions arise from these isomorphic properties. Notably, these decompositions give a natural framework for the eigenvalue problems of quantum models like the discrete Schrödinger and Anderson models as well as solving Poisson problems in high dimensions. Moreover, CP and multilinear SVD can be computed through these alternative decompositions provided that some symmetry constraints hold. These alternative decompositions also provide a simple way to factorize quadricovariance tensors in blind identification of underdetermined mixtures problems.

Solving multilinear systems with the tensor-based iterative methods has several advantages: (1) higher order tensor representation of PDEs preserve low bandwidth thereby keeping the computational cost and memory requirement low and (2) a one-to-one correspondence between the solution and the computational grid facilitate the integration of complicated boundary conditions. Moreover, the tensor representation of eigenvalue problems in quantum statistical mechanics yields visualization tools for studying localization properties of the Anderson model in high dimension.

To improve the applicability of tensor-based methods in solving high-dimensional eigenvalue problems and multilinear systems, we plan to further develop these proposed methods in several research lines. We plan to make the iterative solvers more efficient. Starting with the high-dimensional Poisson problems and eigenvalue problems, a novel method which can operate on each sparse matrix (tensor) blocks as opposed to the whole tensor structure will cut down the memory and operational costs. Extensions of these methods in nonsymmetric and sparse multilinear systems are important since these systems have direct applications in analyzing genes and proteins interaction. Further developments of tensor SVD and tensor EVD for partial symmetric fourth order tensors are needed to provide direct methods in decomposing quadricovariance tensors in prevalent signal processing applications.

Appendix: Proof of Theorem 3.2. Recall the basic definitions.

DEFINITION 5.1 (binary operation). *A binary operation \star on a set \mathbb{G} is a rule that assigns to each ordered pair (A, B) of elements of \mathbb{G} some element of \mathbb{G} .*

DEFINITION 5.2. A group (\mathbb{G}, \star) is a set \mathbb{G} with the operation \star which satisfy the following axioms:

(A1) The binary operation \star is associative, i.e., $(A \star B) \star C = A \star (B \star C)$ for $A, B, C \in \mathbb{G}$.

(A2) There is an element $\mathcal{J} \in \mathbb{G}$ such that $\mathcal{J} \star X = X \star \mathcal{J}$ for all $X \in \mathbb{G}$. This element \mathcal{J} is an identity element for \star on \mathbb{G} .

(A3) For each $A \in \mathbb{G}$, there is an element $\tilde{A} \in \mathbb{G}$ with the property that $\tilde{A} \star A = A \star \tilde{A} = \mathcal{J}$.

Proof. Here we prove by checking each axiom (A1–A3) hold in Definition 5.2.

(A1) Show that the binary operation \star_2 is associative.

Since we know that f is a bijective map with the property that $f(A \star_2 B) = f(A) \cdot f(B)$, we will show $f^{-1}(A \cdot B) = f^{-1}(A) \star_2 f^{-1}(B)$ for $A, B \in \mathbb{M}$.

Let $\mathcal{A}, \mathcal{B}, \mathcal{C} \in \mathbb{T}$ and $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathbb{M}$, where $f(\mathbf{A}) = \mathcal{A}$, $f(\mathbf{B}) = \mathcal{B}$ and $f(\mathbf{C}) = \mathcal{C}$. Then,

$$\begin{aligned} (\mathcal{A} \star_2 \mathcal{B}) \star_2 \mathcal{C} &= f^{-1}(\mathbf{A}) \star_2 f^{-1}(\mathbf{B}) \star_2 f^{-1}(\mathbf{C}) = f^{-1}(\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{C}) = f^{-1}(\mathbf{A} \cdot (\mathbf{B} \cdot \mathbf{C})) \\ &= f^{-1}(\mathbf{A}) \star_2 f^{-1}(\mathbf{B} \cdot \mathbf{C}) = \mathcal{A} \star_2 (f^{-1}(\mathbf{B}) \star_2 f^{-1}(\mathbf{C})) = \mathcal{A} \star_2 (\mathcal{B} \star_2 \mathcal{C}). \end{aligned}$$

Therefore, $(\mathcal{A} \star_2 \mathcal{B}) \star_2 \mathcal{C} = \mathcal{A} \star_2 (\mathcal{B} \star_2 \mathcal{C})$.

(A2) Show that there is an identity element for \star_2 on \mathbb{T} .

Since $\mathbf{I}^{I_1 I_2 \times I_1 I_2} \in \mathbb{M}$ is the identity element in the group, note that we will suppress the superscript of \mathbf{I} in the calculation below. Then we claim that $f^{-1}(\mathbf{I})$ is the identity element for \star_2 on \mathbb{T} .

For every element $\mathcal{A} \in \mathbb{T}$, there exists a matrix $A \in \mathbb{M}$ so that $f^{-1}(\mathbf{A}) = A$. So, we get

$$\mathcal{A} \star_2 f^{-1}(\mathbf{I}) = f^{-1}(\mathbf{A}) \star_2 f^{-1}(\mathbf{I}) = f^{-1}(\mathbf{A} \cdot \mathbf{I}) = f^{-1}(\mathbf{A}) = \mathcal{A}.$$

Similarly,

$$f^{-1}(\mathbf{I}) \star_2 \mathcal{A} = f^{-1}(\mathbf{I}) \star_2 f^{-1}(\mathbf{A}) = f^{-1}(\mathbf{I} \cdot \mathbf{A}) = f^{-1}(\mathbf{A}) = \mathcal{A}.$$

Therefore, $\mathcal{A} \star_2 f^{-1}(\mathbf{I}) = f^{-1}(\mathbf{I}) \star_2 \mathcal{A} = \mathcal{A}$.

Define the tensor \mathcal{E} as follows:

$$(\mathcal{E})_{i_1 i_2 j_1 j_2} = \delta_{i_1 j_1} \delta_{i_2 j_2},$$

where

$$\delta_{lk} = \begin{cases} 1, & l = k, \\ 0, & l \neq k. \end{cases}$$

We claim that $\mathcal{E} = f^{-1}(\mathbf{I})$. By direct calculations, we have

$$\begin{aligned} (\mathcal{E} \star_2 \mathcal{A})_{i_1 i_2 j_1 j_2} &= \sum_{u,v} \epsilon_{i_1 i_2 uv} a_{uv j_1 j_2} = \epsilon_{i_1 i_2 i_1 i_2} a_{i_1 i_2 j_1 j_2} \\ &= \delta_{i_1 i_1} \delta_{i_2 i_2} a_{i_1 i_2 j_1 j_2} = a_{i_1 i_2 j_1 j_2} = \mathcal{A}_{i_1 i_2 j_1 j_2} \end{aligned}$$

and

$$\begin{aligned} (\mathcal{A} \star_2 \mathcal{E})_{i_1 i_2 j_1 j_2} &= \sum_{u,v} a_{i_1 i_2 uv} \epsilon_{uv j_1 j_2} = a_{i_1 i_2 j_1 j_2} \epsilon_{j_1 i_2 j_1 j_2} \\ &= a_{i_1 i_2 j_1 j_2} \delta_{j_1 j_1} \delta_{j_2 j_2} = a_{i_1 i_2 j_1 j_2} = \mathcal{A}_{i_1 i_2 j_1 j_2}. \end{aligned}$$

Thus $\mathcal{E} *_2 \mathcal{A} = \mathcal{A} *_2 \mathcal{E} = \mathcal{A}$ for $\forall \mathcal{A} \in \mathbb{T}$. Therefore $\mathcal{E}_{i_1 i_2 j_1 j_2} = \delta_{i_1 j_1} \delta_{i_2 j_2}$ is the identity element for $*_2$ on \mathbb{T} .

Finally, we know that $f^{-1}(\mathbf{I}^{I_1 I_2 \times I_1 I_2}) = \mathcal{E}$.

(A3) Show that for each $\mathcal{A} \in \mathbb{T}$, there exists an inverse $\tilde{\mathcal{A}}$ such that $\tilde{\mathcal{A}} *_2 \mathcal{A} = \mathcal{A} *_2 \tilde{\mathcal{A}} = \mathcal{E}$.

We define $\tilde{\mathcal{A}} = f^{-1}\{[f(\mathcal{A})]^{-1}\}$ since $f(\mathcal{A}) \in \mathbb{M}$ and f^{-1} is a bijective map from Lemma (3.1). Then,

$$f(\tilde{\mathcal{A}} *_2 \mathcal{A}) = f(\tilde{\mathcal{A}}) \cdot f(\mathcal{A}) = [f(\mathcal{A})]^{-1} \cdot f(\mathcal{A}) = \mathbf{I}^{I_1 I_2 \times I_1 I_2}.$$

From Lemma 3.1 and since $f(\mathcal{E}) = \mathbf{I}^{I_1 I_2 \times I_1 I_2}$, we obtain $\tilde{\mathcal{A}} *_2 \mathcal{A} = \mathcal{E}$. Similarly, we can get $\mathcal{A} *_2 \tilde{\mathcal{A}} = \mathcal{E}$.

It follows that for each $\mathcal{A} \in \mathbb{T}$, there exists an inverse $\tilde{\mathcal{A}}$ such that $\tilde{\mathcal{A}} *_2 \mathcal{A} = \mathcal{A} *_2 \tilde{\mathcal{A}} = \mathcal{E}$.

Therefore, the ordered pair $(\mathbb{T}, *_2)$ is a group where the operation $*_2$ is defined in (2.2). In addition, the transformation $f : \mathbb{T} \rightarrow \mathbb{M}$ in (2.4) is a bijective mapping between groups. Hence, f is an isomorphism. \square

Acknowledgments. C.N. would like to thank Shannon Starr for some fruitful discussions on the quantum models. The authors are indebted to the referees whose suggestions and corrections tremendously improved this manuscript.

REFERENCES

- [1] P.W. ANDERSON, *Absence of diffusion in certain random lattices*, Phys. Rev., 109 (1958), pp. 1492–1505.
- [2] Z. BAI, W. CHEN, R. SCALETTAR, AND I. YAMAZAKI, *Numerical methods for quantum Monte Carlo simulations of the Hubbard model*, in Multi-Scale Phenomena in Complex Fluids, T.Y. Hou, C. Liu, and J.-G. Liu, eds., Higher Education Press, Beijing, China, 2009, pp. 1–114.
- [3] G. BEYLKIN AND M.J. MOHLENKAMP, *Numerical operator calculus in higher dimensions*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 10246–10251.
- [4] G. BEYLKIN AND M.J. MOHLENKAMP, *Algorithms for numerical analysis in high dimensions*, SIAM J. Sci. Comput., 26 (2005), pp. 2133–2159.
- [5] K. BRAMAN, *Third order tensors as linear operators on a space of matrices*, Linear Algebra Appl., 433 (2010), pp. 1241–1253.
- [6] D.S. BURDICK, L.B. MCGOWN, D.W. MILLICAN, AND X.M. TU, *Resolution of multicomponent fluorescent mixtures by analysis of the excitation-emission-frequency array*, J. Chemometrics, 4 (1990), pp. 15–28.
- [7] R. CARMONA, A. KLEIN, AND F. MARTINELLI, *Anderson localization for Bernoulli and other singular potentials*, Comm. Math. Phys., 108 (1987), pp. 41–66.
- [8] J.D. CARROLL AND J.J. CHANG, *Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition*, Psychometrika, 35 (1970), pp. 283–319.
- [9] H. COHN, R. KLEINBERG, B. SZEGEDY, AND C. UMANS, *Group-theoretic algorithms for matrix multiplication*, in Proceedings of the 46th Annual Symposium on Foundations of Computer Science, Pittsburgh, PA, IEEE Computer Society, 2005, pp. 379–388.
- [10] P. COMON, *Tensor decompositions: State of the art and applications*, in Mathematics in Signal Processing V, J.G. McWhirter and I.K. Proudler, eds., Oxford University Press, Oxford, 2001, pp. 1–24.
- [11] P. COMON, G. GOLUB, L.-H. LIM, AND B. MOURRAIN, *Symmetric tensors and symmetric tensor rank*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1254–1279.
- [12] D. COPPERSMITH AND S. WINOGRAD, *Matrix multiplication via arithmetic progressions*, J. Symbolic Comput., 9 (1990), pp. 251–280.
- [13] E. CUTHILL AND J. MCKEE, *Reducing the bandwidth of sparse symmetric matrices*, in Proceedings of the 24th ACM National Conference, New York, 1969, pp. 157–172.

- [14] V. DA SILVA AND L.-H. LIM, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1084–1127.
- [15] L. DE LATHAUWER, *A Survey of Tensor Methods*, ISCAS, Taipei, 2009.
- [16] L. DE LATHAUWER, B. DE MOOR, AND J. VANDEWALLE, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1253–1278.
- [17] L. DE LATHAUWER, B. DE MOOR, AND J. VANDEWALLE, *On the best rank-1 and rank- (R_1, R_2, \dots, R_N) approximation of higher-order tensors*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1324–1342.
- [18] L. DE LATHAUWER, J. CASTAING, AND J.-F. CARDOSO, *Fourth-order cumulant-based blind identification of underdetermined mixtures*, IEEE Trans. Signal Process., 55 (2007), pp. 2965–2973.
- [19] J. DEMMEL, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.
- [20] J. DEMMEL, *Lecture Notes on Cache Blocking*, University of California Berkeley, Berkeley, CA, http://www.cs.berkeley.edu/~demm/lectures/Lect_02_1999b.pdf (2007).
- [21] A. DOOSTAN, G. IACCARINO, AND N. ETEMADI, *A least-squares approximation of high-dimensional uncertain systems*, in Center for Turbulence Research, Annual Research Briefs, Stanford University, Stanford, CA, 2007, pp. 121–132.
- [22] A. EINSTEIN, *The foundation of the general theory of relativity*, in The Collected Papers of Albert Einstein 6, A.J. Kox, M.J. Klein, and R. Schulmann, eds., Princeton University Press, Princeton, NJ, 2007, pp. 146–200.
- [23] L. ERDÖS, M. SALMHOFER, AND H.-T. YAU, *Towards the quantum Brownian motion*, in Mathematical Physics of Quantum Mechanics, Lecture Notes in Physics 690, J. Asch and A. Joyé, eds., Springer, Berlin, 2006, pp. 233–258.
- [24] J.A. GEORGE AND J.W.-H. LIU, *Computer Solution of Large Sparse Positive Definite Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1981.
- [25] W.H. GREUB, *Multilinear Algebra*, Springer-Verlag, Berlin, 1967.
- [26] W. HACKBUSCH AND B.N. KHOROMSKIJ, *Tensor-product approximation to operators and functions in high dimensions*, J. Complexity, 23 (2007), pp. 697–714.
- [27] W. HACKBUSCH, B.N. KHOROMSKIJ, AND E.E. TYRTYSHNIKOV, *Hierarchical kronecker tensor-product approximations*, J. Numer. Math., 13 (2005), pp. 119–156.
- [28] R.A. HARSHMAN, *Foundations of the PARAFAC procedure: Models and conditions for an “explanatory” multi-modal factor analysis*, UCLA Working Papers in Phonetics, 16 (1970), pp. 1–84.
- [29] F.L. HITCHCOCK, *The expression of a tensor or a polyadic as a sum of products*, J. Math. Phys., 6 (1927), pp. 164–189.
- [30] F.L. HITCHCOCK, *Multiple invariants and generalized rank of a p -way matrix or tensor*, J. Math. Phys., 7 (1927), pp. 39–79.
- [31] D. HUNDETRMARK, *A short introduction to Anderson localization*, in Analysis and Stochastics of Growth Processes and Interface Models, P. Mörters, R. Penrose, H. Schwetlick, and J. Zimmer, eds., Oxford University Press, Oxford, 2008, pp. 194–218.
- [32] G.E. ICE, B.C. LARSON, W. YANG, J.D. BUDAI, J.Z. TISCHLER, J.W.L. PANG, R.I. BARABASH, AND W. LIU, *Polychromatic X-ray microdiffraction studies of mesoscale structure and dynamics*, J. Synchrotron Rad., 12 (2005), pp. 155–162.
- [33] G.E. ICE AND B.C. LARSON, *Three-dimensional X-ray structural microscopy using polychromatic microbeams*, MRS Bulletin, 29 (2004), pp. 170–176.
- [34] B.N. KHOROMSKIJ, *Tensor-Structured Numerical Methods in Scientific Computing: Survey on Recent Advances*, Preprint 21/2010, MPI MIS, Leipzig, 2010.
- [35] B.N. KHOROMSKIJ, V. KHOROMSKAIA, AND H.-J. FLAD, *Numerical solution of the Hartree-Fock equation in multilevel tensor-structured format*, SIAM J. Sci. Comput., 33 (2011), pp. 45–65.
- [36] T. KOLDA AND B.W. BADER, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500.
- [37] T. KOLDA, *Orthogonal tensor decompositions*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 243–255.
- [38] W. KIRSCH, *An invitation to random Schrödinger operators*, in Random Schrödinger Operators, Panoramas et Synthèses 25, Société Mathématique de France, Paris, 2008, pp. 1–119.
- [39] J.B. KRUSKAL, *Three-way arrays: Rank and uniqueness of trilinear decompositions with applications to arithmetic complexity and statistics*, Linear Algebra Appl., 18 (1977), pp. 95–138.
- [40] H. KUNZ AND B. SOUILLARD, *Sur le spectre des opérateurs aux différences finies aléatoires*, Comm. Math. Phys., 78 (1980), pp. 201–246.
- [41] W.M. LAI, D. RUBIN, AND E. KREMLP, *Introduction to Continuum Mechanics*, Butterworth-Heinemann, Oxford, 2009.

- [42] C.D. MARTIN AND C.F. VAN LOAN, *A Jacobi-type method for computing orthogonal tensor decompositions*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1219–1232.
- [43] B. NAZER AND R.D. NOWAK, *Sparse interactions: Identifying high-dimensional multilinear systems via compressed sensing*, in Proceedings of the 48th Annual Allerton Conference on Communication, Control and Computation, Monticello, IL, 2010, pp. 1589–1596.
- [44] S. RAGNARSSON AND C. VAN LOAN, *Block tensor unfoldings*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 149–169.
- [45] S. RAGNARSSON AND C. VAN LOAN, *Block tensor and symmetric embeddings*, Linear Algebra Appl., 438, pp. 853–874.
- [46] M. REED AND B. SIMON, *Methods of Modern Mathematical Physics I: Functional Analysis*, Academic Press, New York, 1980.
- [47] N.D. SIDIROPOULOS AND R. BRO, *On the uniqueness of multilinear decomposition of N-way arrays*, J. Chemometrics, 14 (2000), pp. 229–239.
- [48] A. SMILDE, R. BRO, AND G. GIANNAKIS, *Multi-way Analysis. Applications in the Chemical Sciences*, John Wiley and Sons, Chichester, UK, 2004.
- [49] G. STOLZ, *An introduction to the mathematics of Anderson localization*, Contemp. Math., 552 (2010), pp. 71–108.
- [50] V. STRASSEN, *Gaussian elimination is not optimal*, Numer. Math., 13 (1969), pp. 354–356.
- [51] L.R. TUCKER, *The extension of factor analysis to three-dimensional matrices*, in Contributions to Mathematical Psychology, H. Gulliksen and N. Frederiksen, eds., Holt, Rinehardt, & Winston, New York, 1963, pp. 110–127.
- [52] L.R. TUCKER, *Some mathematical notes on three-mode factor analysis*, Psychometrika, 31 (1966), pp. 279–311.
- [53] M.A.O. VASILESCU AND D. TERZOPOULOS, *Multilinear subspace analysis for image ensembles*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 03), Madison, WI, 2003, pp. 93–99.
- [54] J.P. WOLD, R. BRO, A. VEBERG, F. LUNDBY, A.N. NILSEN, AND J. MOAN, *Active photosensitizers in butter detected by fluorescence spectroscopy and multivariate curve resolution*, J. Agric. Food Chem., 54 (2006), pp. 10197–10204.