Approximation of Images via Generalized Higher Order Singular Value Decomposition over Finite-dimensional Commutative Semisimple Algebra

Liang Liao[1], Lun Li[2], Sen Lin[1], Xiuwei Zhang[3], Xin Liu[4], Song Zhao[5], Yan Wang[1], Xinqiang Wang[1], Qi Gao[1], Jingyu Wang[1]

[1] School of Electronics and Information, Zhongyuan University of Technology, China

[2] School of Information Engineering, Zhengzhou University, China

[3] National Engineering Laboratory for Integrated Aerospace-Ground-Ocean Big Data Application Technology, China

[4] Information Center, Yellow River Conservancy Comission, China

[5] School of Intelligent Engineering, Zhengzhou University of Aeronautics, China

liaoliang@ieee.org (or liaoliangis@126.com)

***Abstract*—Low-rank approximation of images via singular value decomposition is well-received in the era of big data. However, singular value decomposition (SVD) is only for order-two data, i.e., matrices. It is necessary to flatten a higher order input into a matrix or break it into a series of order-two slices to tackle higher order data such as multispectral images and videos with the SVD. Higher order singular value decomposition (HOSVD) extends the SVD and can approximate higher order data using sums of a few rank-one components. We consider the problem of generalizing HOSVD over a finite dimensional commutative algebra. This algebra, referred to as a t-algebra, generalizes the field of complex numbers. The elements of the algebra, called t-scalars, are fix-sized arrays of complex numbers. One can generalize matrices and tensors over t-scalars and then extend many canonical matrix and tensor algorithms, including HOSVD, to obtain higher-performance versions. The generalization of HOSVD is called THOSVD. Its performance of approximating multi-way data can be further improved by an alternating algorithm. THOSVD also unifies a wide range of principal component analysis algorithms. To exploit the potential of generalized algorithms using t-scalars for approximating images, we use a pixel neighborhood strategy to convert each pixel to a "deeper-order" t-scalar. Experiments on publicly available images show that the generalized algorithm over t-scalars, namely THOSVD, compares favorably with its canonical counterparts.**

***Index Terms*—higher order singular value decomposition, image analysis, alternating optimization, low-rank approximation, generalization**

I. Introduction

A. Backgrounds

In the era of information deluge, data are often multi-way arrays. For example, numerical multispectral or hyperspectral images and videos are all in a multi-way format.

Concerning the spatial properties of multi-way data, many authors studied tensorial algorithms rather than mere matrix algorithms in data analysis. Hong et al. apply singular value decomposition (SVD) to image recognition and use the results of SVD to represent the features of images [1]. Then, many scholars have accepted the SVD as an essential tool for analyzing images and applied it to image compression [2], [3], watermarking [4], [5], and low-rank approximation [6], [7], to name a few.

However, SVD does not directly involve multi-way data in a non-matrix format. Tensor algorithms, backed by multilinear algebras are more effectively than the SVD in capturing multi-way structural information.

Tensors and their applications have developed since the last century. One pioneer of tenor decompositions and their applications is F. L. Hitchcock [8]. After Hitchcock, Tucker et al. studied tensor decomposition further and proposed the well-received Tucker decomposition algorithms [9]–[11]. Since Tucker's work, many authors have developed various tensor algorithms with applications, including but not limited to signal processing [12], computer vision [13] and image analysis [14], [15].

Among the Tucker-decomposition-based algorithms extending the SVD, higher order singular value decomposition (HOSVD) advertised by Lathauwer et al. is well received [16], [17]. HOSVD can be used for applications such as image fusion [18], [19], image denoising [20] and target detection [21].

However, HOSVD is a higher-order generalization of SVD rather than a "deeper order" generalization. Here, "deeper order" means replacing the complex numbers by the elements of a finite dimensional algebra. Matrices and tensors with entries in a finite dimensional algebra have been investigated. For example, many authors have constructed matrix algorithms over Hamilton's quaternions. Quaternions form a finite-dimensional non-commutative algebra generalizing the field of complex numbers. Quaternion matrices and tensors are used in histopathological image analysis [22], image & video recovery and representations [23]–[25], color image denoising [26], [27], color face recognition [28], [29], color image inpainting [30], [31], to name a few.

However, the multiplication of quaternions is not commutative. The non-commutativity prevents prevents quaternionic matrices from being straightforward analogues of matrices with complex entries. Further, quaternions are only suitable for characterizing objects with four or fewer components. There is no consistent and straightforward extension to objects with five or more dimensions. One might query the existence of a finite-dimensional algebra generalizing complex numbers and allowing a consistent and straightforward extension to higher dimensions? More concretely, can one find matrices or tensors

with fix-sized multi-way arrays as entries, rather than just complex numbers? If yes, one can represent the generalized matrices and tensors by multi-way arrays where the multi-way is two-fold, i.e., a combination of higher-order and deeper-order. Higher-order characterizes their superstructure (rows, columns, and beyond), and deeper order characterizes their substructure (i.e., entries).

Concerning this issue, Kilmer et al. propose the "t-product" model, establishing a finite-dimensional algebra of fix-sized order-one arrays of real numbers. The t-product model establishes matrices over the algebra. The generalized matrices are order-three arrays of real numbers, two orders (higher-order) for indexing rows and columns, the third (deeper-order) used to characterize generalized entries (i.e., t-scalars) [32], [33].

Liao and Maybank extend [15], [34], [35] Kilmer et al.'s generalized scalars from the one-way (i.e., order-one) to multi-way and from real to complex, calling the finite-dimensional commutative algebra a "t-algebra" and the generalized matrices "t-matrices". Using t-matrices instead of their canonical counterparts, one can generalize many matrix algorithms straightforwardly.

This paper validates the t-matrix model and interprets the model via the standard matrix representation theory. It shows that one can generalize many canonical tensor algorithms, such as Higher Order Singular Value Decomposition (HOSVD), by replacing scalars with t-scalars. The generalized tensor algorithm compares favorably with its canonical counterparts for approximating multi-way images.

B. Organization of this paper

The rest of this paper is organized as follows. Section II briefly introduces the basic notions and rationale of canonical tensors and HOSVD, laying the foundations for the generalization over t-algebra.

Section III introduces the finite-dimensional commutative algebra called t-algbera by Liao and Maybank [34]. The elements of the t-algebra are fix-sized arrays generalizing complex numbers and enabling one to build generalized vectors, matrices, and even tensors with entries in the t-algebra. These generalized scalars, vectors, and matrices are called t-scalars, t-vectors, and t-matrices. The standard matrix representation of t-scalars, t-vectors, and t-matrices is also discussed in this section. Then standard t-matrix representation is the basis for many linearity issues for modules over the t-algebra and connects them to their t-scalar-valued versions.

A neighborhood strategy to convert a lower-order image is proposed in Section IV. The neighbourhood strategy exploits the potential of the generalized complex numbers, called t-scalars, and characterizes the spatial constraints of an image.

Section V introduces the generalized tensors over the t-algebra, called g-tensors, and the generalized higher-order singular value decomposition, called THOSVD. THOSVD generalizes the canonical higher-order singular value decomposition (HOSVD) and unifies a wide range of algorithms for principal components analysis. Section V also discusses the generalization of the higher-order orthogonal iteration algorithm HOOI over a t-algebra in order to optimize the performance of THOSVD.

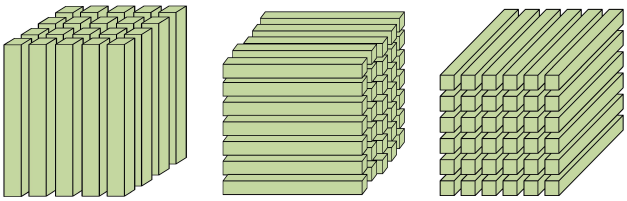
Section VI reports the experimental results on approximating public multi-way images and compares the performances of THOSVD and HOSVD and a wide range of PCA-based algorithms from a unified THOSVD perspective. Experiments show that the generalized algorithms over t-scalars compare favorably with their canonical counterparts over canonical scalars. Finally, we conclude this paper in Section VII.

II. Canonical tensors and higher order singular value decomposition

A. Fundamental notions of canonical tensors

By canonical matrices or canonical tensors, or sometimes just matrices or tensors, we mean the classical matrices or tensors with real or complex numbers as entries. Tensors generalize matrices, vectors, and scalars.

The basic notions can be explained with order-three tensors. Let $\mathcal{X}$ be an order-three tensor in $\mathbb{C}^{D_1\times D_2\times D_3}$. Then, $\mathcal{X}$ has three modes of fibers, i.e., column fibers, row fibers, and tube fibers, respectively obtained by fixing all but one of its indices. If all but two indices are fixed, one has three modes of slices, i.e., horizontal slices, lateral slices and frontal slices Figure 1 shows the column fibers, row fibers, and tube fibers of an order-three tensor.



(a) (b) (c)

Fig. 1. Three modes of fibers of a canonical order-three tensor (a) column fibers (b) row fibers (c) tube fibers

One can also unfold a higher-order tensor to a matrix. More specifically, the mode-k unfolding of a tensor $\mathcal{X}\in\mathbb{C}^{D_1\times\cdots\times D_M}$ is a matrix $X_{(k)}\in\mathbb{C}^{D_k\times D}$ where $D\doteq\frac{\prod_{m=1}^{M}D_m}{D_k}$ for all $k\in[M]\doteq\{1,2,\cdots,M\}$ such that the $(d_1,\cdots,d_M)$-th scalar entry of $\mathcal{X}$ maps the (d_k,j)-th scalar entry of $X_{(k)}$, namely

$$(\mathcal{X})_{d_1,\cdots,d_M}=(X_{(k)})_{d_k,j}\in\mathbb{C} \tag{1}$$

and the following condition holds for all $j\in[D]$,

$$j=1+\sum\nolimits_{m\in[M]\setminus\{k\}}(d_m-1)\cdot J_m \tag{2}$$

where $J_m\doteq\prod_{n\in[m-1]\setminus\{k\}}D_n$.

Namely, the mode-k unfolding $X_{(k)}$ of a tensor has the mode-k fibers as its columns. For example, let $\mathcal{X}$ be a tensor in $\mathbb{C}^{3\times4\times5}$. The mode-1 unfolding is a 3×20 matrix. The mode-2 unfolding is a 4×15 matrix. The mode-3 unfolding is a 5×12 matrix.

Given a tensor $\mathcal{X}\in\mathbb{C}^{D_1\times\cdots\times D_M}$ and a matrix $Y\in\mathbb{C}^{J\times D_k}$, the mode-$k$ multiplication $\mathcal{A}\doteq\mathcal{X}\times_k Y$ is a tensor in $\mathbb{C}^{D_1\times\cdots\times D_{k-1}\times J\times D_{k+1}\times\cdots\times D_M}$ such that the following condition holds for all $k\in[M]$,

$$A_{(k)}=Y\cdot X_{(k)}\ . \tag{3}$$

B. HOSVD: higher-order singular value decomposition

The higher order singular value decomposition (HOSVD) is a variant of the Tucker decomposition of a tensor and generalizes the canonical singular value decomposition (SVD).

Let $\mathcal{X}$ be a tensor in $\mathbb{C}^{D_1\times D_2\times\cdots\times D_M}$. HOSVD decomposes $\mathcal{X}$ as the following multi-mode multiplication

$$\mathcal{X}=\mathcal{S}\times_1 U_1\times_2 U_2\cdots\times_M U_M \tag{4}$$

where the $U_k \in \mathbb{C}^{D_k \times D_k}$ are unitary matrices for all $k \in [M]$, whose i-th column is the i-th leading left singular vector of the mode-k unfolding of $\mathcal{X}$ for all $i \in [D_k]$, and $\mathcal{S} \in \mathbb{C}^{D_1 \times \cdots \times D_M}$, called core tensor of $\mathcal{X}$, is computed as follows

$$\mathcal{S} = \mathcal{X} \times_1 U_1^* \times_2 U_2^* \cdots \times_M U_M^* \tag{5}$$

where the U_k^* denotes the Hermitian transpose of U_k for all $k \in [M]$.

Let $(r_1, r_2, \cdots, r_M)$ be a rank tuple subject to $r_k \leqslant D_k$ for all $k \in [M]$. When $\mathcal{S}$ and $U_1, \cdots, U_M$ are computed, the approximation of $\mathcal{X}$ is given as follows

$$\hat{\mathcal{X}} = \hat{\mathcal{S}} \times_1 \hat{U}_1 \times_2 \hat{U}_2 \cdots \times_M \hat{U}_M \tag{6}$$

where $\hat{\mathcal{S}} \in \mathbb{C}^{r_1 \times \cdots \times r_M}$ denotes the sub-core tensor containing the leading $r_1 \times \cdots \times r_M$ entries of $\mathcal{S}$ and $\hat{U}_k \in \mathbb{C}^{D_k \times r_k}$ denotes the sub-matrix containing the leading r_k columns of U_k for all $k \in [M]$.

The approximation of the tensor $\mathcal{X}$ can also be written as the following multi-mode multiplication

$$\hat{\mathcal{X}} = \mathcal{X} \times_1 P_1 \times_2 P_2 \cdots \times_M P_M \tag{7}$$

where $P_k \doteq \hat{U}_k \cdot \hat{U}_k^* \in \mathbb{C}^{D_k \times D_k}$ denotes the k-th idempotent matrix where $\text{rank}(P_k) = r_k$ holds for all $k \in [M]$.

III. T-scalars and t-algebra

A. Background and notions

Many matrix algorithms or tensor algorithms, including HOSVD, can be straightforwardly generalized to deeper-order versions with elements in a finite dimensional algebra called a t-algebra. We briefly introduce the t-algebra, which is finite-dimensional, commutative, and generalizes complex numbers.

The genesis of the t-algebra is Kilmer et al. s' t-product model [32], [33], in which the generalized scalars are fix-sized order-one real arrays, arranged as the tube fibers of order-three arrays. Kilmer et al. don't explicitly call these tubal fibers "generalized scalars" but instead refer to the order-three arrays with these particularly-purposed tubal fibers as just tensors. However, these "tensors" are essentially matrices of generalized scalars. One can add or multiply these generalized scalars (i.e., fix-sized order-one real arrays) or multiply any generalized scalar with a real or complex number.

Kilmer et al. [32], [33] chose a circular matrix formulation to describe the properties of generalized scalars, matrices, and tensors. In the t-matrix model [34], [35], Liao and Maybank extended Kilmer et al.'s generalized scalars from real one-way arrays to complex multi-way arrays. These generalized multi-way scalars, called t-scalars, form a finite-dimensional commutative algebra, generalizing complex numbers and reserving many well-known properties of complex numbers. For example, one can compute the conjugate, the "real" and the "imaginary" parts of a t-scalar. One can also compute the modulus of a t-scalar, which is nonnegative and self-conjugate. A partial ordering is defined for comparing self-conjugate t-scalars. Liao and Maybank also define generalized matrices, called t-matrices, over t-scalars. The t-matrix operations are analogous to the corresponding canonical matrix operations, but the underlying operations follow the rules defined for t-scalars.

In this paper, we follow the notations, protocols, and symbols in [34], [35] as much as possible. For example, all indices begin from 1 rather than 0, and different subscripts other than fonts are used to denote the data types over t-scalars. Some notations and their descriptions are given in Table I. For those notations not appearing in Table I, we give their descriptions in context, as necessary.

TABLE I
Some notations defined over t-scalars

Notations	Descriptions
$C \equiv \mathbb{C}^{I_1 \times \cdots \times I_N}$	t-algebra of t-scalars
$X_T \in C$	a t-scalar in C
$Z_T, E_T \in C$	zero t-scalar, identity t-scalar
Z_{TM}, I_{TM}	zero t-matrix, identity t-matrix
$X_T + Y_T \in C$	t-scalar addition
$X_T \circ Y_T \in C$	t-scalar multiplication
$X_{TM} \in C^{D_1 \times D_2}$	a t-matrix in $C^{D_1 \times D_2}$
X_{TM}^*	conjugate tranpose of a t-matrix X_{TM}
$X_{TM} + Y_{TM}$	t-matrix addition
$X_{TM} \circ Y_{TM}$	t-matrix multiplication
$\tilde{X}_T \doteq F(X_T)$	multi-way spectral transform of a t-scalar X_T
$\tilde{Y}_{TM} \doteq F(Y_{TM})$	multi-way spectral transform of a t-matrix Y_{TM}
$(X_T)_{i_1, \cdots, i_N} \in \mathbb{C}$	$(i_1, \cdots, i_N)$-th complex entry of a t-scalar $X_T \in C$
$[X_{TM}]_{d_1, d_2} \in C$	(d_1, d_2)-th t-scalar entry of a t-matrix X_{TM}
S^{nonneg}	a set of nonnegative t-scalars
$\|X_{TM}\|_{t,F} \in S^{nonneg}$	t-scalar-valued Frobenius norm of a t-matrix X_{TM}, i.e., a nonnegative t-scalar
$X_T \leq Y_T$ where $X_T, Y_T \in S^{nonneg}$	nonnegative t-scalars satisfying partial order relations "$\leq$"
$\text{rank}_t\, X_{TM} \geq Z_T$	t-scalar-valued rank of a t-matrix, i.e., a nonnegative t-scalar
$X_{GT} \in C^{D_1 \times \cdots \times D_M}$	a g-tensor in $C^{D_1 \times \cdots \times D_M}$

B. T-scalars in a nutshell

T-scalars are fix-sized multi-way arrays of complex numbers. One can add and multiply any pair of these arrays and multiply any of these arrays with a complex number. The t-algebra is the set of these fix-sized arrays.

Let $C \equiv \mathbb{C}^{I_1 \times \cdots \times I_N}$ be the t-algebra whose elements are $I_1 \times \cdots \times I_N$ complex arrays, called t-scalars. A t-scalar is defined by a linear spectral transform and the Hadamard product. Let F be a bijective linear spectral transform. For example, F can be the discrete Fourier transform, discrete cosine transform, or other bijective linear transforms. Then, given any t-scalar $X_T \in C \equiv \mathbb{C}^{I_1 \times \cdots \times I_N}$, its transform $\tilde{X}_T \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ is given by the following multi-mode multiplication

$$\tilde{X}_T \doteq F(X_T) = X_T \times_1 W_1 \cdots \times_N W_N \tag{8}$$

where W_k is a full rank matrix in $\mathbb{C}^{I_k \times I_k}$, i.e., the one-way spectral transform that defines the spectral transform F.

The inverse transform is given by

$$X_T \doteq F^{-1}(\tilde{X}_T) = \tilde{X}_T \times_1 W_1^{-1} \cdots \times_N W_N^{-1} \tag{9}$$

where, for example, W_k can be the $I_k \times I_k$ Fourier-transform matrix or the cosine-transform matrix such that the (m_1, m_2)-th entry of W_k is defined as follow

$$\text{Fourier: } (W_k)_{m_1, m_2} = \exp\left[\frac{-2\pi i}{I_k}(m_1 - 1)(m_2 - 1)\right], \tag{10}$$

$$\text{Cosine: } (W_k)_{m_1, m_2} = \cos\left[\frac{\pi}{I_k}(m_1 - 1)(m_2 - 1/2)\right]. \tag{11}$$

Let $X_T \circ Y_T \in C$ be the multiplication of any two t-scalars X_T and Y_T. By the t-scalar multiplication defined in [34], the following condition holds for all $X_T, Y_T \in C$,

$$F(X_T \circ Y_T) = F(X_T) \odot F(X_T) \tag{12}$$

where $\odot$ is the Hadamard product (i.e., entry-wise product) of two complex arrays of the same size.

It is not difficult to show that the t-algebra C is a commutative semisimple algebra. A non-trivial t-algebra C is semisimple if it is isomorphic to the direct sum of K copies of the field of complex numbers, where

$$K \doteq I_1 \cdots I_N \tag{13}$$

denotes the dimension of C. When $I_1 = \cdots = I_N = 1$, C reduces to the field $\mathbb{C}$ [34], [35].

It is noted by equation (12) that a specific transform defines a corresponding t-scalar multiplication. One can use different transforms to get different t-scalar multiplications.

When the Fourier transform is chosen, the corresponding t-scalar multiplication is called circular convolution [36], which is becoming popular in convolutional networks [37]–[42] in recent years.

C. Matrix representation of t-scalars

Each t-scalar represents a linear operator from C to C since the following condition holds for all $X_T, A_T, B_T \in C$ and $\alpha, \beta \in \mathbb{C}$,

$$X_T \circ \big(\alpha \cdot A_T + \beta \cdot B_T\big) = \alpha \cdot (X_T \circ A_T) + \beta \cdot (X_T \circ B_T) \in C. \tag{14}$$

Thus, each t-scalar is representable by a square matrix in $C^{K \times K}$ (where $K \doteq \dim C = I_1 \cdots I_N$), and has K eigenvalues in $\mathbb{C}$, counting multiplicity.

Let E_T be the identity t-scalar in C such that $E_T \circ X_T = X_T$ holds for all $X_T \in C$. By equation (12), the following identity holds for all $X_T \in C$,

$$F(E_T) \odot F(X_T) = F(X_T) \tag{15}$$

or equivalently, $F(E_T)$ is an array of ones.

A scalar $\lambda \in \mathbb{C}$ is an eigenvalue of a t-scalar $X_T \in C$ if and only if the t-scalar $(X_T - \lambda \cdot E_T)$ is not multiplicatively invertible, or equivalently, the transform

$$F(X_T - \lambda \cdot E_T) \equiv F(X_T) - \lambda \cdot F(E_T) \in \mathbb{C}^{I_1 \times \cdots \times I_N} \tag{16}$$

contains at least one entry equal to zero.

Because $\lambda \cdot F(E_T)$ is an array of λ's, it is equivalent to say that λ is an eigenvalue of X_T if and only if λ is equal to an entry of $F(X_T)$. Specifically, let the entries of $F(X_T)$ be $F_1(X_T), \cdots, F_K(X_T) \in \mathbb{C}$ for any t-scalar $X_T \in C$. Then, these entries of $F(X_T)$ are the K eigenvalues of $X_T \in C$, counting multiplicity.

Because of the set of eigenvalues defines the spectrum of a (bounded) linear operator, and F gives all eigenvalues of any t-scalar $X_T \in C$, we call the mapping F a spectral transform.

Furthermore, any t-scalar $X_T \in C$ is representable by the diagonal matrix formed by its eigenvalues, namely

$$X_T \sim \operatorname{diag}\big[F_1(X_T), \cdots, F_K(X_T)\big] \in \mathbb{C}^{K \times K}\ . \tag{17}$$

If someone uses the Fourier-transform defining F, but another uses the cosine transform defining F, they usually have different diagonal matrices representing the same raw t-scalar. Since the spectral transform F is always determined prior to the genesis of t-scalars, the eigenvalues and the diagonal matrix representation of any t-scalar $X_T \in C$ are uniquely determined.

It is not difficult to follow that all $K \times K$ diagonal matrices form a commutative semisimple subalgebra of the matrix algebra of all $K \times K$ complex matrices. This subalgebra of all $K \times K$ diagonal complex matrices is isomorphic to the t-algebra C. There is one-to-one mapping between any t-scalar X_T and its diagonal matrix representation.

For example, the zero t-scalar Z_T is a $I_1 \times \cdots \times I_N$ array of zeros, its diagonal matrix representation is the $K \times K$ zero matrix. The matrix representatio of the identity t-scalar E_T is the $K \times K$ identity matrix. Let

$$M(X_T) \doteq \operatorname{diag}[F_1(X_T), \cdots, F_K(X_T)] \tag{18}$$

be the diagonal matrix representation of any t-scalar X_T. Then, by the algebra isomorphism mentioned previously, the following relationships hold for all $X_T, Y_T \in C$ and $\alpha, \beta \in \mathbb{C}$, under the t-scalar operations and the usual matrix operations,

$$\begin{aligned} X_T \circ Y_T &\sim\ M(X_T) \cdot M(Y_T)\ , \\ \alpha \cdot X_T + \beta \cdot Y_T &\sim\ \alpha \cdot M(X_T) + \beta \cdot M(Y_T)\ . \end{aligned} \tag{19}$$

Since each t-scalar is representable by a diagonal matrix, one can use the matrix theory to explain many notions of t-scalars.

For example, the conjugate X_T^* of a t-scalar $X_T \in C$, generalizing the conjugate of a complex number, is representable by the Hermitian transpose of the corresponding diagonal matrix. More specifically, the following representation holds

$$X_T^* \sim M(X_T^*) = \operatorname{diag}[\,\overline{F_1(X_T)}, \cdots, \overline{F_K(X_T)}\,] \tag{20}$$

where $\overline{F_k(X_T)} \in \mathbb{C}$ denotes the complex conjugate of $F_k(X_T) \in \mathbb{C}$ for all $k \in [K]$.

A t-scalar X_T is called self-conjugate if and only if $X_T = X_T^*$ holds. It is equivalent to say that the matrix $M(X_T)$ is conjugate symmetric, more specifically,

$$M(X_T) = M(X_T^*) = \operatorname{diag}[\overline{F_1(X_T)}, \cdots, \overline{F_K(X_T)}]\ . \tag{21}$$

In other words, all eigenvalues $F_1(X_T), \cdots, F_K(X_T)$ are real numbers.

The standard matrix representation gives a unique perspective to explain many t-scalar-valued notions in [34], [35]. We follow this perspective to explain our algorithms and formulations in the following sections.

D. T-matrix and block-diagonal matrix representation

A t-matrix is a rectangular array of t-scalars. Since t-scalars are arrays in $\mathbb{C}^{I_1 \times \cdots \times I_N}$, we choose the underlying data structure of a t-matrix in $C^{D_1 \times D_2}$ as an array in $\mathbb{C}^{I_1 \times \cdots \times I_N \times D_1 \times D_2}$. Because of the underlying multi-way array format, many authors call t-matrices tensors [32], [33]. However, they are different from (canonical) tensors with complex entries. For example, any t-matrix Y_{TM} in $C^{D_1 \times D_2}$ is representable by a block matrix $M(Y_{TM})$ in $C^{KD_1 \times KD_2}$, each block is a $K \times K$ diagonal matrix, representing the corresponding t-scalar.

Figure 2 shows the block-diagonal matrix representation $M(Y_{TM})$ of a t-matrix Y_{TM} in $C^{3\times 1} \equiv \mathbb{C}^{2\times 2\times 3\times 1}$. The block-diagonal matrix contains 3×1 blocks, each block a 4×4 diagonal matrix.

To get the matrix representation of a t-matrix Y_{TM}, one needs to compute the eigenvalues of each t-scalar entry, or equivalently, transform $Y_{TM} \in \mathbb{C}^{I_1 \times \cdots \times I_N \times D_1 \times D_2}$ by the N-way Fourier transform. The transform of a t-matrix Y_{TM} is

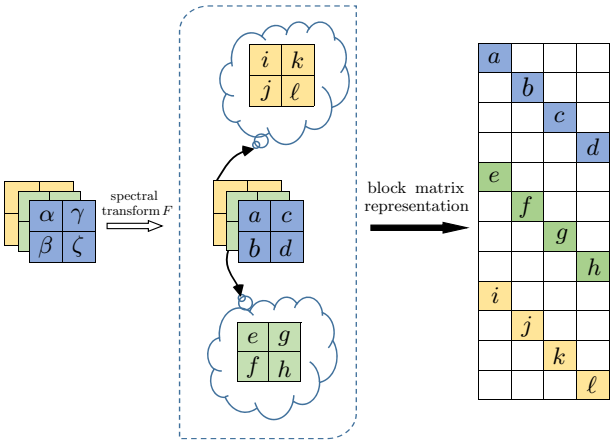


Fig. 2. A diagram of the block-diagonal matrix representation of a t-matrix $Y_{TM} \in C^{3\times 1} \equiv \mathbb{C}^{2\times 2\times 3\times 1}$. Left: a t-matrix Y_{TM} (i.e., a t-vector) with $I_1 \times I_2 = 2 \times 2$ and $D_1 \times D_2 = 3 \times 1$. Middle: transformed t-matrix $F(X_{TM})$. Right: a matrix $M(Y_{TM})$ containing $D_1 \times D_2 = 3 \times 1$ blocks, each block a 4×4 diagonal matrix.

obtained by transforming each t-scalar entry of Y_{TM}. The transformed t-matrix is denoted by $F(Y_{TM})$.

We choose the first N modes of the underlying complex array to accommodate the t-scalar entries of Y_{TM}. The formulations of $F(Y_{TM})$ and $F^{-1}(Y_{TM})$ are consistent with equations (8) and (9), namely

$$\begin{aligned} \tilde{Y}_{TM} &\doteq F(Y_{TM}) = Y_{TM} \times_1 W_1 \cdots \times_N W_N \ , \\ Y_{TM} &\doteq F^{-1}(\tilde{Y}_{TM}) = \tilde{Y}_{TM} \times_1 W_1^{-1} \cdots \times_N W_N^{-1} \ . \end{aligned} \tag{22}$$

The transformm $F(Y_{TM})$ is an array containing all eigenvalues of each t-scalar entry of the t-matrix Y_T. It is easy to convert $F(X_T)$ to the block-diagonal matrix representation $M(X_T)$ as shown in Figure 2.

Analogous to equation (19), the following representations hold for all $\alpha, \beta \in \mathbb{C}$, and t-matrices X_{TM}, Y_{TM} of compatible sizes,

$$\begin{aligned} X_{TM} \circ Y_{TM} &\sim M(X_{TM}) \cdot M(Y_{TM}) \ , \\ \alpha \cdot X_{TM} + \beta \cdot Y_{TM} &\sim \alpha \cdot M(X_{TM}) + \beta \cdot M(Y_{TM}) \ . \end{aligned} \tag{23}$$

The block-diagonal matrix representation can help define and explain many notions of the t-matrix model. For example, the determinant of a t-scalar (or a square t-matrix) is equal to the determinant of its matrix representation. The inverse or pseudo-inverse of a t-scalar (or t-matrix) is representable by the inverse or pseudo-inverse of its matrix representation.

E. T-scalar-valued notions

The matrix representation describes the (canonical) linear properties of t-matrices. Further properties of t-matrices are described in this section. For example, as defined in [34], one can multiply a t-matrix with a t-scalar. Such an operation, generalizing the multiplication of a complex matrix with a complex number, is an operation for characterizing the generalized linear asepct, called the C-linearity, of t-matrices.

More specifically, for all $A_T, B_T \in C$ and $X_{TM}, Y_{TM} \in C^{D_1\times D_2}$, the C-linear sum

$$W_{TM} \doteq A_T \circ X_{TM} + B_T \circ Y_{TM} \tag{24}$$

is a t-matrix in $C^{D_1\times D_2}$ such that the following condition holds

$$[W_{TM}]_{d_1,d_2} = A_T \circ [X_{TM}]_{d_1,d_2} + B_T \circ [Y_{TM}]_{d_1,d_2} \in C \tag{25}$$

where $[\cdot]_{d_1,d_2} : X_{TM} \mapsto [X_{TM}]_{d_1,d_2} \in C$ denotes the (d_1, d_2)-th t-scalar entry of a t-matrix for all $(d_1, d_2) \in [D_1] \times [D_2]$.

Note any linear sum of two t-matrices is just a special case a C-linear sum of the two t-matrices. Specifically, the following identity holds for all $\alpha, \beta \in C$ and $X_{TM}, Y_{TM} \in C^{D_1\times D_2}$,

$$\alpha \cdot X_{TM} + \beta \cdot Y_{TM} \equiv (\alpha \cdot E_T) \circ X_{TM} + (\beta \cdot E_T) \circ Y_{TM} \ . \tag{26}$$

The linearity is just a constrained property of the C-linearity. Therefore, the introduced matrix representation can not fully characterize C-modules. For example, a non-trivial t-scalar multiplication $A_T \circ X_{TM}$ can not be represented by the multplication of the matrices $M(A_T) \in \mathbb{C}^{K\times K}$ and $M(X_{TM}) \in \mathbb{C}^{KD_1\times KD_2}$. Such a multiplication of $M(A_T)$ and $M(X_T)$ is not defined in classical linear algebra.

The linearity and the C-linearity require both scalar-valued notions (for the linearity) and t-scalar-valued notions (for the C-linearity). Further, a t-scalar-valued notion should be reducible to its scalar-valued counterpart.

Before any detailed examples, let S^{nonneg} be the commutative semiring of nonnegative t-scalars. A t-scalar X_T is nonnegative if and only its matrix representation

$$M(X_T) \doteq \mathrm{diag}[F_1(X_T), \cdots, F_K(X_T)]$$

is positive semidefinite. It is equivalent to stating the diagonal entries $F_1(X_T), \cdots, F_K(X_T)$ are all nonnegative real numbers.

The above definition is equivalent to the original definition given in [34]. Namely, a t-scalar X_T is nonnegative if and only if a t-scalar Y_T exists such that $X_T = Y_T \circ Y_T^*$ holds.

It shows that all nonnegative t-scalars are also self-conjugate. The nonnegativity defines a partial order "$\leq$", i.e., a reflexive, antisysmmetric and transitive relationship among self-conjugate t-scalars. More specifically, given any self-conjugate t-scalars X_T and Y_T, the partial order $X_T \leq Y_T$ holds if and only if the t-scalar $(Y_T - X_T)$ is nonnegative. It also shows that the zero t-scalar Z_T is the least element in S^{nonneg}, namely $X_T \geq Z_T$ holds for all nonnegative t-scalar X_T.

Then, many notions are defined as a nonnegative t-scalar. For example, the generalized modulus $|X_T|_t$ of a t-scalar $X_T \in C$ is a nonnegative t-scalar [34] such that its matrix representation is given by

$$|X_T|_t \sim \mathrm{diag}\left(|F_1(X_T)|, \cdots |F_K(X_T)| \right) . \tag{27}$$

where the subscript "t" denotes that the corresponding term is t-scalar-valued, more specifically, a nonnegative t-scalar valued term.

For another example, the trace of a t-scalar X_T is the sum of the eigenvalues of X_T, namely

$$\mathrm{trace}\, X_T \doteq F_1(X_T) + \cdots + F_K(X_T) = \mathrm{trace}\, M(X_T) . \tag{28}$$

Then, one can use the t-scalar-valued modulus $|X_T|_t \in S^{noneneg}$ to define the (canonical) Frobenius norm of a t-scalar $X_T \in C$ by

$$\begin{aligned} \|X_T\|_F &\doteq \sqrt{\mathrm{trace}\, |X_T|_t^2} = \|M(X_T)\|_F \\ &= \sqrt{|F_1(X_T)|^2 + \cdots + |F_K(X_T)|^2} \ \geqslant \ 0 \ . \end{aligned} \tag{29}$$

If the spectral transform F is not isometric, the norm given by equation (29) is not equal to the usual Frobenius norm de-

fined in the spatial domain. Specifically, the following inequality usually holds if the spectral transform F is not isometric.

$$\|X_T\|_F \neq \sqrt{\sum_{i_1,\cdots,i_N} \left|(X_T)_{i_1,\cdots,i_N}\right|^2} \ . \tag{30}$$

Inequality (30) shows that the norm $\|X_T\|_F$ is defined in the transformed domain $F(C)$ rather than the spatial domain C.

Following equation (27), the t-scalar-valued Frobenius norm of a t-matrix X_{TM} is a nonnegative t-scalar given by

$$\|X_{TM}\|_{t,F} \doteq \sqrt{\sum\nolimits_{d_1,d_2} \left|[X_{TM}]_{d_1,d_2}\right|_t^2} \in S^{nonneg} \ . \tag{31}$$

Then, one can define the scalar-valued Frobenius norm of the t-matrix X_{TM} via the t-scalar-valued Frobenius norm as follows

$$\|X_{TM}\|_F \doteq \|M(X_{TM})\|_F = \sqrt{\text{trace}\,\|X_{TM}\|_{t,F}^2} \geqslant 0 \ . \tag{32}$$

Furthermore, if the transform F is not congruent, the norm given by equation (32) is not equal to the usual Frobenius norm of the underlying (canonical) tensor in $\mathbb{C}^{I_1\times\cdots\times I_N\times D_1\times D_2}$.

It is noted that the cosine transform given by equation (11) is isometric, but the Fourier transform is not.

IV. Pixel neighborhood strategy

A. Neighborhood strategy

The above theory provides a paradigm for analyzing multi-way data. This paradigm compares favorably with its canonical counterpart over (canonical) scalars [15], [34], [43]–[48].

To tackle usual images and exploit the potentials of the t-algebra and the t-matrix model, we use a small neighborhood of each pixel to obtain a deeper-order representation of a grayscale image.

A matrix can characterize a grayscale image. Let the size of the matrix be $D_1 \times D_2$. Each pixel has a pixel neighborhood with size $I_1 \times I_2$. If a pixel is at the image's border, at the border of the image, one can pad with "0" where necessary to get an $I_1 \times I_2$ neighborhood of the pixel.

Replacing each pixel by its neighborhood of pixels, one has a multi-way array representation for a grayscale image, i.e., a t-matrix in $C^{D_1\times D_2} \equiv \mathbb{C}^{I_1\times I_2\times D_1\times D_2}$. The use of a multi-way array improves the performance of generalized algorithms over t-scalars.

Figure 3 shows the 3×3 neighborhood strategy, which increases a pixel to an order-two array and then to an order-four array. In this figure, order-four arrays are illustrated in the form of block matrices. Each deeper-order representation of a pixel is specified by a t-scalar.

Given a fixed neighborhood size, there are different neighborhood variants, each with the input pixel at a different position in the neighborhood. For example, an input pixel can be at the center or at the corner of its neighborhood.

Figure 4 shows two types of 3×3 neighborhoods of a raw pixel: the "center" neighborhood and the "inception" neighborhood. In a "center" neighborhood, a raw pixel is at the center of the neighborhood. In an "inception" neighborhood, a raw pixel is in the inceptive corner of the neighborhood.

Our experiments use the inception neighborhood strategy (with the original pixel in the top-left corner) to increase a usual image to a deeper-order version for the following reasons.

If one ues the multi-way Fourier transform defined in equation (10), it is not difficult to prove that the identity t-scalar E_T is a multi-way array whose only non-zero entry is equal to 1 and located at the inceptive position (i.e., $i_1 = \cdots = i_N = 1$). Any t-scalar in the form of $\alpha \cdot E_T$ (where $\alpha \in \mathbb{C}$) is called an inceptive t-scalar in this paper. Then, any inception neighborhood, namely its non-inceptive positions (i.e., $\exists n \in [N], i_n \neq 1$) valued as 0, can be represented as an inception t-scalar in the form $\alpha \cdot E_T$ where $\alpha \in \mathbb{C}$.

An inception t-scalar $\alpha \cdot E_T$ has properties similar to these od the complex number α, in that the following conditions hold for all $(\alpha \cdot E_T)$ and $(\beta \cdot E_T)$, where $\alpha, \beta \in \mathbb{C}$,

$$\begin{aligned} (\alpha \cdot E_T)^* &= \bar{\alpha} \cdot E_T \sim \bar{\alpha} \ , \\ (\alpha \cdot E_T) \circ (\beta \cdot E_T) &= (\alpha \cdot \beta) \cdot E_T \sim (\alpha \cdot \beta) \ , \\ (\alpha \cdot E_T) + (\beta \cdot E_T) &= (\alpha + \beta) \cdot E_T \sim (\alpha + \beta) \ . \end{aligned} \tag{33}$$

Equation (33) shows all inception t-scalars form a sub-algebra of C, isomorphic to the field of complex numbers. If all t-scalars are constrained to be inception t-scalars, an algorithm using inception t-scalars is equivalent to its (canonical) counterpart using the corresponding complex numbers.

Fortunately, the t-scalars obtained by the strategy of neighborhoods (central neighborhoods or inception neighborhoods) are usually not constrained to inception t-scalars. The experiments in this paper show that generalized algorithms using non-constained t-scalars outperform their counterparts using complex numbers (or equivalently using merely inception t-scalars). These experiments are discussed in Section VI.

V. Generalized tensors over t-algebra

A. Generalized tensors

Besides t-matrices, one can establish generalized tensors over t-scalars. We call them g-tensors (generalized tensors) and denote a g-tensor with M-modes by $X_{GT} \in C^{D_1\times\cdots\times D_M}$, where the subscript "$GT$" implies a generalized tensor.

G-tensors are the higher-order generalization of t-matrices and the deeper-order generalization of canonical tensors. Analogous to a canonical tensor, a g-tensor can be unfolded along its k-mode. Specifically, the mode-k unfolding of a g-tensor $X_{GT} \in C^{D_1\times\cdots\times D_M}$ is a t-matrix $X_{TM(k)}$ in $C^{D_k\times D_k^{-1}\prod_{m=1}^M D_m}$ such that the following condition holds for all $j \in \left[D_k^{-1}\prod_{m=1}^M D_m\right]$,

$$[X_{GT}]_{d_1,\cdots,d_M} = [X_{TM(k)}]_{d_k,j} \in C \tag{34}$$

where $[X_{GT}]_{d_1,\cdots,d_M}$ denotes the the $(d_1,\cdots,d_M)$-th t-scalar entry of X_{GT}, $[X_{TM(k)}]_{d_k,j}$ denotes the (d_k,j)-th t-scalar t-scalar entry of $X_{TM(k)}$ and the index j is determined by equation (2).

One can also define the generalized mode-k multiplication of a g-tensor $X_{GT} \in C^{D_1\times\cdots\times D_M}$ with a t-matrix $Y_{TM} \in C^{J\times D_k}$. Their mode-$k$ multiplication $A_{GT} \doteq X_{GT} \circ_k Y_{TM}$, analogous to the mode-$k$ multiplication of a (canonical) tensor with a (canonical) matrix, is a g-tensor in $C^{D_1\times\cdots D_{k-1}\times J\times D_{k+1}\times D_M}$ such that the following condition holds

$$A_{GT} = X_{GT} \circ_k Y_{TM} \ \Leftrightarrow \ A_{TM(k)} = Y_{TM} \circ X_{TM(k)} \ . \tag{35}$$

B. Serialization problem: little-endianess versus big-endianess

A convenient approach is to organize a t-matrix as a multi-mode array of complex numbers, some modes for the t-scalar entries and the others for the t-matrix. The principle of this approach also applies to a g-tensor.

For example, Kilmer et al. [32], [33] organize a t-matrix as a three-mode array, the first two modes for its rows and columns, the last for t-scalar entries. The modes chosen for t-scalars depend on the underlying structure of the data. Our implementation chooses the leading modes for t-scalars and differs from Kilmer et al.'s implementation. Our implementation

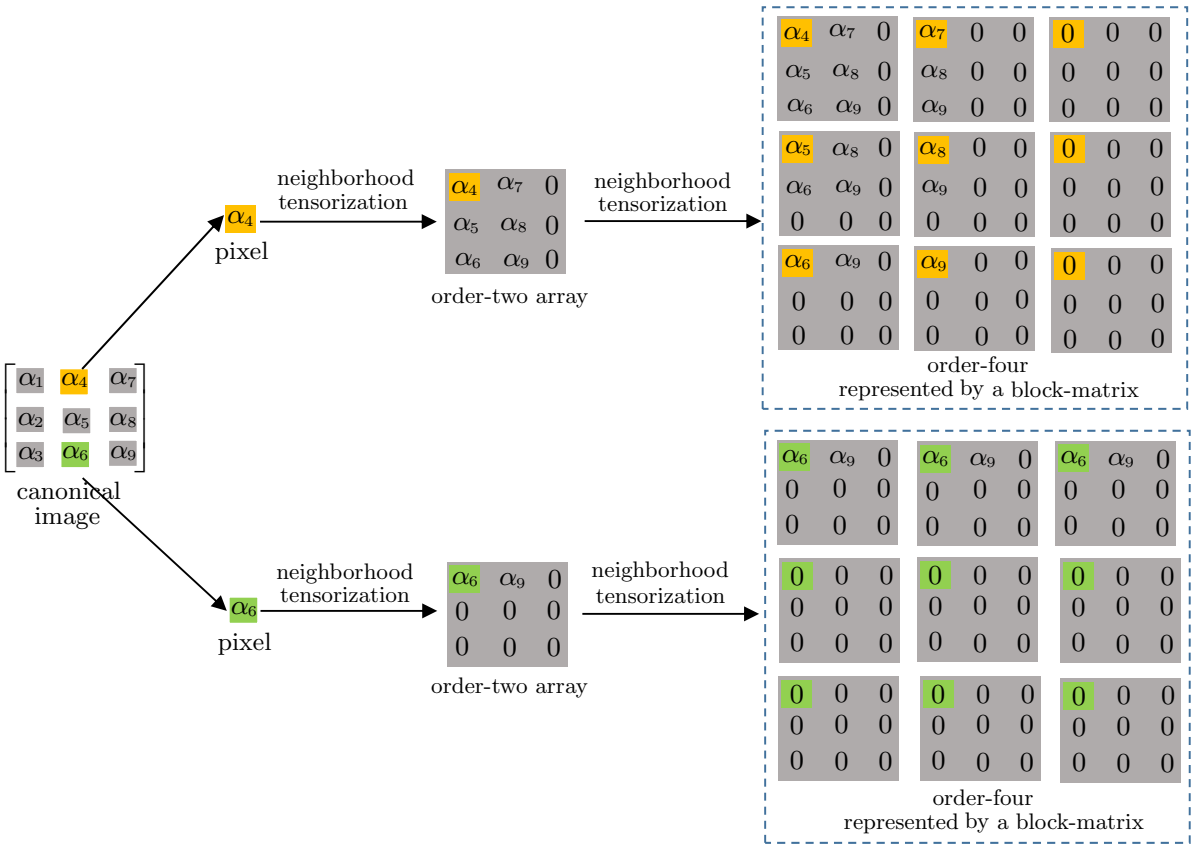


Fig. 3. The 3×3 neighborhood of each pixel is used to increase the orders of a grayscale image. Two example pixels α_4 and α_6 are released by order-two arrays and then by order-four arrays.

uses the little-endian protocol (reverse lexicographic ordering), and Kilmer et al. use the big-endian protocol (co-lexicographic ordering) [49], [50].

Although the Endian wars have never been quenched [51], the little-endian protocol offers the consistency of the transform of a t-matrix (or g-tensor) given by equation (22) and the transformation of a t-scalar given by equation (8). Using the little-endian protocol, we call the leading modes for t-scalars the "deeper orders" and the rest the "higher orders."

To this end, we denote

$$\mathbb{C}^{(I_1 \times \cdots I_N) \times [D_1 \times \cdots \times D_M]} \equiv C^{D_1 \times \cdots \times D_M} \tag{36}$$

to emphasize the first N modes of the underlying array are for t-scalars.

Then, given a g-tensor $X_{GT} \in C^{D_1 \times \cdots \times D_M}$, the underlying array of the generalized mode-k unfolding $X_{TM(k)}$ is an element in $\mathbb{C}^{(I_1 \times \cdots \times I_N) \times [D_k \times D_k^{-1} \prod_{m=1}^{M} D_m]}$.

C. THOSVD: t-algebra based HOSVD

THOSVD (T-algebra based HOSVD) is a generalization of HOSVD (Higher Order Singular Value Decomposition) over t-scalars. Specifically, given a g-tensor $X_{GT} \in C^{D_1 \times \cdots \times D_M}$, the g-tensor can be written by the following generalized multi-mode multiplication

$$X_{GT} = S_{GT} \circ_1 U_{TM,1} \circ_2 U_{TM,2} \cdots \circ_M U_{TM,M} \tag{37}$$

where $U_{TM,m} \in C^{D_M \times D_M}$ denotes the orthogonal t-matrix whose columns are the left singular t-vectors of the generalized mode-k unfolding of X_{GT} for all $m \in [M]$ and $S_{GT} \in C^{D_1 \times \cdots \times D_M}$ is the core g-tensor given by

$$S_{GT} = X_{GT} \circ_1 U^*_{TM,1} \circ_2 U^*_{TM,2} \cdots \circ_M U^*_{TM,M} \tag{38}$$

where $U^*_{TM,m}$ denotes the conjugate transpose of the t-matrix $U_{TM,m}$ such that the matrix representation $M(U^*_{TM,m})$ is the Hermitian transpose of $M(U_{TM,m})$ for all $m \in [M]$.

Let $\hat{U}_{TM,k} \in C^{D_k \times r_k}$ be the t-matrix containing the first r_k columns of the t-matrix $U_{TM,k} \in C^{D_k \times D_k}$. Then, one can define the following idempotent t-matrix

$$P_{TM,k} = \hat{U}_{TM,k} \circ \hat{U}^*_{TM,k} \in C^{D_k \times D_k} \ . \tag{39}$$

such that $M(P_{TM,k})$ is an idempotent matrix in $\mathbb{C}^{KD_k \times KD_k}$ for all $k \in [M]$.

The t-scalar-valued rank of the t-matrix $P_{TM,k}$ is given by

$$\operatorname{rank}_t P_{TM,k} = r_k \cdot E_T \geq Z_T \ . \tag{40}$$

The t-scalar-valued rank of a t-matrix is defined in the appendix of [34]. Interested readers are referred to this definition for more details.[1]

One can also define the (canonical) rank of a t-matrix $X_{TM} \in C^{D_1 \times D_2}$ as follows.

$$\operatorname{rank} X_{TM} \doteq \operatorname{rank} M(X_{TM}) \in [0, K \cdot \min(D_1, D_2)] \ . \tag{41}$$

and the following condition holds for all t-matrices X_{TM},

$$\operatorname{rank} X_{TM} = \operatorname{trace}(\operatorname{rank}_t X_{TM}) \ . \tag{42}$$

[1]The definition in [34] uses the notation $\operatorname{rank}(\cdot)$ rather than $\operatorname{rank}_t(\cdot)$ to denote the t-scalar-valued rank.

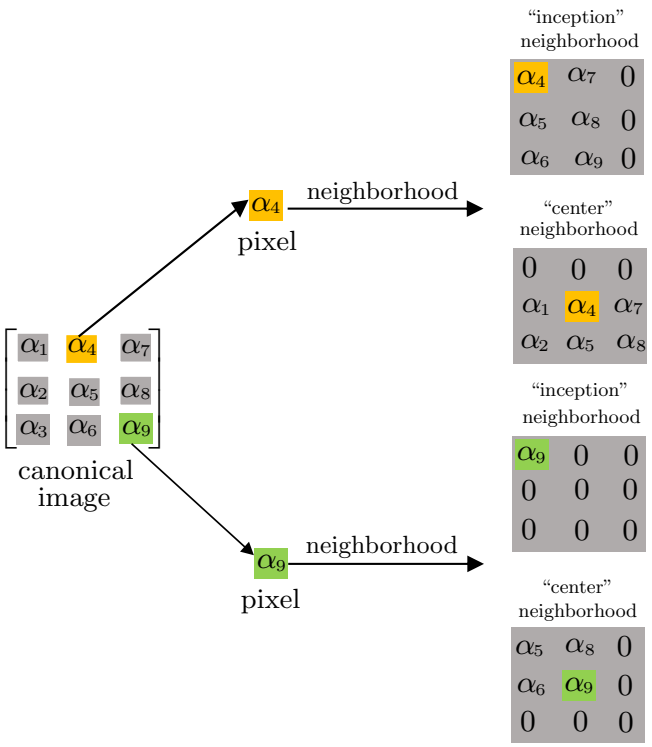


Fig. 4. Two types of neighborhoods of a raw pixel, one called "center" neighborhood with the pixel at the center, the other called "inceptive" neighborhood with the pixel at the inceptive corner

Thus, the canonical rank of the idempotent matrix $P_{TM,k}$ is given as follows

$$\operatorname{rank} P_{TM,k} \doteq \operatorname{rank} M(P_{TM,k}) = \operatorname{trace}(r_k \cdot E_T) = K \cdot r_k \ . \quad (43)$$

When the condition $r_k = D_k$ holds, the t-matrix $P_{TM,k}$ is the identity t-matrix in $C^{D_k \times D_k}$ and $M(P_{TM,k})$ is the identity matrix in $\mathbb{C}^{KD_k \times KD_k}$.

Then, given the positive integers $r_1, \cdots, r_M$, one has the idempotent t-matrices $P_{TM,1}, \cdots, P_{TM,M}$ and can project the g-tensor X_{GT} on a low-dimensional sub-module of $C^{D_1 \times \cdots \times D_M}$ as follows

$$\hat{X}_{GT} = X_{GT} \circ_1 P_{TM,1} \cdots \circ_M P_{TM,M} \ . \quad (44)$$

Let the low-dimensional sub-module be $\mathcal{M}$. The t-scalar-valued dimension of $\mathcal{M}$ is given by

$$\begin{aligned} \dim_t \mathcal{M} &= (r_1 \cdot r_2 \cdots r_M) \cdot E_T \\ &\leq (D_1 \cdot D_2 \cdots D_M) \cdot E_T \ . \end{aligned} \quad (45)$$

or equivalently, the (canonical) dimension of $\mathcal{M}$ is given by

$$\begin{aligned} \dim \mathcal{M} &= \operatorname{trace}\left(\dim_t \mathcal{M}\right) = K \cdot (r_1 \cdot r_2 \cdots r_M) \\ &\leqslant K \cdot (D_1 \cdot D_2 \cdots D_M) \ . \end{aligned} \quad (46)$$

The t-scalar-valued rank given by equation (45) is an inception t-scalar. The t-scalar-valued rank of a t-matrix had not to be just an inception t-scalar. However, we only discuss inception-tscalar-valued rank in this paper.

D. Local alternating optimization

The optimal $(r_1, \cdots, r_M)$-approximation of a g-tensor X_{GT} is a solution $\hat{X}_{GT}$ minimizing the following function under the partial order "$\leq$" on nonnegative t-scalars.

$$\{P^*_{TM,k}\}_{k=1}^M = \operatorname{argmin}_{\{P_{TM,k}\}_{k=1}^M} \|X_{GT} - \hat{X}_{GT}\|_{t,F} \quad (47)$$

where $\hat{Y}_{GT}$, the $(r_1, \cdots, r_M)$-approximation, is given by equation (44) and $\|\cdot\|_{t,F} : Y_{GT} \mapsto \|Y_{GT}\|_{t,F}$ denotes the t-scalar-valued Frobenius norm of a g-tensor Y_{GT} such that the following condition holds for all g-tensors Y_{GT}

$$\|Y_{GT}\|_{t,F} = \|Y_{TM(1)}\|_{t,F} \in S^{nonneg} \ . \quad (48)$$

To characterize the linearity rather than C-linearity, one can define the following (canonical) norm

$$\|Y_{GT}\|_F \doteq \|M(Y_{TM(1)})\|_F = \sqrt{\operatorname{trace} \|Y_{GT}\|^2_{t,F}} \geqslant 0 \ . \quad (49)$$

Then, equation (47) is equivalent to the following (canonical) minimization problem under the total order "$\leqslant$" on nonnegative real numbers

$$\{P^*_{TM,k}\}_{k=1}^M = \operatorname{argmin}_{\{P_{TM,k}\}_{k=1}^M} \|X_{GT} - \hat{X}_{GT}\|_F \ . \quad (50)$$

Analogous to the well-received conclusions [52], [53], there is no known deterministic global optimizer of equation (1). However, the ALS (Alternating Least Squares) algorithm called HOOI (Higher Order Orthogonal Iteration) can be generalized over t-scalars for locally minimizing equation (50) or equation (47). The generalized algorithm, called THOOI (t-algebra based HOOI), is a straightforward generalization of HOOI as shown in Algorithm 1. When $I_1 = \cdots = I_N = 1$, THOOI reduces to HOOI.

Algorithm 1 THOOI: t-algebra based HOOI

Input: a g-tensor $X_{GT} \in C^{D_1 \times \cdots \times D_M}$ and M initial idempotent t-matrices computed as in equation (39) by THOSVD such that $\operatorname{rank}_t P_{TM,k} = r_k \cdot E_T$ holds for all $k \in [M]$.

Output: the locally optimized g-tensor approximation $\hat{X}_{GT}$.

1: **repeat**
2: **for all** $k \in [M]$ **do**
3: $P_{TM,k} \leftarrow I_{TM} \in C^{D_m \times D_m}$ ◁ *set the k-th idempotent t-matrix to the identity t-matrix*
4: $\hat{X}_{GT} \leftarrow X_{GT} \circ_1 P_{TM,1} \cdots \circ_M P_{TM,M}$
5: $P_{TM,k} \leftarrow \hat{U}_{TM,k} \circ \hat{U}^*_{TM,k}$ where $\hat{U}_{TM,k}$ contains the leading r_k left singular t-vectors of the generalized mode-k unfolding of $\hat{X}_{GT}$.
6: **end for**
7: **until** convergence is obtained or the maximum number of iterations is reached
8: **return** the updated $P_{TM,1}, \cdots, P_{TM,M}$ and the optimized approximation $\hat{X}_{GT}$

E. Unification of principal component analysis algorithms

A wide range of PCA-based algorithms can be unified by HOSVD [54] or, more generally, THOSVD. Specifically, given Q multi-way arrays $\mathcal{X}_1, \cdots, \mathcal{X}_Q$ in $\mathbb{C}^{(I_1 \times \cdots \times I_N) \times [D_1 \times \cdots \times D_M]}$, one can concatenate them as a multi-way array

$$\mathcal{X} \overset{cat}{=} (\mathcal{X}_1, \cdots, \mathcal{X}_Q) \quad (51)$$

in $\mathbb{C}^{(I_1 \times \cdots \times I_N) \times [D_1 \times \cdots \times D_M \times Q]}$.

If $\mathcal{X}_1 + \cdots + \mathcal{X}_Q$ is not zero, one can use $\mathcal{X}_k \leftarrow (\mathcal{X}_k - \bar{\mathcal{X}})$ to update each sample array, where $\bar{\mathcal{X}} \doteq Q^{-1} \cdot (\mathcal{X}_1 + \cdots + \mathcal{X}_Q)$ denotes the mean of the Q arrays.

To have a consistent notation, we re-denote the multi-way array by $\mathcal{X} \in \mathbb{C}^{D_1 \times \cdots \times D_J}$ where $J = N + M + 1$. Then, a wide range of PCA-based algorithms on $\mathcal{X}$ can be unified by HOSVD in the following multi-linear multiplication,

$$\hat{\mathcal{X}} = \mathcal{X} \times_1 P_1 \cdots \times_J P_J \quad (52)$$

where $P_i \doteq \hat{U}_i \hat{U}_i^* \in \mathbb{C}^{D_i \times D_i}$ denotes the i-th idempotent matrix and $\hat{U}_i \in \mathbb{C}^{D_i \times r_i}$ denotes the matrix containing the r_i leading left singular vectors of the mode-i unfolding of $\mathcal{X}$ for all $i \in [J]$. If $r_i = D_i$, the idempotent matrix P_i is full rank, or equivalently, the identity matrix I_{D_i} in $\mathbb{C}^{D_i \times D_i}$.

If P_i is rank deficient for all $i < J$ and the only full rank idemponent matrix is P_J, then the approximation $\hat{\mathcal{X}} \stackrel{cat}{=} (\hat{\mathcal{X}}_1, \cdots, \hat{\mathcal{X}}_Q)$ is the PCA reconstruction of $\mathcal{X} \stackrel{cat}{=} (\mathcal{X}_1, \cdots, \mathcal{X}_Q)$.

When P_2 is rank deficient, and the other idempotent matrices are of full rank, the arrays $\hat{\mathcal{X}}_1, \cdots, \hat{\mathcal{X}}_Q$ are the so-called 2DPCA reconstructions of $\mathcal{X}_1, \cdots, \mathcal{X}_Q$. When only P_1 and P_2 are rank-deficient, $\hat{\mathcal{X}}_1, \cdots, \hat{\mathcal{X}}_Q$ are the reconstructions of the so-called $(2D)^2$PCA approximations proposed by Zhang and Zhou et al [55].

Further, when $P_1, \cdots, P_{J-1}$ are rank-deficient and P_J is of full rank, equation (52) gives the reconstruction of raw MPCA (multi-linear PCA), a multi-way generalization of PCA, 2DPCA and $(2D)^2$PCA proposed by Lu et al [56].

However, since PCA and 2DPCA are already generalized over t-scalars, see TPCA and T-2DPCA in [15], [34], [57], we are interested in unifying them via THOSVD and generalizing Lu's MPCA over t-scalars. This unification provides a panoramic view of seemingly unrelated algorithms and might help in the design of new feature extractors.

The unification is straightforward. Given Q multi-way arrays $\mathcal{X}_1, \cdots, \mathcal{X}_Q \in \mathbb{C}^{(I_1 \times \cdots \times I_N) \times [D_1 \times \cdots \times D_M]}$ such that their sum is zero, these arrays can be characterized by Q g-tensors in $C^{D_1 \times \cdots \times D_M}$ over t-scalars in $\mathbb{C}^{I_1 \times \cdots \times I_N}$.

The Q g-tensors can be concatenated and characterized by a higher-order g-tensor X_{GT} in $C^{D_1 \times \cdots \times D_M \times D_{M+1}}$ (where $D_{M+1} \doteq Q$). Then, similar to equation (44), the g-tensor X_{GT} is approximated by the following generalized multi-mode multiplication

$$\hat{X}_{GT} = X_{GT} \circ_1 P_{TM,1} \cdots \circ_{(M+1)} P_{TM,(M+1)} . \qquad (53)$$

If $P_{TM,(M+1)}$ is rank-deficient, and the other t-matrices are of full rank, the approximation $\hat{X}_{GT} \stackrel{cat}{=} (\hat{X}_1, \cdots, \hat{X}_Q)$ gives the TPCA approximations of the Q g-tensors $X_1, \cdots, X_Q$ in $C^{D_1 \times \cdots \times D_M}$.

When the only rank-deficient idempotent t-matrix is $P_{TM,2}$, equation (53) yields the approximations by T-2DPCA. T-2DPCA generalizes Yang's 2DPCA and can be found with more details in [34]. When the only rank-deficient idempotent t-matrices are $P_{TM,1}$ and $P_{TM,2}$, equation (53) generalizes the result given by $(2D)^2$PCA. Furthermore, when the only full-rank idempotent t-matrix is $P_{(M+1)}$, equation (53) generalizes the results by MPCA. We call the generalized algorithm TM-PCA (t-algebra based MPC).

If there are two or more rank-deficient idempotent t-matrices, equation (53) cannot solve the minimization described in equation (50). However, one can use Algorithm 1 to optimize the results given by equation (53).

VI. Experimental Verifications

In this section, two types of experiments are described. One type is the "vertical" experiments in which we use HOSVD to approximate multi-spectral images. We also use THOSVD with the same parameters to approximate the deeper-order versions of these multispectral images.

The other type is the "horizontal" experiments in which THOSVD approximates raw multispectral images, without the pixel neighborhood strategy, as HOSVD does on the same raw images.

All the deeper-order versions of raw images are generated using the "inception" neighborhood strategy. The transform used for defining the t-scalar multiplication is based on the Fourier transform. The results of experiments using the "center" neighbor strategy and the cosine-based transform are omitted since they are similar to those mentioned above.

A. "Vertical" experiments

Two public multispectral images are used in the "vertical" experiments of image approximation using HOSVD and THOSVD. One is the "food" image, the other the "superballs" image [2].

Each image has 31 channels and is a $512 \times 512 \times 31$ array of real numbers that defines an order-three canonical tensor. The tensor is approximated using HOSVD.

To exploit the potential of THOSVD in the "vertical" experiments, it is necessary to extend the raw images to their deeper-order versions. With the neighborhood strategy, each grayscale of a multispectral image is increased to a 3×3 array of grayscales. Thus, each $512 \times 512 \times 31$ multispectral image is increased to a $(3 \times 3) \times [512 \times 512 \times 31]$ array of real numbers, i.e., $I_1 \times I_2 = 3 \times 3$ and $D_1 \times D_2 \times D_3 = 512 \times 512 \times 31$.

Figure 5 shows the increase of a $D_1 \times D_2 \times D_3$ tensor (multispectral image) to its order-five version. Each mode-3 fiber (i.e., a vector) in $\mathbb{C}^{D_3}$, is increased to a generalized mode-3 fiber (i.e., a t-vector) in $C^{D_3} \equiv \mathbb{C}^{(I_1 \times I_2) \times [D_3]}$.

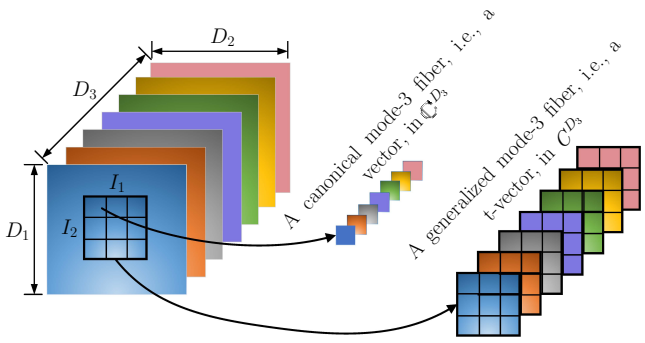


Fig. 5. A multispectral image (with $I_1 \times I_2$ inception pixel neighborhoods) to the order-five version, where each mode-3 fiber, i.e., a vector in $\mathbb{C}^{D_3}$, is increased to a t-vector in $C^{D_3} \equiv \mathbb{C}^{(I_1 \times I_2) \times [D_3]}$.

Using the neighborhood strategy for a second time, one can increase a multispectral image to the size $(3 \times 3 \times 3 \times 3) \times [512 \times 512 \times 31]$.

Different variants with different t-scalar settings are used in this "vertical" experiment. Their settings are listed in Table II. Among these methods, HOSVD is the special case of THOSVD with t-scalars reduced to order-zero (i.e., canonical scalars). THOSVD uses order-two t-scalars, and THOSVD-A uses order-four t-scalars. If THOOI is used, the optimized versions are denoted by by THOSVD-OP and THOSVD-A-OP.

The peak signal-to-noise ratio (PSNR) is used to measure the quality of the approximations. The larger the PSNR, the better the approximation. For a multi-way array $\mathcal{X}$ and its approximation $\hat{\mathcal{X}}$, the PSNR is given by

$$PSNR = 20 \log_{10} \frac{MAX \cdot \sqrt{N^{entry}}}{\|\mathcal{X} - \hat{\mathcal{X}}\|_F} \qquad (54)$$

[2] https://www1.cs.columbia.edu/CAVE/databases/multispectral/

TABLE II
HOSVD/THOSVD VARIANTS

method	t-scalar size ($I_1 \times \cdots \times I_N$)	t-scalar order	if locally optimized?
HOSVD	–	0	No
HOSVD-OP	–	0	Yes
THOSVD	3×3	2	No
THOSVD-A	$3 \times 3 \times 3 \times 3$	4	No
THOSVD-OP	3×3	2	Yes
THOSVD-A-OP	$3 \times 3 \times 3 \times 3$	4	Yes

where N^{entry} denotes the number of scalar entries of the given array, $\|\cdot\|_F : \mathcal{X} \mapsto \sqrt{\sum_{d_1,\cdots,d_M} |(\mathcal{X})_{d_1,\cdots,d_M}|^2}$ denotes the Frobenius norm of a canonical tensor in $\mathbb{R}^{D_1 \times \cdots \times D_M}$, and MAX denotes the maximum possible value of the scalar entries.

Figure 6 shows the heatmaps of PSNRs using HOSVD, THOSVD, THOSVD-A with different values of rank tuple (r_1, r_2, r_3) on approximating the "food" multispectral image and the "superballs" multispectral image.

When r_1, r_2, and r_3 are increased, the PSNR of the approximation also increases. The generalized methods THOSVD and THOSVD-A consistently outperform the canonical counterpart HOSVD with the same parameters. The tabulated results in Figure 6 provide a quantitative comparison. For example, when $(r_1, r_2, r_3) = (500, 500, 31)$, THOSVD outperforms HOSVD by 22.9 dB (i.e., 98.55 dB - 75.65 dB) and THOSVD-A outperforms HOSVD by 38.06 dB (i.e., 113.71 dB - 75.65 dB) on the "food" image. The approximation results on the "superballs" image are consistent with the conclusion drawn from the results on the "superballs" image.

Figure 6 also shows that the generalized algorithm using deeper-order t-scalars outperforms the version of the algorithm using shallower-order t-scalars. In both the "food" image and the "superballs" image, THOSVD-A (using order-four t-scalars) outperforms THOSVD (using order-two t-scalars). Also note, HOSVD is a special case of THOSVD with the "shallowest order" t-scalars (i.e., order-zero t-scalars). HOSVD yields the lowest PSNRs in the "vertical" experiments.

B. "Vertical" experiments using local optimization

A comparison of the approximations by optimized algorithms HOSVD-OP, THOSVD-OP, and THOSVD-A-OP is given in Figure 7. Figure 7(a) shows the results on the "food" image and 7(b) shows the results on the "superballs" image.

The results are consistent with the observation that generalized algorithms outperform their canonical counterpart, and generalized algorithms using high-order t-scalars outperform those using low-order t-scalars. Some representative quantitative results are also tabulated in Figure 7. For example, on the "food" image, when $(r_1, r_2, r_3) = (180, 180, 25)$, THOSVD-OP outperforms HOSVD-OP by 17.20 dB (i.e., 67.02 dB – 49.82 dB), and THOSVD-A-OP outperforms THOSVD-OP by 0.59 dB (i.e., 67.61 dB – 67.02 dB).

The tabulated results in Figure 7 also include the PSNRs yielded by HOSVD, THOSVD, and THOSVD-A (i.e., the algorithms without the alternating optimization). It shows that the optimized algorithms HOSVD-OP, THOSVD-OP, and THOSVD-A-OP respectively outperform non-optimized algorithms HOSVD, THOSVD, and THOSVD-A.

C. "Horizontal" experiments on USC-SIPI images

It is of interest to compare HOSVD and THOSVD on approximating the same raw images. For example, a color image is a three-way array of real numbers. It can be regarded as either an order-three tensor of real numbers or a t-matrix of order-one t-scalars.

When regarded as an order-three tensor, a color image can be approximated by HOSVD along each of the three modes. When regarded as a t-matrix, a color image can be approximated by THOSVD along each of the two generalized modes. For example, given a $256 \times 256 \times 3$ color image, one can model it as a 256×256 t-matrix (using the big-endian protocal), with each t-scalar entry being an order-one array containing three real numbers. To approximate such a t-matrix, THOSVD requires two parameters $r_1 \in [512]$ and $r_2 \in [512]$.

Although the quality of a t-matrix approximation is determined by $\min(r_1, r_2)$, we give the results of THOSVD using the two parameters r_1 and r_2 to compare with the results of HOSVD using the same parameters.

Also worthy of notice is, with two parameters r_1 and r_2, THOSVD is equivalent to TSVD introduced in [34] or Kilmer et al.'s t-SVD introduced in [32]. In the following, we use TSVD to interpret the approximating process of THOSVD. More specifically, let X_{TM} be a t-matrix in $\in C^{D_1 \times D_2}$ (where $D_1 = D_2 = 512$). The compact TSVD of X_{TM} is given by

$$X_{TM} = U_{TM} \circ S_{TM} \circ V_{TM}^* \tag{55}$$

where the t-matrix $U_{TM} \in C^{D_1 \times \min(D_1, D_2)}$ contains the left singular t-vectors of the t-matrix X_{TM}, and the t-matrix $V \in C^{D_2 \times \min(D_1, D_2)}$ contains the right singular t-vectors of X_{TM} such that the following condition holds

$$X_{TM}^* \circ X_{TM} = V_{TM}^* \circ V_{TM} = I_{TM}$$

where I_{TM} denotes the $\min(D_1, D_2) \times \min(D_1, D_2)$ identity t-matrix, or equivalently, the matrix $M(I_{TM})$ is the (canonical) identity matrix in $\mathbb{C}^{K \min(D_1, D_2) \times K \min(D_1, D_2)}$.

The t-matrix S_{TM} denotes the $\min(D_1, D_2) \times \min(D_1, D_2)$ diagonal t-matrix whose (i, j)-th t-scalar entry is given by

$$[S_{TM}]_{i,j} = \delta_{i,j} \cdot \beta_{T,i}$$

where $\delta_{i,j}$ denotes the Kroneck delta and $\beta_{T,i}$ is the i-th leading singular t-scalar value of X_{TM} such that the following partial order of nonnegative t-scalars holds for all $i \in [\min(D_1, D_2) - 1]$,

$$\beta_{T,i} \geq \beta_{T,(i+1)} \geq Z_T \ .$$

Then, given two parameters r_1 and r_2, one can approximate the t-matrix X_{TM} as follows

$$\hat{X}_{TM} = U_{TM} \circ \hat{S}_{TM} \circ V_{TM}^* \ . \tag{56}$$

where $\hat{S}_{TM}$ denotes the low-rank approximation of S_{TM} with the parameters $r_1, r_2 \in [\min(D_1, D_2)]$ such that the following condition holds

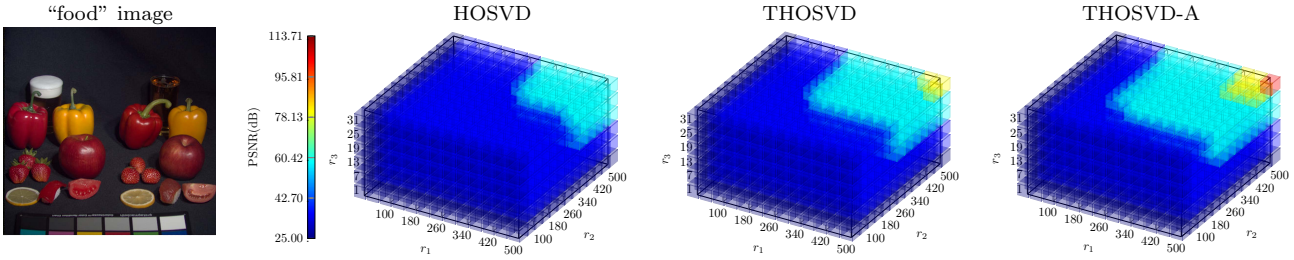
$$[\hat{S}_{TM}]_{i,j} = \begin{cases} [S_{TM}]_{i,j} & \text{if } i \in [r_1] \text{ and } j \in [r_2] \\ Z_T & \text{otherwise} \end{cases} \ . \tag{57}$$

One can find more details of TSVD in [34] and [32].

To have a fair comparison with the same parameters, r_1 and r_2, HOSVD approximates a given multispectral image $\mathcal{X} \in \mathbb{C}^{D_1 \times D_2 \times D_3}$ as follows

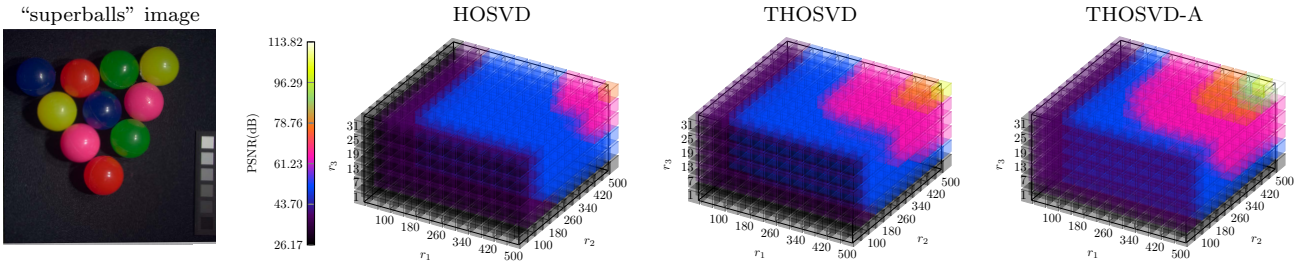
$$\hat{\mathcal{X}} = \mathcal{X} \times_1 (\hat{U}_1 \hat{U}_1^*) \times_2 (\hat{U}_2 \hat{U}_2^*) \times_3 I_{D_3}$$

where I_{D_3} denotes the $D_3 \times D_3$ identity matrix, $\hat{U}_1 \in \mathbb{C}^{D_1 \times r_1}$ and $\hat{U}_2 \in \mathbb{C}^{D_2 \times r_2}$ are semi-orthogonal, the columns of $\hat{U}_1$ are



Quantitative comparison of PSNRs by HOSVD, THOSVD and THOSVD-A on the "food" multispectral image with some representative parameter tuples (r_1, r_2, r_3)

method	rank tuple (r_1, r_2, r_3)											
	$r_3 = 7$				$r_3 = 19$				$r_3 = 31$			
	$r_1 = 20$	180	340	500	20	180	340	500	20	180	340	500
	$r_2 = 20$	180	340	500	20	180	340	500	20	180	340	500
HOSVD	35.410	47.250	49.220	49.910	35.470	49.910	55.640	62.250	35.470	50.000	56.220	70.770
THOSVD	36.040	49.250	50.610	50.790	36.110	53.470	60.490	63.380	36.110	53.720	63.250	95.260
THOSVD-A	36.440	50.250	50.990	51.040	36.530	56.510	62.830	63.890	36.530	57.130	69.040	113.820



Quantitative comparison of PSNRs by HOSVD, THOSVD and THOSVD-A on the "superballs" multispectral image with some representative parameter tuples (r_1, r_2, r_3)

method	rank tuple $(r_1, r_2.r_3)$											
	$r_3 = 7$				$r_3 = 19$				$r_3 = 31$			
	$r_1 = 20$	180	340	500	20	180	340	500	20	180	340	500
	$r_2 = 20$	180	340	500	20	180	340	500	20	180	340	500
HOSVD	35.410	47.250	49.220	49.910	35.470	49.910	55.640	62.250	35.470	50.000	56.220	70.770
THOSVD	36.040	49.250	50.610	50.790	36.110	53.470	60.490	63.380	36.110	53.720	63.250	95.260
THOSVD-A	36.440	50.250	50.990	51.040	36.530	56.510	62.830	63.890	36.530	57.130	69.040	113.820

Fig. 6. A "vertical" comparison of the PSNRs by HOSVD, THOSVD and THOSVD-A with different parameter tuples (r_1, r_2, r_3) on the "food" and "superballs" multispectral image (a) RGB version of the "food" iamge (b) Results of different algorithms on the "food" image (c) RGB version of the "superballs" image (d) Results of different algorithms on the "superballs" image

the r_1 leading left singular vectors of the mode-1 unfolding of $\mathcal{X}$, and the columns of $\hat{U}_2$ are the r_2 leading left singular vectors of the mode-2 unfolding of $\mathcal{X}$.

Two public USC-SIPI RGB images are used in the "horizontal" experiment. One is the "house" image, the other is the "airplane" image [3]. The "house" image size is $256 \times 256 \times 3$, and the "airplane" image size is $512 \times 512 \times 3$.

Figure 8 shows the PSNRs by HOSVD and THOSVD for approximating the "house" image and the "airplane" image. The PSNRs with some representative rank tuples (r_1, r_2) are also tabulated in the figure. It shows that, on the same raw data, THOSVD consistently outperforms HOSVD. For example, when $(r_1, r_2) = (210, 210)$, on the "house" image, THOSVD outperforms HOSVD by 13.29 dB (62.30 dB − 49.01 dB).

[3]https://sipi.usc.edu/database/

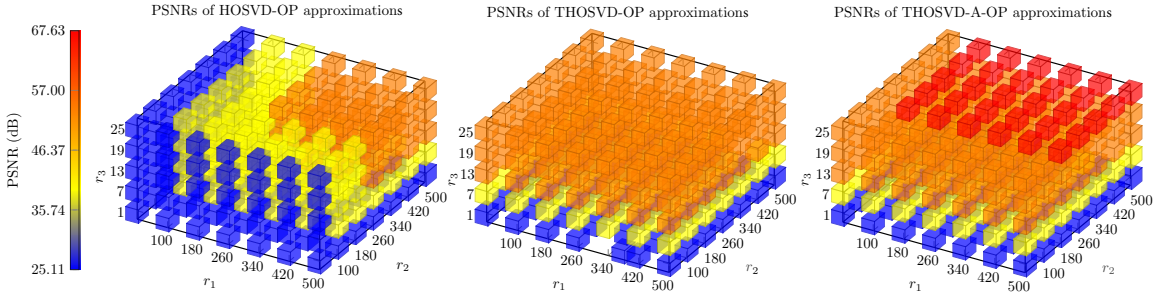
D. Unification of principal component analysis algorithms

This section unifies a wide range of principal component analysis algorithms and compares their performances for approximating images in vertical and horizontal experiments.

The CIFA-10 image set is used in the experiments. The image set contains 60000 RGB images. Each image is a $32 \times 32 \times 3$ array [4]. We use the first 1000 images of this image set for the experiments. The 1000 CIFA images can be organized as a $32 \times 32 \times 3 \times 1000$ array of real numbers, namely 3000 grayscale images of size 32×32.

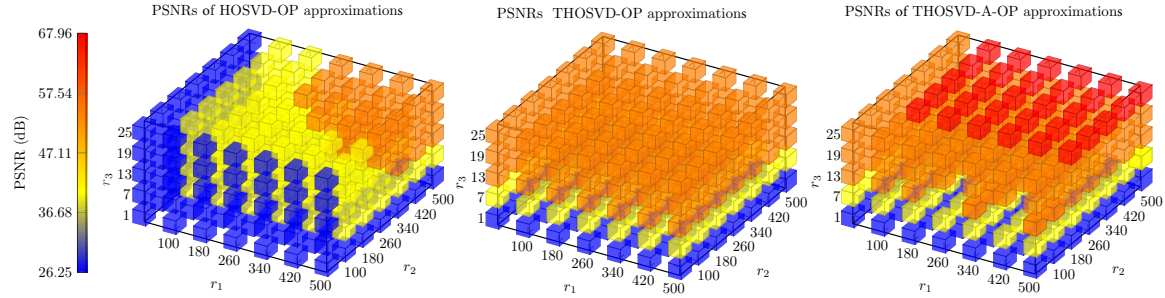
Using the 3×3 neighborhood strategy as shown in Figure 3, one can increase each grayscale image from order-two to order-four, yielding a g-tensor of four modes in $C^{32\times32\times3\times1000}$, i.e., a $(3 \times 3) \times [32 \times 32 \times 3 \times 1000]$ underlying array of real numbers for the selected CIFA-10 images.

[4]https://www.cs.toronto.edu/~kriz/cifar.html



method	rank tuple (r_1, r_2, r_3)											
	$r_3 = 1$				$r_3 = 13$				$r_3 = 25$			
	$r_1 = 20$	180	340	500	20	180	340	500	20	180	340	500
	$r_2 = 20$	180	340	500	20	180	340	500	20	180	340	500
HOSVD	25.00	25.48	25.48	25.48	32.23	49.09	55.04	56.63	32.24	49.61	58.46	66.07
HOSVD-OP	25.11	25.48	25.48	25.48	32.42	49.20	55.08	56.63	32.43	49.82	58.48	66.07
THOSVD	25.06	25.51	25.51	25.51	32.63	52.23	56.55	57.12	32.63	53.39	62.82	67.10
THOSVD-OP	25.51	25.51	25.51	25.51	56.91	57.09	57.12	57.12	66.38	67.02	67.10	67.10
THOSVD-A	25.11	25.53	25.53	25.53	32.95	54.45	57.20	57.36	32.96	56.77	65.89	67.63
THOSVD-A-OP	25.53	25.53	25.53	25.53	57.16	57.35	57.36	57.36	67.02	67.61	67.63	67.63

(a) results on the "food" image



method	rank tuple (r_1, r_2, r_3)											
	$r_3 = 7$				$r_3 = 13$				$r_3 = 19$			
	$r_1 = 20$	180	340	500	20	180	340	500	20	180	340	500
	$r_2 = 20$	180	340	500	20	180	340	500	20	180	340	500
HOSVD	35.41	47.25	49.22	49.91	35.47	49.66	54.60	58.41	35.47	49.91	55.64	62.25
HOSVD-OP	35.55	47.36	49.25	49.91	35.61	49.83	54.69	58.42	35.61	50.09	55.74	62.26
THOSVD	36.04	49.25	50.61	50.79	36.11	53.00	58.16	59.50	36.11	53.47	60.49	63.38
THOSVD-OP	50.64	50.79	50.79	50.79	59.07	59.49	59.50	59.50	62.91	63.34	63.38	63.38
THOSVD-A	36.44	50.25	50.99	51.04	36.52	55.58	59.48	59.90	36.53	56.51	62.83	63.89
THOSVD-A-OP	50.90	51.04	51.04	51.04	59.46	59.90	59.90	59.90	63.36	63.87	63.89	63.89

(b) results on the "superballs" image

Fig. 7. A comparison of the results by optimized algorithms HOSVD-OP, THOSVD-OP, and THOSVD-A-OP on the "food" image and the "superballs" image (a) results on the "food" image (b) results on the "superballs" image

By reusing the neighborhoods as shown in Figure 5, one can further have a g-tensor in $\mathcal{C}^{32\times32\times3\times1000}$ of t-scalars in $\mathbb{C}^{3\times3\times3\times3}$, i.e., an underlying order-eight array of real numbers.

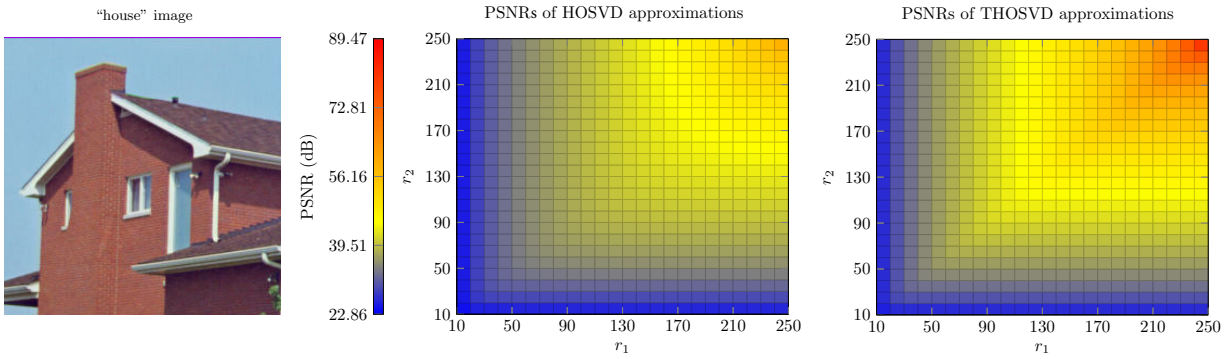
No matter what size the underlying array is, the g-tensor has four modes, two modes for the rows and columns of the underlying images, one mode for the RGB changes, the fourth mode for image samples. Let each image sample be subtracted from their mean. Then, by using THOSVD on the row and column modes, one generalizes $(2D)^2$PCA [55]. We call the generalized algorithm T-$(2D)^2$PCA.

Note that MPCA is a higher-order generalization of $(2D)^2$PCA, while T-$(2D)^2$PCA is a deeper-order generalization of $(2D)^2$PCA, totally different from MPCA. Further, if one of the two modes mentioned above is left unhandled, T-$(2D)^2$PCA reduces to T-2DPCA, generalizing Yang's 2DPCA [58]. Interested readers are referred to [34] for a performance comparison between 2DPCA and T-2DPCA.

Figure 9 shows the PSNRs by $(2D)^2$PCA and T-$(2D)^2$PCA using either order-two t-scalars or order-four t-scalars. The observation from this figure is consistent with those found in other experiments — the generalized algorithm outperforms its canonical counterpart, and an algorithm using "deeper-order" t-scalars outperforms its counterpart using "shallower-order" t-scalars.

One can further optimize the results of $(2D)^2$PCA and T-$(2D)^2$PCA with the alternating least-squares algorithms (HOOI and THOOI).

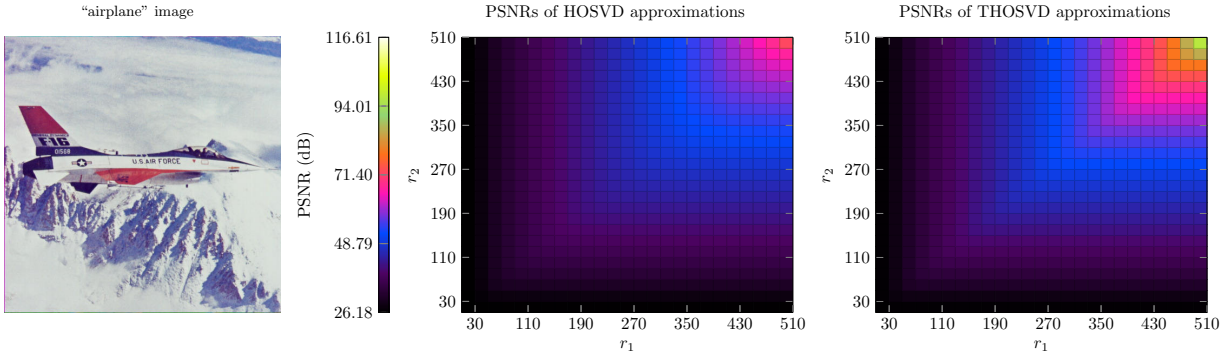
Table III lists the quantitative PSNRs of related algorithms, using or not using alternating optimization, with some representative parameters. It shows that alternating optimization



PSNRs with some representative rank tuples (r_1, r_2) on the "house" image

method	rank tuple (r_1, r_2)						
	(10,10)	(50,50)	(90,90)	(130,130)	(170,170)	(210,210)	(250,250)
HOSVD	22.86	32.90	37.08	40.53	44.17	49.01	61.44
THOSVD	24.53	35.73	41.56	46.86	53.03	62.30	89.47

(a) results on the "house" image



PSNRs with some representative rank tuples (r_1, r_2) on the "airplane" image

method	rank tuple (r_1, r_2)						
	(30,30)	(110,110)	(190,190)	(270,270)	(350,350)	(430,430)	(510,510)
HOSVD	26.18	35.14	40.90	45.89	51.44	58.60	79.46
THOSVD	27.11	37.05	44.09	50.68	58.84	71.08	116.61

(b) results on the "airplane" image

Fig. 8. A comparison of the PSNRs (dB) by HOSVD and THOSVD on the "house" image and the "airplane" image

can increase performance. Further, the tabulated results are consistent with the conclusion found in previous experiments — an algorithm using "deeper-order" t-scalars outperforms its counterpart using "shallower-order" t-scalars.

E. "Horizontal" comparison

The comparison of related algorithms on the same images is conducted in a "horizontal" experiment. The SVHN (Street View House Number) image set is chosen for the experiments. The SVHN image set contains over sixty hundred thousand $32 \times 32 \times 3$ RGB digit images [5]. The first 100 images from its training set are chosen for the experiments.

To unify the PCA-based approximations, one needs to organize the mean-subtracted images into a $32 \times 32 \times 3 \times 100$ array of real numbers. This underlying order-four array can be considered a canonical tensor of four modes (i.e., rows, columns, color channels, and samples) or a g-tensor of three modes (i.e., rows, columns, and samples) with the mode of color channels chosen to characterize t-scalars. Also worthy of notice is that the little-endian protocal mentioned in Section V-B requires permuting the grayscale indices of the underlying arrays, transforming their sizes from $32 \times 32 \times 3$ to $3 \times 32 \times 32$.

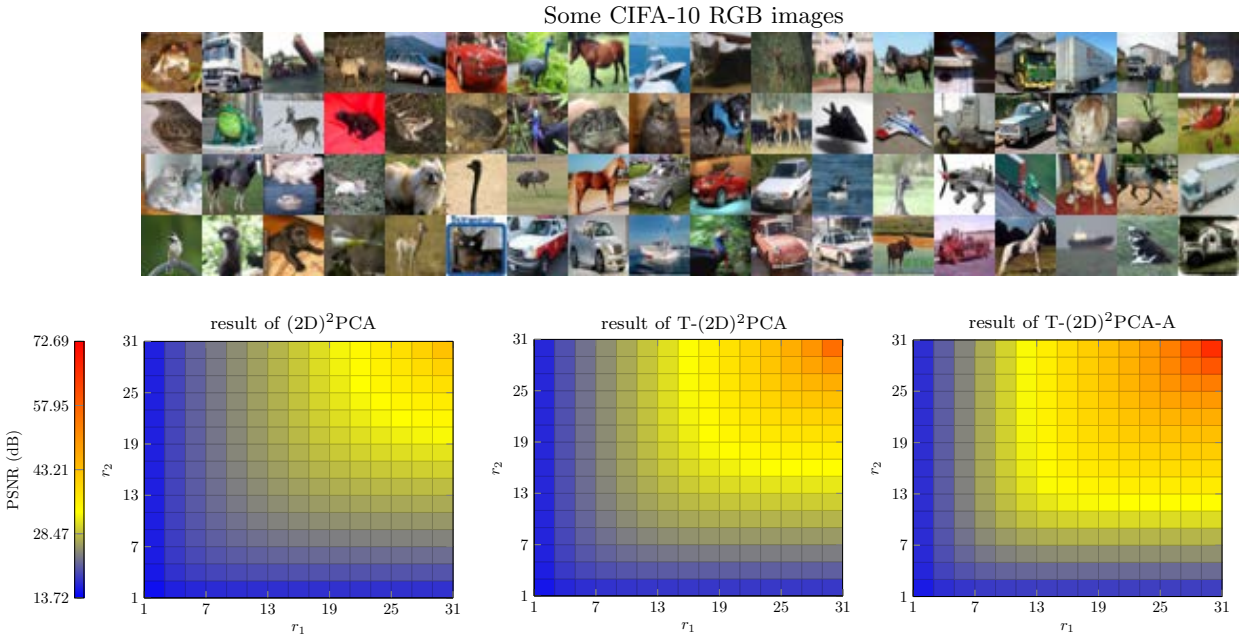
The top row of Figure 10 shows the chosen SVHN images. The first two sub-figures of the second row shows the PSNRs by $(2D)^2$PCA and the PSNR gain of T-$(2D)^2$PCA on the same SVHN images. The last two sub-figures of the second row show the PSNRs on the same images by the optimized algorithms $(2D)^2$PCA-OP and T-$(2D)^2$PCA-OP. Some quantitative PSNRs with representative parameters r_1, r_2 are also given in Figure 10.

The "horizontal" comparison shows that generalized algorithms T-$(2D)^2$PCA and T-$(2D)^2$PCA outperform their canonical counterparts $(2D)^2$PCA and $(2D)^2$PCA on the same SVHN images.

Another "horizontal" experiment compares the performances of PCA and TPCA. TPCA is a "deeper-order" generation of PCA [15], [34]. Therefore, one can employ HOSVD and THOSVD to realize PCA and TPCA, respectively.

Given a $32 \times 32 \times 3 \times 100$ array formed by the 100 mean-subtracted SVHN image samples, PCA on these samples is

[5] http://ufldl.stanford.edu/housenumbers/



PSNRs by $(2D)^2$PCA, T-$(2D)^2$PCA and T-$(2D)^2$PCA-A with some representative parameter tuple (r_1, r_2)

method	parameter tuple (r_1, r_2)							
	(3,3)	(7,7)	(11,11)	(15,15)	(19,19)	(23,23)	(27,27)	(31,31)
$(2D)^2$PCA	16.59	20.25	23.21	26.06	29.21	32.84	37.30	45.78
T-$(2D)^2$PCA	17.08	21.45	25.35	29.43	33.90	38.77	44.93	64.81
T-$(2D)^2$PCA-A	17.65	23.03	28.89	34.59	38.98	44.12	53.03	72.69

Fig. 9. A comparison of the PSNRs by $(2D)^2$PCA and T-$(2D)^2$PCA on 1000 CIFA images where T-$(2D)^2$PCA is with order-two t-scalars, and T-$(2D)^2$PCA-A is withh order-four t-scalars

TABLE III
A quantitative comparison of the PSNRs with some representative parameters using or not using alternating optimization

method	parameter tuple (r_1, r_2)								optimized?
	(3, 3)	(7, 7)	(11, 11)	(15, 15)	(19, 19)	(23, 23)	(27, 27)	(31, 31)	
$(2D)^2$PCA	16.59	20.25	23.21	26.06	29.21	32.84	37.30	45.78	no
$2D^2$PCA-OP	16.61	20.26	23.21	26.06	29.22	32.84	37.31	45.78	yes
T-$(2D)^2$PCA	17.08	21.45	25.35	29.43	33.9	38.77	44.93	64.81	no
T-$(2D)^2$PCA-OP	19.73	24.47	28.54	32.63	37.04	41.78	48.03	67.91	yes
T-$(2D)^2$PCA-A	17.65	23.03	28.89	34.59	38.98	44.12	53.03	72.69	no
T-$(2D)^2$PCA-A-OP	20.42	26.19	32.18	37.78	42.08	47.27	56.25	77.12	yes

equivalent to HOSVD conducted merely with the sample mode of the underlying array. The $32 \times 32 \times 3 \times 100$ array can also be interpreted as a g-tensor in $C^{32\times32\times100}$ (i.e., 32 rows, 32 columns, 100 samples, and each t-scalar entry containing three numbers). Analogous to PCA, TPCA is equivalent to THOSVD performed merely with the "sample" mode of the g-tensor.

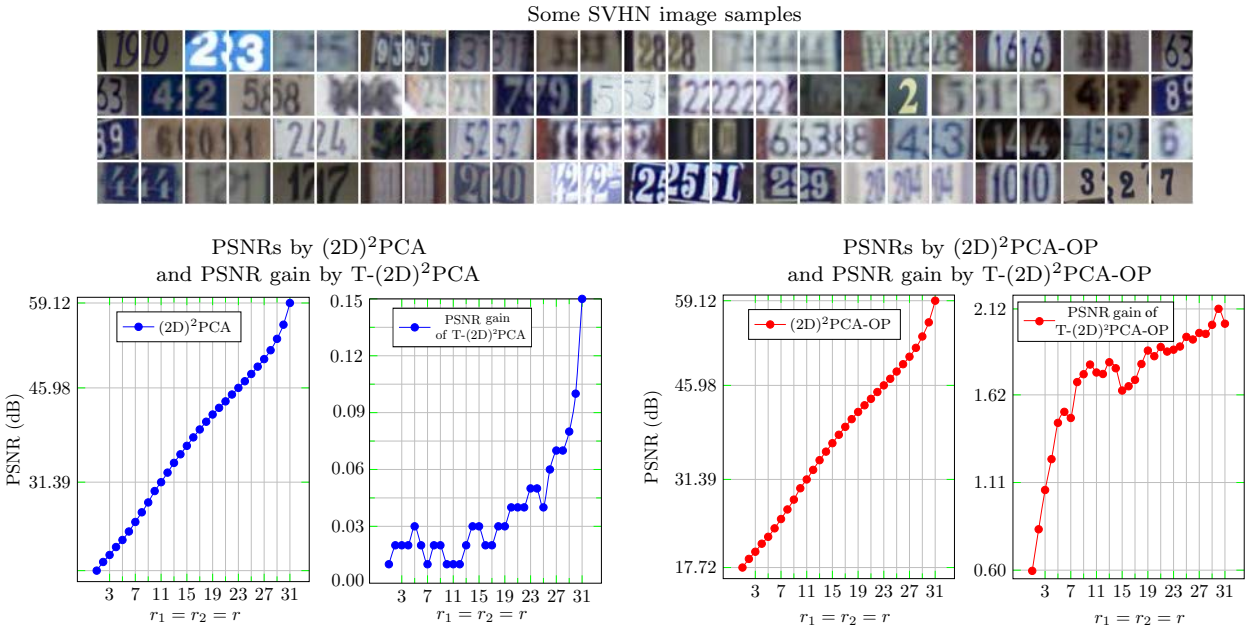
It is noted that to implement specific g-tensors using the little-endian protocol discussed in Section V-B, one must permute the underlying $32 \times 32 \times 3 \times 100$ array to an underlying $3 \times 32 \times 32 \times 100$ array.

The PSNR curves yielded by PCA and TPCA on the same image SVHN samples are shown in Figure 11. A quantitative comparison of PCA and TPCA with some representative parameter $r \in [99]$ is also tabulated in Figure 11. It shows that TPCA consistently outperforms PCA with different r on the same SVNH images. It reconfirms the conclusion that an algorithm established over t-scalars outperforms its counterpart over canonical scalars (i.e., complex numbers).

It is noted that the alternating optimization algorithm (HOOI or THOOI) does not increase the performances of PCA and TPCA.

VII. Conclusions

This paper discusses a generalization of the higher-order singular value decomposition (HOSVD) over a finite-dimensional commutative algebra C called t-algebra. The elements of the algebra, called t-scalars, are fix-sized multi-way arrays of complex numbers The arrays generalize the complex numbers. Vectors, matrices, and even tensors can be built over the t-algebra and behave analogously to the canonical counterparts over complex numbers. The generalized vectors, matrices, and tensors over t-scalars are called t-vectors, t-matrices, and g-tensors. The modules formed by t-vectors, t-matrices, or g-tensors are the modules over both complex numbers and t-scalars.



A quantitative PSNR comparison of
(2D)²PCA and T-(2D)²PCA with some representative parameter tuples on the identical SVHN image set

method	parameter tuple (r_1, r_2)							
	(3,3)	(7,7)	(11,11)	(15,15)	(19,19)	(23,23)	(27,27)	(31,31)
(2D)²PCA	20.13	25.23	31.39	37.01	41.86	45.98	50.43	59.12
T-(2D)²PCA	20.15	25.24	31.40	37.04	41.89	46.03	50.50	59.27

A quantitative PSNR comparison of
optimzied algorithms with some representative parameter tuples on the identical SVHN image set

method	parameter tuple (r_1, r_2)							
	(3,3)	(7,7)	(11,11)	(15,15)	(19,19)	(23,23)	(27,27)	(31,31)
(2D)²PCA-OP	20.17	25.24	31.39	37.01	41.86	45.98	50.44	59.12
T-(2D)²PCA-OP	21.24	26.73	33.14	38.66	43.74	47.87	52.42	61.16

Fig. 10. A "horizontal" comparision of the results on the same SVHN images by (2D)²PCA and T-(2D)²PCA with or without alternating optimization

This module duality requires t-scalar-valued notions to describe its C-linear aspect and canonical-scalar-valued notions to describe its linear aspect. To describe the linear aspect of t-algebras and t-matrices over algebra, we introduce the standard matrix representation of t-scalars and t-matrices. With the matrix representation, a t-scalar is represented by a diagonal matrix whose diagonal entries are the eigenvalues of the t-scalar as a linear operator. As a result, a t-matrix is representable by a (canonical) block-diagonal matrix. The proposed matrix representation links many t-scalar-valued notions to their canonical versions in matrix theory.

With the introduced matrix representation, this paper shows how to generalize HOSVD (Higher Order Singular Value Decomposition) to THOSVD (t-algebra based HOSVD). THOSVD generalizes SVD and HOSVD [17] and unifies a wide range of PCA-based algorithms, such as PCA, 2DPCA [58], MPCA [56], and their generalized versions over t-scalars, i.e., TPCA [15], [34], T-2DPCA [34], and TMPCA.

The approximation by THOSVD is usually not optimal. Instead, this paper proposes a generalization of the alternating optimization algorithm HOOI (Higher Order Orthogonal Iteration) over t-scalars. The experiments on public data show that the generalized alternating optimization algorithm THOOI (t-algbera based HOOI) improves the results of THOSVD.

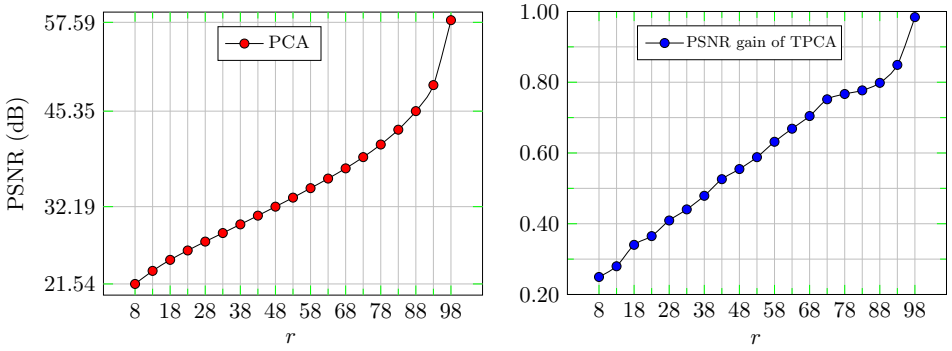
This paper also introduces a pixel neighborhood strategy to extend a "shallower-order" image to its "deeper-order" version. If the pixel neighborhood strategy is nested, one can further increase the image's order. Experiments show that the generalized algorithms using deeper-order t-scalars compare favorably with their counterparts using shallower-order t-scalars. The canonical algorithms are the special cases with the "shallowest order" t-scalars (the order-zero t-scalars, i.e., complex numbers).

The theory of t-vectors, t-matrices, g-tensors provides a consistent, straightforward framework for generalizing many (canonical) matrix and tensor algorithms.

Acknowledgments

Liang Liao and Stephen John Maybank contribute equally to the theory of t-algebra, t-scalars, t-vectors, t-matrices, and g-tensors. Liao Liao designs the experiments of this paper. Other authors help significantly with the experiments presented in this paper. Liang Liao would like to thank the Birkbeck Institute of Data Analytics for free using the high-performance

PSNR by PCA and PSNR gain by TPCA with different r on the same SVHN image set



PSNRs by PCA and TPCA with some representative parameter r on the identical SVHN image set

method	parameter r									
	8	18	28	38	48	58	68	78	88	98
PCA	21.54	24.86	27.38	29.75	32.19	34.74	37.47	40.75	45.35	57.89
TPCA	21.79	25.20	27.78	30.22	32.74	35.37	38.17	41.52	46.15	58.87

Fig. 11. A horizontal comparison of PSNRs by PCA and TPCA on the same SVHN image samples

computing facilities during Liao's visit to Birkbeck College, University of London.

The experiments were designed by Liang Liao, who also analyzed the experiment results. Lun Li and Sen Lin implemented these experiments. They are equally contributing authors. Other authors also significantly contributed to the experiments. This work was partially supported by the National Natural Science Foundation of China grant No. U1404607, the High-end Foreign Experts Program grants GDW20186300351, G20200226015, G2021026018L of the Ministry of Science and Technology of China, the Open Research Program of the National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology with grant number No. 20200206, and the Key Technologies Research & Development Programs of Henan with grant number 222102210016

References

[1] Z. Q. Hong, "Algebraic feature extraction of image for recognition," *Pattern Recognition*, vol. 24, no. 3, pp. 211–219, 1991.

[2] M. Tian, S.-W. Luo, and L.-Z. Liao, "An investigation into using singular value decomposition as a method of image compression," in *2005 International Conference on Machine Learning and Cybernetics*, 8, Ed., 2005, pp. 5200–5204.

[3] M. Dabass, S. Vashisth, and R. Vig, "Lossy color image compression technique using reduced bit plane-quaternion svd," in *2019 9th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, 2019, pp. 504–509.

[4] R. K. Senapati, S. Srivastava, and P. Mankar, "Rst invariant blind image watermarking schemes based on discrete tchebichef transform and singular value decomposition," *Optik - International Journal for Light and Electron Optics*, vol. 45, no. 4, pp. 3331–3353, 2020.

[5] Z. Zainol, J. S. Teh, M. Alawida, A. Alabdulatif *et al.*, "Hybrid svd-based image watermarking schemes: A review," *IEEE Access*, vol. 9, pp. 32 931–32 968, 2021.

[6] K. Batselier, W. Yu, L. Daniel, and N. Wong, "Computing low-rank approximations of large-scale matrices with the tensor network randomized svd," *SIAM Journal on Matrix Analysis and Applications*, 2017.

[7] M. Li, W. Bi, J. T. Kwok, and B.-L. Lu, "Large-scale nyström kernel matrix approximation using randomized svd," *IEEE transactions on neural networks and learning systems*, vol. 26, no. 1, pp. 152–164, 2014.

[8] F. L. Hitchcock, "Multiple invariants and generalized rank of a p-way matrix or tensor," *Journal of Mathematics and Physics*, vol. 7, no. 1-4, pp. 39–79, 1928.

[9] L. R. Tucker, "Implications of factor analysis of three-way matrices for measurement of change," *In Problems in Measuring Change(C W. Harris ed.),University of Wisconsin Press*, pp. 122–137, 1963.

[10] ——, "The extension of factor analysis to three-dimensional matrices," *In Contributions to Mathematical Psychology(H. Gulliksen and N. Frederiksen, eds.)*, pp. 109–127, 1964.

[11] ——, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, pp. 279–311, 1966.

[12] N. D. Sidiropoulos, L. D. Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3551–3582, 2017.

[13] Y. Panagakis, J. Kossaifi, G. G. Chrysos, J. Oldfield, M. A. Nicolaou, A. Anandkumar, and S. Zafeiriou, "Tensor methods in computer vision and deep learning," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 863–890, 2021.

[14] R. K. Renu, V. Sowmya, and K. P. Soman, "Pre-processed hyperspectral image analysis using tensor decomposition techniques," *Advances in Signal Processing and Intelligent Recognition Systems*, pp. 205–216, 2019.

[15] Y. Ren, L. Liao, S. J. Maybank, Y. Zhang, and X. Liu, "Hyperspectral image spectral-spatial feature extraction via tensor principal component analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 9, pp. 1431–1435, 2017.

[16] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Blind source separation by higher-order singular value decomposition," in *Proc. EUSIPCO*, vol. 1, 1994, pp. 175–178.

[17] L. D. Lathauwer, "A multilinear singular value decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.

[18] T. J. Gregor, "Higher order singular value decomposition of tensors for fusion of registered images," *Journal of Electronic Imaging*, vol. 20, no. 1, pp. 9–15, 2011.

[19] J. Liang, H. Yang, L. Ding, and X. Zeng, "Image fusion using higher order singular value decomposition," *IEEE Trans Image Process*, vol. 21, no. 5, pp. 2898–2909, 2012.

[20] X. Zhang, Z. Xu, N. Jia, W. Yang, Q. Feng, W. Chen, and Y. Feng, "Denoising of 3d magnetic resonance images by using higher-order singular value decomposition," *Medical Image Analysis*, vol. 19, no. 1, pp. 75–86, 2015.

[21] X. Geng, L. Ji, Y. Zhao, and F. Wang, "A small target detection method for the hyperspectral image based on higher order singular value decomposition (hosvd)," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 6, pp. 1305–1308, 2013.

[22] J. Shi, X. Zheng, J. Wu, B. Gong, Q. Zhang, and S. Ying, "Quaternion grassmann average network for learning representation of histopathological image," *Pattern Recognition*, vol. 89, pp. 67–76, 2019.

[23] J. Miao, K. I. Kou, and W. Liu, "Low-rank quaternion tensor completion for recovering color videos and images," *Pattern Recognition*, vol. 107, 2020.

[24] J. Miao and K. I. Kou, "Color image recovery using low-rank quaternion matrix completion algorithm," *IEEE Transactions on Image Processing*, vol. 31, pp. 190–201, 2021.

[25] K. M. Hosny and M. M. Darwish, "New set of multi-channel orthogonal moments for color image representation and recognition," *Pattern Recognition*, vol. 88, pp. 153–173, 2019.

[26] Y. Chen, X. Xiao, and Y. Zhou, "Low-rank quaternion approximation for color image processing," *IEEE Transactions on Image Processing*, vol. 29, pp. 1426–1439, 2019.

[27] Y. Yu, Y. Zhang, and S. Yuan, "Quaternion-based weighted nuclear norm minimization for color image denoising," *Neurocomputing*, vol. 332, pp. 283–297, 2019.

[28] C. Zou, K. I. Kou, and Y. Wang, "Quaternion collaborative and sparse representation with application to color face recognition," *IEEE Transactions on image processing*, vol. 25, no. 7, pp. 3287–3302, 2016.

[29] C. Zou, K. I. Kou, L. Dong, X. Zheng, and Y. Y. Tang, "From grayscale to color: Quaternion linear regression for color face recognition," *IEEE Access*, vol. 7, pp. 154 131–154 140, 2019.

[30] J. Miao and K. I. Kou, "Quaternion-based bilinear factor matrix norm minimization for color image inpainting," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5617–5631, 2020.

[31] Z. Jia, M. K. Ng, and G.-J. Song, "Robust quaternion matrix completion with applications to image inpainting," *Numerical Linear Algebra with Applications*, vol. 26, no. 4, p. e2245, 2019.

[32] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra and its Applications*, vol. 435, no. 3, pp. 641–658, 2011.

[33] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 1, pp. 148–172, 2013.

[34] L. Liao and S. J. Maybank, "Generalized visual information analysis via tensorial algebra," *Journal of Mathematical Imaging and Vision*, vol. 62, pp. 560–584, 2020.

[35] ——, "General data analytics with applications to visual information analysis: A provable backward-compatible semisimple paradigm over t-algebra," *arXiv preprint arXiv:2011.00307*, pp. 1–53, 2020.

[36] B. G. Osgood, *Lectures on the Fourier Transform and its Applications*. American Mathematical Soc., 2019, vol. 33.

[37] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[38] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*. Springer, 2014, pp. 818–833.

[39] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015.

[40] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[42] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016, ch. 9.1 The convolution operation.

[43] Z. Zhang and S. Aeron, "Exact tensor completion using t-SVD," *IEEE Transactions on Signal Processing*, vol. 65, no. 6, pp. 1511–1526, 2017.

[44] J. Hou, F. Zhang, H. Qiu, J. Wang, Y. Wang, and D. Meng, "Robust low-tubal-rank tensor recovery from binary measurements," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[45] Q. Jiang and M. Ng, "Robust low-tubal-rank tensor completion via convex optimization," in *Proc. IJCAI*, 2019, pp. 2649–2655.

[46] R. Dian and S. Li, "Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization," *IEEE Transactions on Image Processing*, vol. 28, no. 10, pp. 5135–5146, 2019.

[47] M. Cheng, L. Jing, and M. K. Ng, "Tensor-based low-dimensional representation learning for multi-view clustering," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2399–2414, 2018.

[48] M. Yin, J. Gao, S. Xie, and Y. Guo, "Multiview subspace clustering via tensorial t-product representation," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 3, pp. 851–864, 2018.

[49] S. V. Dolgov and D. V. Savostyanov, "Alternating minimal energy methods for linear systems in higher dimensions," *SIAM Journal on Scientific Computing*, vol. 36, no. 5, pp. A2248–A2271, 2014.

[50] Y. Liu, *Tensors for Data Processing: Theory, Methods, and Applications*. Elsevier, 2021, ch. 1.2.3.

[51] D. Cohen, "On holy wars and a plea for peace," *Computer*, vol. 14, no. 10, pp. 48–54, 1981.

[52] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-$(r_1, r_2, \cdots, r_n)$ approximation of higher-order tensors," *SIAM journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324–1342, 2000.

[53] P. M. Kroonenberg and J. D. Leeuw, "Principal component analysis of three-mode data by means of alternating least squares algorithms," *Psychometrika*, vol. 45, no. 1, pp. 69–97, 1980.

[54] B. N. Sheehan and Y. Saad, "Higher order orthogonal iteration of tensors (HOOI) and its relation to PCA and GLRAM," in *Proceedings of the 2007 SIAM International Conference on Data Mining*. SIAM, 2007, pp. 355–365.

[55] D. Zhang and Z.-H. Zhou, "$(2D)^2$PCA: Two-directional two-dimensional PCA for efficient face representation and recognition," *Neurocomputing*, vol. 69, no. 1-3, pp. 224–231, 2005.

[56] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE transactions on Neural Networks*, vol. 19, no. 1, pp. 18–39, 2008.

[57] L. Liao, X. Zhang, X. Wang, S. Lin, and X. Liu, "Generalized image reconstruction over t-algebra," in *Proceedings of the 2021 3rd International Conference on Advances in Computer Technology*. IEEE, 2021, pp. 387–392.

[58] J. Yang, D. Zhang, A. F. Frangi, and J.-y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 1, pp. 131–137, 2004.