

**GENERAL DATA ANALYTICS
WITH APPLICATIONS TO VISUAL INFORMATION ANALYSIS:
A PROVABLE BACKWARD-COMPATIBLE SEMISIMPLE
PARADIGM OVER T-ALGEBRA**

LIANG LIAO AND STEPHEN JOHN MAYBANK

ABSTRACT. We consider a novel backward-compatible paradigm of general data analytics over a recently-reported semisimple algebra called t-algebra. We study the generalized matrix framework over the t-algebra by representing the elements of t-algebra by fixed-sized multi-way arrays of complex numbers and the algebraic modules over the t-algebra by a finite number of direct-product factors. Over the t-algebra, many algorithms are generalized in a straightforward manner using this new semisimple paradigm. To demonstrate the new paradigm’s performance and its backward-compatibility, we generalize some canonical algorithms for visual pattern analysis. Experiments on public datasets show that the generalized algorithms compare favorably with their canonical counterparts.

CONTENTS

List of Figures	2
1. Introduction	3
1.1. Motivation	3
1.2. Background and Related Work	4
1.3. Contributions of This Work and Organization of This Article	6
2. T-algebra and T-scalars	7
2.1. T-algebra	7
2.2. T-scalars: Generalization of Complex Numbers	10
3. Decomposition of T-algebra via Direct Product	14
3.1. Idempotence	14
3.2. Direct Product and Semisimplicity	14
4. Generalized Matrices over T-algebra and Beyond	17
4.1. T-matrix	17

4.2.	Semisimplicity and Decomposability of A Module of T-matrices	18
4.3.	Optimization	22
5.	Applications of T-matrices in General Visual Information Analysis	23
5.1.	Tensorial Representation of T-matrices	23
5.2.	T-matrix Representation of Legacy Visual Information	23
5.3.	Generalized Low-rank Approximation over C	26
5.4.	Generalized Least-squares over C	27
5.5.	Generalized Principal Component Analysis over C	29
6.	Experimental Verifications	33
6.1.	Generalized Low-rank Approxiamtion	33
6.2.	Generalized Least-squares	37
6.3.	Generalized Principal Component Analysis	38
7.	Conclusion	42
	Acknowledgments	43
	Credits	44
	Open Source	44
	Contact Information	44
	Appendix: Generalization of Supervised Classification and Neural Networks	44
7.1.	Generalized Supervised Classification	44
	TNN: Generalized Nearest Neighbor	45
	TCNN: Generalized Convolutional Neural Networks	47
	References	47

LIST OF FIGURES

5.1	A t-matricization strategy of a 2D legacy image using 3×3 “inception” neighborhoods —— from an input order-two array in $\mathbb{C}^{4 \times 4}$ (i.e., canonical matrix) to an order-four array in $C^{4 \times 4} \equiv \mathbb{C}^{3 \times 3 \times 4 \times 4}$ (i.e., a t-matrix).	24
5.2	Reuse the neighborhood strategy to extend to a legacy grey image (i.e., an array in $\mathbb{R}^{4 \times 4}$) to a low-order compounded image (i.e., an array in $\mathbb{R}^{3 \times 3 \times 4 \times 4}$) and, then to a higher-order compounded image (i.e., an array in $\mathbb{R}^{3 \times 3 \times 3 \times 3 \times 4 \times 4}$).	25

6.1 A comparison of low-rank approximation by SVD and TSVD on the “baboon” image	34
6.2 PSNR (dB) heat maps of TSVD approximations with different t-matrix ranks, characterized by the tuple r_1, r_2, r_3 on the RGB “baboon” image	34
6.3 A comparison of approximations by TSVD, using inception slice and different array orders, and SVD where $N^{entry} = 65536$.	50
6.4 A comparison of approximations by TSVD with $K \in \{9, 81\}$ and SVD with $K = 1$ where PSNRs are computed with $N^{entry} = 65536 \times K$.	50
6.5 A quantitative comparison of approximations in PSNR by canonical least squares and generalized least squares on the ORL image dataset	51
6.6 PSNRs by TPCA with various parameters (r_1, r_2, r_3) , or equivalently, $H_T = \sum_{k=1}^3 r_k \cdot Q_{T,k}$, on the CIFAR-10 images.	51
6.7 PSNRs by PCA and TPCA with various paramters r on the CIFAR-10 images	52
6.8 A comparison of PRNS by PCA and TPCA variants with various t-scalar orders on the CIFAR-10 images. TPCA-I = TPCA with order-two t-scalars, TPCA-II = TPCA with order-four t-scalars, TPCA-III = TPCA with order-six t-scalars, TPCA-1 = TPCA with order-three t-scalars, TPCA-2 = TPCA with order-five t-scalars, TPCA-3 = TPCA with order-seven t-scalars	52
7.1 Generalized classification over C with or without pooling of features.	53
7.2 A diagram of generalized CNN (Convolutional Neural Network) model over C .	53

1. INTRODUCTION

1.1. Motivation. In the big-data deluge era, the canonical matrix and tensor paradigm over an algebraically closed field plays an essential role in many areas, including but not limited to machine learning, computer vision, pattern analysis, and statistic inference. Under the canonical matrix and tensor paradigm, observed data are given in the form of high-order arrays of canonical scalars (i.e., real or complex numbers). For example, an RGB image is a real number array of order three, two orders for the image’s spatial measures, and a third for the image’s spectral measures. An RGB image is also said to have three modes or three-way. A color video sequence of images is of order four, with three orders for spatial-spectral measures and the fourth-order chronological tempo.

Therefore, it is a natural question of whether there exists an extension of the field \mathbb{C} over which a generalized matrix and tensor paradigm can be established and backward-compatible to the canonical paradigm over a field. Fortunately, the answer is yes, but one had to sacrifice at least one of the axioms of a field to obtain something extended.

1.2. Background and Related Work. The most well-known generalization of the field of complex numbers is probably the ring $M_n(\mathbb{C})$ of all $n \times n$ (where n is a positive integer) matrices over complex numbers under the usual matrix addition and multiplication.

The field of complex numbers specializes the ring $M_n(\mathbb{C})$ for $n = 1$. However, when $n \geq 2$, the matrix ring $M_n(\mathbb{C})$ is not a field. Two axioms of a field are sacrificed, (i) not all non-zero matrices are multiplicatively invertible, and (ii) the multiplication is non-commutative.

Besides the matrix ring, some hypercomplex number systems also generalize complex numbers. Among these hypercomplex number systems, well-known is Hamilton’s \mathbb{H} of quaternions, which, up to isomorphism, is a real division subring and subalgebra of $M_2(\mathbb{C})$ [11, 26, 13]. However, the multiplication of quaternions is not commutative.

Most hypercomplex number systems, including Hamilton’s quaternions, are all subalgebras of Clifford algebra, and the fruits of generating complex numbers to obtain something extended. However, Clifford algebra’s hypercomplex number systems are not suitable for general data analytics partially because they are either non-commutative or incompatible with many canonical notions such as euclidean norms. These hypercomplex number systems so far only find narrow niches in geometry and geometry-related branches of physics and computer sciences [12, 1].

To have a well-defined extension of the field \mathbb{R} other than \mathbb{C} , Kilmer et al. proposed a tensorial model called “t-product” for characterizing the multi-way structures of higher-order data. In the “t-product” model, a circulant matrix representation is chosen for its formulation [17, 18, 29, 28].

In the “t-product” model, the generalized scalars are fixed-sized first-order arrays of real numbers. Equipped with a circular-convolution multiplication, a scalar multiplication, and an entry-wise addition, these generalized scalars form a finite-dimensional commutative unital real algebra R .

With the circulant matrix representation over the generalized scalars, many authors have studied and extended the “t-product” model. Gleich et al. [10] investigate the generalized eigenvalues and eigenvectors of matrices over the algebra R and show how the standard power method for finding an eigenvector and the standard Arnoldi method for constructing an orthogonal basis for a Krylov subspace can both be generalized over R . Braman et al. [4] show that the vectors over R form a free module.

Kilmer and Martin show that many properties and structures of canonical matrices and vectors can be generalized. Their examples include transposition, orthogonality, and the singular value decomposition (SVD). The tensor SVD is used to compress tensors. A tensor-based method for image de-blurring is also described. Kilmer et al. [17] generalize the inner product of two vectors, suggesting a notion of the angle between two vectors with elements in R , and define a notion of orthogonality for two vectors. A generalization of the Gram-Schmidt method for generating an orthonormal set of vectors is also studied [17].

Zhang et al. [29] use the tensor SVD to efficiently store video sequences and fill in missing entries in video sequences. Zhang et al. [28] use a randomized version of the tensor SVD to produce low-rank approximations to matrices. Ren and Liao et al. [25] define a tensor version of principal component analysis and extract features from hyperspectral images. The features are classified using standard methods such as support vector machines and nearest neighbors. Liao et al. [21] generalize a sparse representation classifier to tensor data and apply the generalized classifier to image data such as numerals and faces. Chen et al. [5] use a four-dimensional HOSVD (Higher-Order Singular Value Decomposition), one generalization of the matrix singular value decomposition over canonical scalars, to detect changes in a time sequence of hyperspectral images. The K-means clustering algorithm is used to classify the pixel values as changed or unchanged. Fan et al. [9] model a hyperspectral image as the sum of an ideal image, a sparse noise term, and a Gaussian noise term. A product of two low-rank tensors models the ideal image. The low-rank tensors are estimated by minimizing a penalty function obtained by adding the squared errors in a fit of the hyperspectral image to penalty terms for the sparse noise and the sizes of the two low-rank tensors. Lu et al. [23, 22] approximate a third-order tensor using the sum of a low-rank tensor and a sparse tensor. Under suitable conditions, the low-rank tensor and the sparse tensor are recovered exactly.

However, the formulation in circulant matrices is not straightforwardly compatible with the canonical formulation in standard matrices. The elements of real algebra R so far remain as first-order arrays of real numbers. To represent and extend the existing theories via a straightforward compatible approach, Liao and Maybank et al. proposed a framework called “t-matrix” [20, 25] via modules over an algebra C . In the t-matrix framework, generalized scalars are represented by fixed-sized multi-way arrays of complex numbers. These complex arrays can be added in the usual way, but there is no definition of multiplication satisfying the axioms of a field such as \mathbb{R} or \mathbb{C} . However, multiplication based on multi-way circular convolution has many but not all of the properties of a field. Multi-way circular convolution differs from the multiplication in a field in that an infinite number of elements have no multiplicative inverse. These complex arrays form a finite-dimensional commutative algebra C under

the vector addition, scalar multiplication, and convolution-based multiplication. The elements of the algebra C generalize complex numbers and are referred to as t-scalars.

The bijective map by the multi-way Fourier transform shows that the algebra C of t-scalars under the convolution-based multiplication is isomorphic to an algebra of complex arrays of the same size in which the Hadamard product defines the multiplication.

In effect, the algebra, mapped by the Fourier transform, splits into a finite number of copies of \mathbb{C} . This splitting allows the construction of generalized algorithms for analyzing tensorial data without data unraveling. The so-called t-matrices with t-scalar entries have many properties of canonical real or complex matrices. In particular, t-matrices can be scaled with a real or complex number, added and multiplied. There are an additive identity and a multiplicative identity of the algebra C . The generalized rank of a t-matrix is defined by a nonnegative t-scalar, which generalizes the canonical rank of a real or complex matrix, and is a nonnegative element of a partially ordered set of self-conjugate t-scalars. A given t-matrix is invertible if and only if it is square and of full rank over C . The t-matrices include but are not limited to the generalizations of unitary matrices and Hermitian matrices.

1.3. Contributions of This Work and Organization of This Article. This article introduces semisimplicity, a concept in algebra and other algebraical disciplines, to general data analytics, with visual information analysis applications. Launching with a few postulates, one has a generalized paradigm over a semisimple algebra. Using the generalized paradigm, data analytics can be more effective than with the canonical paradigm over real or complex numbers.

This article shows that the semisimple algebra C , called “t-algebra”, generalizes the field \mathbb{C} and can be represented as a direct product of a finite number of simple algebras all isomorphic to the field \mathbb{C} . The semisimplicity of the t-algebra C allows a straightforward backward-compatible generalization of many canonical linear or multilinear structures and algorithms over \mathbb{C} . In the direct product representation of the t-algebra C , its idempotent elements play a critical role. Via the idempotent elements of the t-algebra C , many generalized algebraic notions, including but not limited to generalized scalars (called t-scalars), generalized rank, generalized norm, generalized orthogonality, are reducible to the corresponding canonical notions defined over the field \mathbb{C} . Analogous to their canonical counterparts, generalized matrices over C , called t-matrices, can be scaled, added, multiplied, conjugate transposed, and inverted or pseudo-inverted, in a way backward-compatible to their canonical counterparts defined over \mathbb{C} . The t-algebra C and the t-matrix framework over it allow us to establish a generalized “semisimple” paradigm of data analytics, which is backward-compatible with the canonical paradigm over the field \mathbb{C} .

To demonstrate the “semisimple” paradigm on general visual information analysis, we propose spatial solutions for elevating lower-order visual information to higher-order and pooling higher-order information to lower-order. With the proposed spatial solutions, we adopt generalized algorithms to represent, approximate or analyze images data. Our experiments using the generalized algorithms on public datasets show a provable performance increase compared with the corresponding canonical algorithms’ results. We also give principles on generalizing canonical algorithms and models, including but not limited to CNN (Convolutional Neural Network) for classifying visual patterns. Besides visual information, if appropriate topological information of each data point is known, the “semisimple” paradigm also applies to non-spatially-constrained data.

The remainder of this article is organized as follows. The generalized scalars, called t-scalars, their set called t-algebra, and the generalization of complex numbers are described in Section 2. The idempotent t-scalars and the semisimplicity and decomposability of the t-algebra are discussed in Section 3. Generalized matrices with entries of t-scalars, the semisimplicity and decomposability of the modules, and generalized minimization over the t-algebra are discussed in Section 4. In Section 5, we discuss and demonstrate the principles of applying the semisimple paradigm to generalized visual information analytics. In Section 6, we give provable experimental verifications on public datasets, where results by generalized algorithms compare favorably with the canonical counterparts. We conclude this article in Section 7. Finally, a brief discussion on adopting the proposed paradigm on supervised classification and neural network is given in an appendix.

2. T-ALGEBRA AND T-SCALARS

This work is a continued effort to complete the t-scalar and t-matrix paradigm proposed by Liao and Maybank [20]. The notations, index protocols, symbols, and others in the existing work [20] are followed as much as possible.

For example, all indices begin from 1 rather than 0. Different symbol subscripts other than symbol fonts are used for different data types since there are many data types rather than just canonical scalars, vectors, matrices, and tensors. Interested readers are referred to [20] for more details of these symbol subscripts. For the notations not consistent with those or not yet appearing in [20], we give their descriptions when necessary.

2.1. T-algebra. The t-algebra C also referred to as the ring of t-scalars in [20], generalizes the field of complex numbers \mathbb{C} . It shows that many structures over C are algebraically semisimple and can be defined as a direct product of a finite number of

simple factors over \mathbb{C} . We discuss the semisimplicity and the decomposability of C with more details later in Section 3.

The genesis of the t-algebra and its elements, called t-scalars, are from the following several postulates.

Definition 2.1 (Multi-way array). The generalized scalars, called t-scalars, are order- N arrays of complex numbers belonging to the set $C \equiv \mathbb{C}^{I_1 \times \cdots \times I_N}$.

Definition 2.2 (Addition). The addition of t-scalars is identified with the addition bestowed to linear space namely, given any two t-scalars $X_T, Y_T \in C$, their addition $A_T \doteq X_T + Y_T \in C$ is given by the following complex-entry-wise addition.

$$(A_T)_{i_1, \dots, i_N} = (X_T)_{i_1, \dots, i_N} + (Y_T)_{i_1, \dots, i_N} \in \mathbb{C}, \forall (i_1, \dots, i_N) \in [I_1] \times \cdots \times [I_N] \quad (2.1)$$

where $(X_T)_{i_1, \dots, i_N}$ denotes the (i_1, \dots, i_N) -th complex entry of X_T for all $X_T \in C$ and $[I_n] \doteq \{1, \dots, I_n\}$ for all $n = 1, \dots, N$.

Definition 2.3 (Scalar multiplication). For each a t-scalar $X_T \in C$ and each scalar $\lambda \in \mathbb{C}$, the scalar multiplication $Y_T \doteq \lambda \cdot X_T \in C$ is given by the following entry-wise complex multiplication.

$$(Y_T)_{i_1, \dots, i_N} = \lambda \cdot (X_T)_{i_1, \dots, i_N} \in \mathbb{C}, \forall i_1, \dots, i_N. \quad (2.2)$$

Definition 2.4 (Convolutional multiplication). The convolutional multiplication of a pair of t-scalars is defined by N -way circular convolution — for each pair of two t-scalars $X_T, Y_T \in C$, the product $A_T \doteq X_T \circ Y_T \in C$ is given as follows.

$$(A_T)_{i_1, \dots, i_N} = \sum_{m_1=1}^{I_1} \cdots \sum_{m_N=1}^{I_N} (X_T)_{m_1, \dots, m_N} \cdot (Y_T)_{m'_1, \dots, m'_N} \in \mathbb{C} \quad (2.3)$$

where $m'_n = \text{mod}(i_n - m_n, I_n) + 1$ for all $n \in [N] \doteq \{1, \dots, N\}$.

The product of p copies of t-scalar X_T is also denoted by the shorthand notation X_T^p where $p \geq 1$ is an integer.

The zero t-scalar and the identity t-scalar. T-scalars, under the addition, form an abelian group. The additive identity, denoted by Z_T , is the array of zeros, namely $(Z_T)_{i_1, \dots, i_N} \equiv 0, \forall i_1, \dots, i_N$. It is easy to verify that the t-scalar multiplication is associative, commutative, and distributive to the addition. The multiplicative identity, denoted by E_T , is a t-scalar whose inception entry, with the subscript indices $i_1 = \cdots = i_N = 1$, is equal to 1 and all other entries equal to 0.

Remark 2.5 (T-algebra). The addition and scalar multiplication show that t-scalars form a linear space of dimension $K \doteq I_1 \times \cdots \times I_N$. Under the addition and t-scalar multiplication, C is also a commutative ring. Then, by the definition of algebra, C is a finite-dimensional commutative unital algebra over the field \mathbb{C} . However, C is not a division algebra because not all non-zero t-scalars in C are multiplicatively invertible.

For example, all t-scalars with identical complex entries are not multiplicatively invertible. In other words, C can not be a field or even a skew field. On the other hand, the t-algebra C generalizes the field \mathbb{C} of complex numbers such that C reduces to \mathbb{C} when $I_1 = I_2 = \dots = I_N = 1$.

An equivalence of Definition 2.4 is given by the Hadamard product via the Fourier transform. The equivalence is guaranteed by the convolution theorem [3]. More precisely, the Fourier transform is an isomorphism of algebra $F : (C, +, \cdot, \circ) \rightarrow (C, +, \cdot, *)$ such that for all $X_T, Y_T \in C$, the following condition holds.

$$F(X_T \circ Y_T) = F(X_T) * F(Y_T) \in \mathbb{C}^{I_1 \times \dots \times I_N} \quad (2.4)$$

where $*$ denotes the Hadamard multiplication and is given by entry-wise multiplication of complex arrays $F(X_T)$ and $F(Y_T)$ in $\mathbb{C}^{I_1 \times \dots \times I_N}$. More precisely, let $\tilde{X}_T \doteq F(X_T)$, $\tilde{Y}_T \doteq F(Y_T)$ and $\tilde{C}_T \doteq F(X_T) * F(Y_T)$. Then, $(\tilde{C}_T)_{i_1, \dots, i_N} = (\tilde{X}_T)_{i_1, \dots, i_N} \cdot (\tilde{Y}_T)_{i_1, \dots, i_N} \in \mathbb{C}$ for all i_1, \dots, i_N .

The Fourier transform is an isomorphism defined by the N -mode multiplication of tensors, which sends each element $X_T \in C$ to $\tilde{X}_T \in C$ as follows.

$$\tilde{X}_T \doteq F(X_T) \doteq X_T \times_1 W_{mat}^{(I_1)} \cdots \times_k W_{mat}^{(I_n)} \cdots \times_N W_{mat}^{(I_N)} \in C \equiv \mathbb{C}^{I_1 \times \dots \times I_N} \quad (2.5)$$

where $W_{mat}^{(I_n)} \in \mathbb{C}^{I_n \times I_n}$ denotes the $I_n \times I_n$ Fourier matrix for all $n \in [N]$.

The (m_1, m_2) -th complex entry of the matrix $W_{mat}^{(I_n)}$ is given by

$$\left(W_{mat}^{(I_n)} \right)_{m_1, m_2} = e^{2\pi i \cdot (m_1 - 1) \cdot (m_2 - 1) \cdot I_n^{-1}} \in \mathbb{C}, \quad \text{for all } m_1, m_2. \quad (2.6)$$

The inverse transform $F^{-1} : (C, +, \cdot, *) \rightarrow (C, +, \cdot, \circ)$ is given by the following N -mode multiplication for tensors as follows.

$$X_T \doteq F^{-1}(\tilde{X}_T) = \tilde{X}_T \times_1 \left(W_{mat}^{(I_1)} \right)^{-1} \cdots \times_n \left(W_{mat}^{(I_n)} \right)^{-1} \cdots \times_N \left(W_{mat}^{(I_N)} \right)^{-1} \in C \equiv \mathbb{C}^{I_1 \times \dots \times I_N} \quad (2.7)$$

where $\left(W_{mat}^{(I_n)} \right)^{-1}$ denotes the inverse of the matrix $W_{mat}^{(I_n)}$ for all $n \in [N]$.

By definition, the Fourier transform is a linear mapping, and the following equality holds for all $X_T \in C \equiv \mathbb{C}^{I_1 \times \dots \times I_N}$,

$$\|X_T\|_F = K^{-1/2} \cdot \|\tilde{X}_T\|_F \quad (2.8)$$

where $K \doteq I_1 \times \dots \times I_N$ and $\|\cdot\|_F$ denotes the canonical Frobenius norm for the normed linear space $C \equiv \mathbb{C}^{I_1 \times \dots \times I_N}$.

2.2. T-scalars: Generalization of Complex Numbers. The t-algebra C reduces to the field \mathbb{C} when $I_1 = \dots = I_N = 1$. Besides the fundamental operations addition and multiplication, one can generalize more notions of complex numbers over the t-algebra C .

Conjugation. One of the generalizations is the notion of conjugation over C . The conjugation is an involutory antiautomorphism φ of C such that the following conditions hold

$$\begin{aligned}\varphi(E_T) &= E_T \\ \varphi(\varphi(X_T)) &= X_T \\ \varphi(X_T \circ Y_T) &= \varphi(Y_T) \circ \varphi(X_T) \\ \varphi(\alpha \cdot X_T + \beta \cdot Y_T) &= \bar{\alpha} \cdot \varphi(X_T) + \bar{\beta} \cdot \varphi(Y_T)\end{aligned}\tag{2.9}$$

for all $\alpha, \beta \in \mathbb{C}$ and $X_T, Y_T \in C$.

Also, note that the antiautomorphism φ is also automorphic since the t-algebra C is commutative. In other words, the antiautomorphism condition $\varphi(X_T \circ Y_T) = \varphi(Y_T) \circ \varphi(X_T)$ is equivalent to the automorphism condition $\varphi(X_T \circ Y_T) = \varphi(X_T) \circ \varphi(Y_T)$.

Let the map $\varphi : C \rightarrow C, X_T \mapsto \varphi(X_T)$ be a homomorphism from C to itself, such that

$$(\varphi(X_T))_{i_1, \dots, i_N} = \overline{(X_T)_{m_1, \dots, m_N}} \in \mathbb{C}\tag{2.10}$$

where $m_n \doteq \text{mod}(1 - i_n, I_n) + 1$ for all $n \in [N]$.

The homomorphism φ defined as in equation (2.10) satisfies all the conditions of the notion of conjugation. When $I_1 = \dots = I_N = 1$, the conjugation φ over C reduces to the conjugation over complex numbers.

To comply with the standard notation of *-algebra, we use the notation $X_T^* \doteq \varphi(X_T)$ for the conjugate of a t-scalar $X_T \in C$.¹

Self-conjugate. The conjugate of a t-scalar can be used for characterizing a particular type of t-scalars — a t-scalar X_T is called self-conjugate if $X_T^* = X_T$. It is immediately verified that Z_T and E_T are both self-conjugate. A necessary and sufficient condition for a t-scalar X_T to be self-conjugate is that the Fourier transform $F(X_T)$ is a real array [20].

Let the set of self-conjugate t-scalars be $C^{sc} = \{X_T \in C \mid X_T^* = X_T\}$. The set C^{sc} under the addition, scalar multiplication, and t-scalar multiplication is a subalgebra of C . The t-algebra C is a free complex algebra of dimension $K \doteq I_1 \times \dots \times I_N$. On the other hand, the subalgebra C^{sc} is a free real algebra of the same dimension K over \mathbb{R} since the subalgebra C^{sc} is isomorphic to the algebra $F(C^{sc}) \doteq \{F(X_T) \mid X_T \in C^{sc}\}$. That C^{sc} is a K -dimensional real algebra does not necessarily mean that all

¹The original notation of the conjugate of a t-scalar X_T in [20] is $\text{conj}(X_T)$.

self-conjugate t-scalar must be of real arrays. The subalgebra C^{sc} contains real arrays if the parameters I_1, \dots, I_N are in the set $\{1, 2\}$.

When $I_1 = \dots = I_N = 1$, the subalgebra C^{sc} reduces to the field \mathbb{R} of real numbers. As a generalization of real numbers, C^{sc} helps establish many fundamental notions over C .

Real part and imaginary part. Each t-scalar X_T is representable by two unique self-conjugate t-scalars. More precisely, the equality $X_T \equiv \frac{X_T + X_T^*}{2} + i \cdot \frac{X_T - X_T^*}{2i}$ holds for all $X_T \in C$.

The self-conjugate t-scalars $\frac{X_T + X_T^*}{2}$ and $\frac{X_T - X_T^*}{2i}$ are respectively called the real part and the imaginary part of X_T . Let $Re(X_T) \doteq \frac{X_T + X_T^*}{2}$ and $Im(X_T) \doteq \frac{X_T - X_T^*}{2i}$ for all $X_T \in C$. Then, the following equations hold for all t-scalars $X_T, Y_T \in C$,

$$\begin{aligned} X_T^* &= Re(X_T) - i \cdot Im(X_T) \\ X_T^* \circ X_T &= Re(X_T)^2 + Im(X_T)^2 \\ X_T + Y_T &= (Re(X_T) + Re(Y_T)) + i \cdot (Im(X_T) + Im(Y_T)) \\ X_T \circ Y_T &= (Re(X_T) \circ Re(Y_T) - Im(X_T) \circ Im(Y_T)) + i \cdot (Im(X_T) \circ Re(Y_T) + Re(X_T) \circ Im(Y_T)). \end{aligned} \tag{2.11}$$

Nonnegative t-scalar. Over the subalgebra C^{sc} , one can generalize the notion of nonnegative real numbers — a t-scalar $Y_T \in C^{sc}$ is said nonnegative if and only if there exists a t-scalar $X_T \in C$ such that the condition $Y_T = X_T^* \circ X_T$ holds. It is easy to verify that both Z_T and E_T are nonnegative. Furthermore, any t-scalar in the form of $Re(X_T)^2 + Im(X_T)^2$ is nonnegative. A t-scalar $X_T \in C^{sc}$ is nonnegative iff its Fourier transform $F(X_T)$ only contains nonnegative real entries [20].

Let the set of nonnegative t-scalars be $S^{nonneg} \doteq \{Y_T \mid Y_T = X_T^* \circ X_T, X_T \in C\}$. The set S^{nonneg} is a commutative submonoid of C^{sc} under the t-scalar addition and the t-scalar multiplication. When $I_1 = \dots = I_N = 1$, the monoid S^{nonneg} reduces to the monoid of nonnegative real numbers.²

Partial order. The field \mathbb{R} is a totally ordered set under the usual binary relation “ \leqslant ”. To be a well-behaved generation of the field \mathbb{R} , the algebra C^{sc} needs to be ordered under a binary relation “ \leq ”, defined for at least some pairs of its elements. The notion of S^{nonneg} can help define such the relation “ \leq ”. More formally, the binary relation ‘ \leq ’ on C^{sc} defines a proper subset of the Cartesian product $C^{sc} \times C^{sc}$ such that $X_T \leq Y_T$ if and only if $(Y_T - X_T) \in S^{nonneg}$ for all $X_T, Y_T \in C^{sc}$.

By this definition, it is immediately verified that $Z_T \leq X_T$ for all $X_T \in S^{nonneg}$, namely, the t-scalar Z_T is the least element of S^{nonneg} . The relation $Z_T \leq X_T$ is synonymous with the claim that the t-scalar X_T is nonnegative.

²A monoid is a set equipped with an associative binary operation and an identity element [15].

Nonpositive t-scalar. The binary relation “ \leq ” is reflexive, antisymmetric, and transitive. Those properties qualify “ \leq ” a relation of a partial order. The partial order helps define nonpositive t-scalars —— a t-scalar $X_T \in C^{sc}$ is called nonpositive if and only if $X_T \leq Z_T$, or equivalently, $-1 \cdot X_T \in S^{nonneg}$.

Let $S^{nonpos} \doteq \{-X_T \mid X_T \in S^{nonneg}\} \equiv \{X_T \mid X_T \leq Z_T, X_T \in C^{sc}\}$ be the set of nonpositive t-scalars. The set S^{nonpos} is a monoid under the addition, with the additive identity Z_T being the greatest element of S^{nonpos} . A t-scalar X_T is nonpositive iff its Fourier transform $F(X_T)$ is a nonpositive real array.

Both S^{nonneg} and S^{nonpos} are proper subsets of C^{sc} . There is usually a “gap” between S^{nonneg} and S^{nonpos} such that $S^{nonneg} \cup S^{nonpos} \neq C^{sc}$ unless $I_1 = \dots = I_N = 1$. Given two self-conjugate t-scalars X_T and Y_T , if and only if their subtraction falls in this “gap”, namely $X_T - Y_T \notin (S^{nonneg} \cup S^{nonpos})$, X_T and Y_T are called incomparable under the partial order “ \leq ”.

When $I_1 = \dots = I_N = 1$, the partial order under the relation “ \leq ” reduces to the usual total order of real numbers under the relation “ \leqslant ”. The set S^{nonneg} reduces to the interval $[0, +\infty)$ of real numbers, and S^{nonpos} reduces to the interval $(-\infty, 0]$ of real numbers.

Nonnegative p -th root of a nonnegative t-scalar. For each integer $p \geq 1$ and nonnegative t-scalar Y_T , there is a unique nonnegative t-scalar X_T such that $Y_T = X_T^p$. The proof of the unique existence for $p = 2$ is given in [20], and the proof for $p > 2$ can be given analogously. The nonnegative t-scalar X_T is called the p -th arithmetic root of the nonnegative t-scalar Y_T and is denoted by

$$X_T \doteq \sqrt[p]{Y_T} \equiv Y_T^{1/p}. \quad (2.12)$$

Norm of a t-scalar. The notions of nonnegative t-scalars and nonnegative roots help define the norm of a t-scalar, also called the absolute value of a t-scalar —— for all t-scalar $X_T \in C$, its norm $|X_T| \doteq r(X_T)$ is a nonnegative t-scalar defined by

$$|X_T| \doteq r(X_T) \doteq \sqrt[2]{X_T^* \circ X_T} \equiv \sqrt[2]{\operatorname{Re}(X_T)^2 + \operatorname{Im}(X_T)^2} \in S^{nonneg}. \quad (2.13)$$

It is easy to verify that the following equalities, analogous to their canonical counterparts, hold for all $X_T, Y_T \in C$ and $\alpha \in \mathbb{C}$,

$$\begin{aligned} r(\alpha \cdot X_T) &= |\alpha| \cdot r(X_T) \\ r(X_T) &= Z_T \text{ iff } X_T = Z_T \\ r(X_T \circ Y_T) &= r(X_T) \circ r(Y_T) \\ r(X_T + Y_T) &\leq r(X_T) + r(Y_T) \end{aligned} \quad . \quad (2.14)$$

When $I_1 = \dots = I_N = 1$, the norm $r(\cdot)$ reduces to the absolute value of a complex number.

Inner product of two t-scalars. Following the vein of equation (2.13), one can define the notion of orthogonality for a pair of t-scalars. First, the polarization identity

$$X_T^* \circ Y_T = \frac{1}{4} \cdot \left(r^2(X_T + Y_T) - i \cdot r^2(X_T + i \cdot Y_T) - r^2(X_T - Y_T) + i \cdot r^2(X_T - i \cdot Y_T) \right) \quad (2.15)$$

holds for all t-scalars $X_T, Y_T \in C$.

By the polarization identity, we define the inner product of any pair t-scalars $X_T, Y_T \in C$ by

$$\psi(X_T, Y_T) = X_T^* \circ Y_T . \quad (2.16)$$

The following identities, analogous to their canonical counterparts for a linear space, hold for all $X_T, Y_T, A_T, B_T \in C$,

$$\begin{aligned} r(X_T) &= \sqrt[2]{\psi(X_T, X_T)} \\ \psi(X_T, Y_T) &= (\psi(Y_T, X_T))^* . \quad (2.17) \\ \psi(A_T \circ X_T, B_T \circ Y_T) &= \psi(A_T, B_T) \circ \psi(X_T, Y_T) \\ \psi(X_T + Y_T, A_T + B_T) &= \psi(X_T, A_T) + \psi(X_T, B_T) + \psi(Y_T, A_T) + \psi(Y_T, B_T) \end{aligned}$$

The inner product $\psi : (X_T, Y_T) \mapsto X_T^* \circ Y_T$ is employed to define the notion of orthogonal t-scalars, which is used in the decomposition of the t-algebra C to a finite number of simple algebras. Two t-scalars $X_T, Y_T \in C$ are said to be orthogonal over the t-algebra C , iff their inner product is equal to Z_T , more precisely,

$$\psi(X_T, Y_T) = Z_T . \quad (2.18)$$

The condition $\psi(X_T, Y_T) = Z_T$ is equivalent to the condition $X_T \circ Y_T = Z_T$ for all $X_T, Y_T \in C$. The trivial case of t-scalar orthogonality is that Z_T is orthogonal to all t-scalars. In non-trivial cases of t-scalar orthogonality where X_T and Y_T are not equal to Z_T , both X_T and Y_T must be non-invertible.

Since the t-algebra C is a ring and a linear space, the notion of inner product $\psi : C \times C \rightarrow C$ over the ring C has a canonical counterpart over the linear space C . The canonical inner product $\langle \cdot, \cdot \rangle : C \times C \rightarrow \mathbb{C}$ is a sesquilinear form defined by

$$\langle X_T, Y_T \rangle \doteq \sum_{(i_1, \dots, i_N) \in [I_1] \times \dots \times [I_N]} \overline{(X_T)_{i_1, \dots, i_N}} \cdot (Y_T)_{i_1, \dots, i_N} \quad (2.19)$$

for all t-scalars $X_T, Y_T \in C$.

Two t-scalars X_T and Y_T , as two elements of a linear space, are said orthogonal over the linear space if and only if $\langle X_T, Y_T \rangle = 0$.

Since the t-algebra C is both a linear space and a ring, orthogonality over the linear space C is a companion notion of orthogonality over the ring C . Orthogonality over the

ring is a sufficient condition for orthogonality over the linear space. When $I_1 = \dots = I_N = 1$, the two notions of orthogonalities become identical.

3. DECOMPOSITION OF T-ALGEBRA VIA DIRECT PRODUCT

3.1. Idempotence. The orthogonality introduced by equation (2.18) plays an essential role in decomposing the t-algebra C to a finite number of simple algebras. To this end, we introduce the notion of idempotence over the ring C . An element $P_T \in C$ is called idempotent iff $P_T \circ P_T = P_T$.

It is easy to follow that a t-scalar P_T is idempotent iff its Fourier transform $F(P_T)$ is an array with entries either 0 or 1, and the t-scalars Z_T and E_T are idempotent. All idempotent t-scalars are nonnegative and form a multiplicative monoid with the identity E_T .

Let $S^{idem} \doteq \{X_T \mid X_T \circ X_T = X_T, X_T \in C\}$ be the set of all idempotent t-scalars. The cardinality of S^{idem} is equal to 2^K , where $K \doteq I_1 \times \dots \times I_N$. It also shows that, given any idempotent t-scalar P_T , $(E_T - P_T)$ is also idempotent. Besides being both idempotent, P_T and $E_T - P_T$ are orthogonal. Namely, $\psi(P_T, E_T - P_T) = Z_T$ holds for all $P_T \in S^{idem}$.

Let $P_T^\perp \doteq E_T - P_T$ for all $P_T \in S^{idem}$. Then, it shows that each t-scalar $Y_T \in C$ can be written as a sum of two orthogonal constituents in the form

$$Y_T = P_T \circ Y_T + P_T^\perp \circ Y_T \equiv P_T \circ Y_T + (E_T - P_T) \circ Y_T \quad (3.1)$$

such that $\psi(P_T \circ Y_T, P_T^\perp \circ Y_T) = Z_T$ for all $P_T \in S^{idem}$ and $Y_T \in C$.

Further, the following equalities hold for all $\lambda \in \mathbb{C}$ and $X_T, Y_T \in C$.

$$\begin{aligned} X_T^* &= (P_T \circ X_T)^* + (P_T^\perp \circ X_T)^* \\ r(X_T) &= r(P_T \circ X_T) + r(P_T^\perp \circ X_T) \\ \lambda \cdot X_T &= \lambda \cdot (P_T \circ X_T) + \lambda \cdot (P_T^\perp \circ X_T) \\ X_T + Y_T &= (P_T \circ X_T + P_T \circ Y_T) + (P_T^\perp \circ X_T + P_T^\perp \circ Y_T) \quad . \\ \psi(X_T, Y_T) &= \psi(P_T \circ X_T, P_T \circ Y_T) + \psi(P_T^\perp \circ X_T, P_T^\perp \circ Y_T) \\ X_T \circ Y_T &= (P_T \circ X_T) \circ (P_T \circ Y_T) + (P_T^\perp \circ X_T) \circ (P_T^\perp \circ Y_T) \\ X_T \leq Y_T &\Leftrightarrow P_T \circ X_T \leq P_T \circ Y_T \text{ and } P_T^\perp \circ X_T \leq P_T^\perp \circ Y_T \end{aligned} \quad (3.2)$$

3.2. Direct Product and Semisimplicity. By the observations as in equation (3.2), the t-algebra C is a direct product of two algebras $C(P_T)$ and $C(P_T^\perp)$, written as follows.

$$C = C(P_T) \times C(P_T^\perp) \quad (3.3)$$

where $C(P_T) \doteq P_T \circ C \doteq \{P_T \circ Y_T \mid Y_T \in C\}$ and $C(P_T^\perp) \doteq P_T^\perp \circ C \doteq \{P_T^\perp \circ Y_T \mid Y_T \in C\}$ for all $P_T \in S^{idem}$.³

For all $P_T \in S^{idem}$, under the t-scalar addition and multiplication, $C(P_T)$ and $C(P_T^\perp)$ are both principal ideals of the underlying ring of C . It implies that $C(P_T)$ and $C(P_T^\perp)$ are closed under the t-scalar addition and multiplication.

Under the t-scalar addition, $C(P_T)$ is a subgroup of the underlying additive group of C . On the other hand, the equality $X_T \circ P_T = X_T$ holds for all $X_T \in C(P_T)$. Hence, $C(P_T)$ is a ring with the additive identity Z_T and the multiplicative identity P_T . A similar conclusion is obtained that $C(P_T^\perp)$ is also a ring with the additive identity Z_T and the multiplicative identity P_T^\perp .

However, usually, neither $C(P_T)$ nor $C(P_T^\perp)$ is a subring of C since E_T is not an element of either $C(P_T)$ or $C(P_T^\perp)$ unless that one of $C(P_T)$ and $C(P_T^\perp)$ is equal to C , and the other is just a singleton set $\{Z_T\}$.

Orthogonal algebras. The algebra $C(P_T^\perp)$ is the orthogonal complement of the algebra $C(P_T)$ in the sense that

$$\begin{aligned} C(P_T) \cap C(P_T^\perp) &= \{Z_T\} \\ C(P_T^\perp) &\equiv \{X_T \in C \mid \psi(X_T, Y_T) = Z_T, \forall Y_T \in C(P_T)\} \end{aligned} \quad . \quad (3.4)$$

Primitive idempotent t-scalars. By equation (3.1), each idempotent t-scalar P_T can be written as a sum of two orthogonal idempotent t-scalars, more precisely,

$$P_T = X_T + Y_T \quad (3.5)$$

such that $X_T, Y_T \in S^{idem}$ and $\psi(X_T, Y_T) = Z_T$ for all $P_T \in S^{idem}$.

An idempotent t-scalar P_T is called primitive, if and only if P_T can not be written as a sum of two non-zero orthogonal idempotent t-scalars. By definition, it is easy to show that Z_T is a primitive idempotent t-scalar.

There are $K \doteq I_1 \times \cdots \times I_N$ non-zero primitive idempotent t-scalars. Let $Q_{T,1}, \dots, Q_{T,K}$ be these t-scalars, and $S^{pidem} \doteq \{Q_{T,1}, \dots, Q_{T,K}\}$ be the set of them. It is easy to verify that the Fourier transform $F(Q_{T,k})$ contains only one entry of 1 and other entries of 0 for all $Q_{T,k} \in S^{pidem}$.

Any two elements of S^{pidem} are orthogonal to each other and are incomparable under the partial order “ \leq ”. Each element of S^{pidem} is multiplicatively non-invertible and one of the K minimal elements of the poset $S^{idem} \setminus \{Z_T\}$. Furthermore, the following

³For most finite-dimensional algebraic structures, the notions of direct product and direct sum are equivalent. However, on the underlying ring of C , the two notions are not equivalent. Interested readers are referred to [19] for a relevant discussion.

identities hold

$$E_T = \sum_{k=1}^K Q_{T,k} \quad (3.6)$$

and

$$Y_T \equiv Y_T \circ E_T \equiv \sum_{k=1}^K Y_T \circ Q_{T,k}, \quad \forall Y_T \in C. \quad (3.7)$$

These primitive elements $Q_{T,1}, \dots, Q_{T,K}$ play an essential role in decomposing C and all C -modules.

Via direct product, one can discuss the semisimplicity of C and the modules over C . Semisimplicity is a concept with a rigorous definition in mathematical disciplines such as linear algebra, abstract algebra, representation theory, etc. In brief, a semisimple object can be represented as a non-trivial direct product of simple objects. Simple objects are non-representable by a non-trivial direct product.

In this article, we are only concerned with semisimple algebras and modules. One can find a rigorous definition of semisimple algebras and semisimple modules in Chapter 4 of a recent book by Karin Erdmann and Thorsten Holm [8] or equivalently in Chapter IX of Thomas Hungerford's time-honored book on algebra [16]. For the reader's convenience, the relevant definition of semisimplicity is given as follows.

Definition 3.1 (Simple algebra). A non-zero algebra is called simple or irreducible if the algebra has no two-sided ideals besides the zero ideal and itself.

Definition 3.2 (Semisimple algebra). If a non-zero algebra is a direct product of simple algebras, then the algebra is called semisimple.

A simple algebra is a special case of semisimple algebras in the sense that, up to isomorphism, a simple algebra can be written as a direct product of the zero algebra and itself.

It immediately follows that a non-zero algebra $C(P_T) \doteq P_T \circ C$ is simple if and only if the non-zero idempotent t-scalar P_T is primitive, namely, $P_T \in S^{pidem}$. It is immediately verified that a non-trivial semisimple t-algebra C is the direct product of K orthogonal factors as follows.

$$C = C(Q_{T,1}) \times \cdots \times C(Q_{T,K}) \doteq \prod_{k=1}^K C(Q_{T,k}) \quad (3.8)$$

where $C(Q_{T,k}) \doteq Q_{T,k} \circ C \doteq \{Q_{T,k} \circ X_T \mid X_T \in C\}$ for all $Q_{T,k} \in S^{pidem}$.

By definition, the field \mathbb{C} of complex numbers is also a one-dimensional algebra. It is easy to follow that each algebra $C(Q_{T,k})$ is isomorphic to the algebra \mathbb{C} for all $Q_{T,k} \in S^{pidem}$. Hence, the following isomorphism holds in the form of a direct product

$$C \doteq \prod_{k=1}^K C(Q_{T,k}) \cong \prod_{k=1}^K \mathbb{C}. \quad (3.9)$$

Orthogonality series. Following the vein of equation (3.8), it shows that the set $S^{pidem} \doteq \{Q_{T,1}, \dots, Q_{T,K}\}$ is a generating-set of C . More precisely, each t-scalar $Y_T \in C$ is a linear combination of $Q_{T,1}, \dots, Q_{T,K}$ in the following form

$$Y_T \equiv Y_T \circ E_T = \sum_{k=1}^K Y_T \circ Q_{T,k} \equiv \sum_{k=1}^K \tau(Q_{T,k}, Y_T) \cdot Q_{T,k} \quad (3.10)$$

where

$$\tau(Q_{T,k}, Y_T) \doteq K \cdot \langle Q_{T,k}, Y_T \rangle \quad (3.11)$$

is the k -th complex coordinate of $Y_T \in C$ in terms of $Q_{T,k} \in S^{pidem}$.

It is easy to follow the primitive idempotent t-scalars $Q_{T,1}, \dots, Q_{T,K}$ are orthogonal basis vectors of the underlying vector space of C , and the Gram matrix of $Q_{T,1}, \dots, Q_{T,K}$ is given by $G_{mat} \doteq (\langle Q_{T,k}, Q_{T,k'} \rangle) = K^{-1} \cdot I_{mat}$ where I_{mat} denotes the $K \times K$ identity matrix.

By the nature of equation (3.9), following equalities hold for all $\lambda \in \mathbb{C}$ and $X_T, Y_T \in C$,

$$\begin{aligned} X_T^* &= \sum_{k=1}^K \overline{\tau(Q_{T,k}, Y_T)} \cdot Q_{T,k} \\ r(X_T) &= \sum_{k=1}^K |\tau(Q_{T,k}, Y_T)| \cdot Q_{T,k} \\ \lambda \cdot X_T &= \sum_{k=1}^K (\lambda \cdot \tau(Q_{T,k}, Y_T)) \cdot Q_{T,k} \\ X_T + Y_T &= \sum_{k=1}^K (\tau(Q_{T,k}, X_T) + \tau(Q_{T,k}, Y_T)) \cdot Q_{T,k} \quad (3.12) \\ \psi(X_T, Y_T) &\doteq X_T^* \circ Y_T = \sum_{k=1}^K (\overline{\tau(Q_{T,k}, X_T)} \cdot \tau(Q_{T,k}, Y_T)) \cdot Q_{T,k} \\ X_T \circ Y_T &= \sum_{k=1}^K (\tau(Q_{T,k}, X_T) \cdot \tau(Q_{T,k}, Y_T)) \cdot Q_{T,k} \\ X_T \leq Y_T &\Leftrightarrow \tau(Q_{T,k}, X_T) \leq \tau(Q_{T,k}, Y_T), \forall k \in [K]. \end{aligned}$$

These equalities in equation (3.12) are analogous to those in equation (3.2).

4. GENERALIZED MATRICES OVER T-ALGEBRA AND BEYOND

4.1. T-matrix. With various notions defined on C , one can establish algebraic structures over C . For example, one can define matrices over C , which generalizes matrices over \mathbb{C} (i.e., complex matrices). A matrix over C , called t-matrix, is a rectangular array of t-scalars arranged in rows and columns. T-matrices follow the same algebraic principles of complex matrices and hence are backward-compatible to complex matrices [20].

For instance, for each t-matrix $X_{TM} \in C^{M_1 \times M_2}$, let $(X_{TM})_{m_1, m_2}$ be its (m_1, m_2) -th t-scalar entry for all $(m_1, m_2) \in [M_1] \times [M_2]$. Then, some operations on t-matrices are given as follows.

T-matrix addition. The t-matrix addition $X_{TM} + Y_{TM}$ for all $X_{TM}, Y_{TM} \in C^{M_1 \times M_2}$ is a t-matrix in $C^{M_1 \times M_2}$ such that

$$(X_{TM} + Y_{TM})_{m_1, m_2} = (X_{TM})_{m_1, m_2} + (Y_{TM})_{m_1, m_2} \in C, \forall (m_1, m_2) \in [M_1] \times [M_2]. \quad (4.1)$$

T-scalar multiplication. The t-scalar multiplication $\lambda_T \circ X_{TM} \in C^{M_1 \times M_2}$ for all $\lambda_T \in C$, $X_{TM} \in C^{M_1 \times M_2}$ is a t-matrix in $C^{M_1 \times M_2}$ such that

$$(\lambda_T \circ X_{TM})_{m_1, m_2} = \lambda_T \circ (X_{TM})_{m_1, m_2} \in C, \forall m_1, m_2. \quad (4.2)$$

T-matrix multiplication. The t-matrix multiplication $X_{TM} \circ Y_{TM}$ for all $X_{TM} \in C^{M_1 \times M}$, $Y_{TM} \in C^{M \times M_2}$ is a t-matrix in $C^{M_1 \times M_2}$ such that

$$(X_{TM} \circ Y_{TM})_{m_1, m_2} = \sum_{m=1}^M (X_{TM})_{m_1, m} \circ (Y_{TM})_{m, m_2} \in C, \forall m_1, m_2. \quad (4.3)$$

Scalar multiplication. The scalar multiplication $\lambda \cdot X_{TM}$ for all $\lambda \in \mathbb{C}$, $X_{TM} \in C^{M_1 \times M_2}$ is a t-matrix in $C^{M_1 \times M_2}$ such that

$$(\lambda \cdot X_{TM})_{m_1, m_2} = \lambda \cdot (X_{TM})_{m_1, m_2} \in C, \forall m_1, m_2. \quad (4.4)$$

Conjugate transpose of a t-matrix. The conjugate transpose X_{TM}^* (the original notation in [20] is X_{TM}^H) of a t-matrix $X_{TM} \in C^{M_1 \times M_2}$ is a t-matrix in $C^{M_2 \times M_1}$ such that

$$(X_{TM}^*)_{m_2, m_1} = (X_{TM})_{m_1, m_2}^* \in C, \forall m_1, m_2. \quad (4.5)$$

Multiplication of a matrix and a t-scalar. The multiplication $Y_{TM} \doteq Y_{mat} \ltimes X_T$ is a t-matrix in $C^{M_1 \times M_2}$ for all $Y_{mat} \in \mathbb{C}^{M_1 \times M_2}$ and $X_T \in C$ such that the (m_1, m_2) -th t-scalar entry of the product Y_{TM} is given by

$$(Y_{TM})_{m_1, m_2} = (Y_{mat})_{m_1, m_2} \cdot X_T \in C, \forall m_1, m_2 \quad (4.6)$$

where $(Y_{mat})_{m_1, m_2}$ denotes the (m_1, m_2) -th complex entry of the matrix Y_{mat} .

Equation (4.6) extends equation (4.4) since the former reduces to the latter when $M_1 = M_2 = 1$. One has the notion of t-vector via the notion of t-matrix —— a t-matrix $X_{TM} \in C^{M_1 \times M_2}$ reduces to a t-vector in $X_{TV} \in C^{M_1 \times 1} \equiv C^{M_1}$ when $M_2 = 1$.

All t-matrices of the same size form a module over the ring C by equations (4.1) and (4.2). They also form a vector space over the field \mathbb{C} by equations (4.1) and (4.4).

4.2. Semisimplicity and Decomposability of A Module of T-matrices. The semisimplicity of a module over an algebra can be defined analogously to the semisimplicity of an algebra. Here, we rephrase the definition of module semisimplicity given in [8] as follows.

Definition 4.1 (Simple module). A non-zero module is called simple or irreducible if it has no submodule besides the zero submodule and itself.

Definition 4.2 (Semisimple module). A non-zero module is called semisimple if the module is a direct product of simple modules.

Every simple module is a special case of semisimple module since, up to isomorphism, a simple module can always be written as a direct product of the zero submodule and itself.

A non-trivial C -module of can be written as a direct product via a finite number of primitive idempotent elements. More concretely, let $G \equiv C^{M_1 \times M_2}$ be a non-trivial module over $C \equiv \mathbb{C}^{I_1 \times \cdots \times I_N}$ and $Q_{T,1}, \dots, Q_{T,K}$ be the primitive idempotent elements of C . Following the direct product in equation (3.8), the module G is semisimple and hence a direct product by K simple submodules as follows.

$$G = G(Q_{T,1}) \times \cdots \times G(Q_{T,K}) \doteq \prod_{k=1}^K G(Q_{T,k}) \quad (4.7)$$

where $G(Q_{T,k}) \doteq Q_{T,k} \circ G \doteq \{Q_{T,k} \circ X_{TM} \mid X_{TM} \in G\}$ for all $k \in [K]$.

These submodules are simple such that none of them can be written as a direct product of two non-trivial proper submodules. Furthermore, the submodules $Q_{T,1}, \dots, Q_{T,K}$ are orthogonal to each other.

One can also extend the notion of inner product ψ over C to over $G \equiv C^{M_1 \times M_2}$. The extended inner product is a C -sesquilinear form $\psi : G \times G \rightarrow C$ defined as follows.

$$\psi(X_{TM}, Y_{TM}) \doteq \sum_{m_1, m_2} (X_{TM})_{m_1, m_2}^* \circ (Y_{TM})_{m_1, m_2} \in C, \forall X_{TM}, Y_{TM} \in G. \quad (4.8)$$

Orthogonality on C -module. Two t-matrices X_{TM}, Y_{TM} are said orthogonal on G iff their inner product is equal to Z_T , namely, $\psi(X_{TM}, Y_{TM}) = Z_T$. It follows that any two submodules $G(Q_{T,k})$ and $G(Q_{T,k'})$ with $k \neq k'$ in equation (4.7), are orthogonal in the following sense

$$\psi(X_{TM}, Y_{TM}) = Z_T, \forall X_{TM} \in G(Q_{T,k}), Y_{TM} \in G(Q_{T,k'}). \quad (4.9)$$

Orthogonality series. In the light of equation (3.10), for all t-matrix $Y_{TM} \in G \equiv C^{M_1 \times M_2}$, the t-matrix Y_{TM} can be written as a unique series as follows

$$Y_{TM} = \sum_{k=1}^K Y_{TM,k} \doteq \sum_{k=1}^K Y_{mat,k} \ltimes Q_{T,k} \doteq \sum_{k=1}^K f_k(Y_{TM}) \ltimes Q_{T,k}. \quad (4.10)$$

where $f_k(Y_{TM}) \doteq Y_{mat,k} \in \mathbb{C}^{M_1 \times M_2}$ is the k -th matrix constituent, such that the (m_1, m_2) -th complex entry of $Y_{mat,k}$ is given by

$$(Y_{mat,k})_{m_1, m_2} = \tau(Q_{T,k}, (Y_{TM})_{m_1, m_2}) \doteq K \cdot \langle Q_{T,k}, (Y_{TM})_{m_1, m_2} \rangle \in \mathbb{C} \quad (4.11)$$

for all $(k, m_1, m_2) \in [K] \times [M_1] \times [M_2]$.

It shows that the t-matrices $Y_{TM,k} \doteq Y_{mat,k} \times Q_{T,k}$ and $Y_{TM,k'} \doteq Y_{mat,k'} \times Q_{T,k'}$ are orthogonal on the module G and its underlying vector space for all $k \neq k'$, more

symbolically, the following equalities hold for all $k \neq k'$,

$$\begin{aligned} \langle Y_{TM,k}, Y_{TM,k'} \rangle &= 0 \\ \psi(Y_{TM,k}, Y_{TM,k'}) &= Z_T \\ Y_{TM,k} \cap Y_{TM,k'} &= \{Z_T\} \end{aligned} \quad (4.12)$$

Equation (4.10) is called the orthogonality series of Y_{TM} .

For $X_{TM} = \sum_{k=1}^K X_{mat,k} \ltimes Q_{T,k} \in G$, $Y_{TM} = \sum_{k=1}^K Y_{mat,k} \ltimes Q_{T,k} \in G$, $\lambda_T \in C$ and $\alpha \in \mathbb{C}$, it is easy to verify the following equalities hold

$$\begin{aligned} X_{TM}^* &= \sum_{k=1}^K X_{mat,k}^* \ltimes Q_{T,k} \\ \alpha \cdot X_T &= \sum_{k=1}^K (\alpha \cdot X_{mat,k}) \ltimes Q_{T,k} \\ \lambda_T \circ X_T &= \sum_{k=1}^K (\tau(Q_{T,k}, \lambda_T) \cdot X_{mat,k}) \ltimes Q_{T,k} \\ X_{TM} + Y_{TM} &= \sum_{k=1}^K (X_{mat,k} + Y_{mat,k}) \ltimes Q_{T,k} \\ \psi(X_{TM}, Y_{TM}) &= \sum_{k=1}^K \langle X_{mat,k}, Y_{mat,k} \rangle \ltimes Q_{T,k} \\ X_{TM} \circ Y_{TM} &= \sum_{k=1}^K (X_{mat,k} \cdot Y_{mat,k}) \ltimes Q_{T,k} \end{aligned} \quad (4.13)$$

It shows that these equalities in equation (4.13) are analogous to those in equations (3.2) and (3.12).

Equation (4.13) shows that an operation or a notion on a t-matrix $X_{TM} = \sum_{k=1}^K X_{mat,k} \ltimes Q_{T,k}$ is reducible to its canonical counterparts $X_{mat,k}$ for all $k \in [K]$. This helps define more notions of t-matrices.

Singular value decomposition of a t-matrix. One can give the notion of singular value decomposition on a t-matrix. Given a t-matrix $Y_{TM} = \sum_{k=1}^K Y_{mat,k} \ltimes Q_{T,k} \in C^{M_1 \times M_2}$, let the singular value decomposition (SVD) of the k -th matrix constituent $Y_{mat,k}$ of the t-matrix Y_{TM} be

$$Y_{mat,k} = U_{mat,k} \cdot S_{mat,k} \cdot V_{mat,k}^*, \quad \forall k \in [K] \quad (4.14)$$

where $M \doteq \min(M_1, M_2)$, $U_{mat,k} \in \mathbb{C}^{M_1 \times M}$, $V_{mat,k} \in \mathbb{C}^{M_2 \times M}$ and $S_{mat,k} \doteq \text{diag}(\lambda_1^{(k)}, \dots, \lambda_M^{(k)})$ such that $U_{mat,k}^* \cdot U_{mat,k} = V_{mat,k}^* \cdot V_{mat,k} = I_{mat}$ and $\lambda_1^{(k)} \geq \dots \geq \lambda_M^{(k)} \geq 0$ for all $k \in [K]$.

Then, the following t-matrices are given

$$\begin{aligned} U_{TM} &\doteq \sum_{k=1}^K U_{mat,k} \ltimes Q_{T,k} \in C^{M_1 \times M} \\ V_{TM} &\doteq \sum_{k=1}^K V_{mat,k} \ltimes Q_{T,k} \in C^{M_2 \times M} \\ S_{TM} &\doteq \sum_{k=1}^K S_{mat,k} \ltimes Q_{T,k} \doteq \text{diag}(\lambda_{T,1}, \dots, \lambda_{T,M}) \in C^{M \times M} \end{aligned} \quad (4.15)$$

It follows that the following equality holds

$$Y_{TM} = U_{TM} \circ S_{TM} \circ V_{TM}^* \quad (4.16)$$

where

$$\begin{aligned} U_{TM}^* \circ U_{TM} &= V_{TM}^* \circ V_{TM} = I_{TM} \doteq \text{diag}(E_T, \dots, E_T), \\ \lambda_{T,1} &\geq \dots \geq \lambda_{T,M} \geq Z_T . \end{aligned} \quad (4.17)$$

Equation (4.16), called TSVD (Tensorial Singular Value Decomposition), is a higher-order generalization of a matrix's singular value decomposition. When $I_1 = \dots = I_N = 1$, equation (4.16) reduces to the canonical singular value decomposition of a matrix. When $I_1 > 1$ and $I_2 = \dots = I_N = 1$, equation (4.16) reduces to Kilmer's version of SVD called t-SVD in [18].

Pseudoinverse of a t-matrix. The notion of pseudoinverse can also be defined via the orthogonality series in equation (4.10).

The pseudoinverse of a t-matrix $Y_{TM} = \sum_{k=1}^K Y_{mat,k} \ltimes Q_{T,k} \in C^{M_1 \times M_2}$ is given by

$$Y_{TM}^+ = \sum_{k=1}^K Y_{mat,k}^+ \ltimes Q_{T,k} \quad (4.18)$$

where $Y_{mat,k}^+$ denotes the canonical Moore-Penrose inverse of $Y_{mat,k}$ for all $k \in [K]$.⁴

Rank of a t-matrix. The notions of TSVD and pseudoinverse of a t-matrix helps define the rank of a t-matrix. Let $Y_{TM} = U_{TM} \circ S_{TM} \circ V_{TM} \in C^{M_1 \times M_2}$ be the compact TSVD of Y_{TM} , namely S_{TM} is a diagonal t-matrix in $C^{M \times M}$ and $M \doteq \min(M_1, M_2)$, the rank of Y_{TM} is given by

$$\text{rank}(Y_{TM}) \doteq \text{trace}(S_{TM} \circ S_{TM}^+) \quad (4.19)$$

where $\text{trace}(\cdot)$ returns the sum of diagonal t-scalar entries of a square t-matrix.

It show that the inequality $Z_T \leq \text{rank}(Y_{TM}) \leq M \cdot E_T$ holds for all $Y_{TM} \in C^{M_1 \times M_2}$. On the other hand, for all t-matrix $Y_{TM} \doteq \sum_{k=1}^K Y_{mat,k} \ltimes Q_{T,k} \in C^{M_1 \times M_2}$, one has the following equality

$$\text{rank}(Y_{TM}) = \sum_{k=1}^K \text{rank}(Y_{mat,k}) \cdot Q_{T,k} . \quad (4.20)$$

When $I_1 = \dots = I_N = 1$, equation (4.19) reduces to the rank of a $M_1 \times M_2$ complex matrix. When $M_1 = M_2 = 1$, equation (4.19) reduces to the rank of a t-scalar, i.e., an idempotent t-scalar in S^{idem} . Namely, there are 2^K different possible values as the rank of a t-scalar where $K \doteq I_1 \times \dots \times I_N$.

If and only if a t-scalar $X_T \in C$ is multiplicatively invertible, the rank of X_T is equal to E_T , i.e., the greatest element of the poset S^{idem} . Hence, the invertible t-scalar X_T is also equivalently called of full rank.

Furthermore, the rank of a t-matrix $Y_{TM} \in C^{M_1 \times M_2}$ is reducible to the ranks of its singular values. Namely, let the generalized singular values of Y_{TM} be $\lambda_{T,1}, \dots, \lambda_{T,M}$.

⁴The pseudoinverse can be of any type. However, in this article, we only discuss the Moore-Penrose inverse and its generalization over C .

where $M \doteq \min(M_1, M_2)$, the following equality holds

$$\text{rank}(Y_{TM}) = \sum_{k=1}^M \text{rank}(\lambda_{T,k}) . \quad (4.21)$$

Generalized Frobenius norm. Another fundamental notion for general data analytics is the norm of a t-matrix. Similar to its canonical counterparts, a t-matrix can have different types of norms. Among them is the so-called Frobenius norm — given a t-matrix $X_{TM} \in C^{M_1 \times M_2}$, its Frobenius norm, analogous to its canonical counterpart, is defined as follows

$$r(X_{TM})_F = \sqrt[2]{\psi(X_{TM}, X_{TM})} \in S^{\text{nonneg}} . \quad (4.22)$$

When $M_2 = 1$, $X_{TM} \in C^{M_1 \times M_2}$ reduces to a t-vector denoted by X_{TV} . The norm in equation (4.22) in this context is denoted by $r(X_{TV})_2 \in S^{\text{nonneg}}$.

Generalized distance. Equation (4.22) help extend the notion of distance over C . For each pair of t-matrices $X_{TM}, Y_{TM} \in G \equiv C^{M_1 \times M_2}$, a generalized distance between X_{TM} and Y_{TM} is defined by a nonnegative t-scalar as follows

$$d(X_{TM}, Y_{TM}) = r(X_{TM} - Y_{TM})_F \in S^{\text{nonneg}} . \quad (4.23)$$

It is easy to prove that the following conditions, analogous to the axioms of canonical distance, hold for all $X_{TM}, Y_{TM}, X'_{TM} \in G$,

$$\begin{aligned} d(X_{TM}, Y_{TM}) &= d(Y_{TM}, X_{TM}) \\ d(X_{TM}, Y_{TM}) = Z_T &\Leftrightarrow X_{TM} = Y_{TM} . \\ d(X_{TM}, X'_{TM}) \leq d(X_{TM}, Y_{TM}) + d(Y_{TM}, X'_{TM}) \end{aligned} \quad (4.24)$$

Remark 4.3 (Generalized metric space). The pair (G, d) generalizes the notion of metric space, and the function d , called t-metric, sends any pair of elements in G to an element of the poset S^{nonneg} .

4.3. Optimization. One can minimize a nonnegative-tscalar-valued function characterized by the generalized notions over C .⁵

Given a function $f : A \rightarrow S^{\text{nonneg}}$, if the range $f(A)$, a poset, has the least element under partial order “ \leq ”, namely $\inf f(A) \doteq \min_{\alpha \in A} f(\alpha) \in f(A)$, and f is an injection, there is a unique element $\hat{\alpha} \in A$ satisfying $f(\hat{\alpha}) = \inf f(A)$. In this scenario, the notation $\hat{\alpha} \doteq \operatorname{argmin}_{\alpha \in A} f(\alpha)$ is used.

The function $f : \alpha \mapsto f(\alpha) \in S^{\text{nonneg}}$ can also be written in the following form

$$f(\alpha) = \sum_{k=1}^K \tau(Q_{T,k}, f(\alpha)) \cdot Q_{T,k} \quad (4.25)$$

⁵A maximization problem can be reformulated as a minimization problem. Hence, we only discuss the minimization problems over C in this subsection.

where $\tau(Q_{T,k}, f(\alpha)) \doteq K \cdot \langle Q_{T,k}, f(\alpha) \rangle \geq 0$ for all $k \in [K]$.

If the range $f(A)$ has the least element, the unique element $\inf f(A) \doteq \min_{\alpha \in A} f(\alpha)$ can be written as follows

$$\inf f(A) \doteq \min_{\alpha \in A} f(\alpha) = \sum_{k=1}^K \left(\min_{\alpha \in A} \tau(Q_{T,k}, f(\alpha)) \right) \cdot Q_{T,k} . \quad (4.26)$$

In other words, one can seek a minimizer to $f(\alpha)$ via investigating the minimizers to a finite number of canonical subfunctions $\tau(Q_{T,1}, f(\alpha)), \dots, \tau(Q_{T,K}, f(\alpha))$.

5. APPLICATIONS OF T-MATRICES IN GENERAL VISUAL INFORMATION ANALYSIS

5.1. Tensorial Representation of T-matrices. A t-matrix is an order-two array of t-scalars. On the other hand, each t-scalar entry of a t-matrix can be represented by a fixed-sized order- N complex array. Thus, a convenient numerical representation of a t-matrix is an order- $(N + 2)$ array of complex numbers.

There are many equivalent tensorial representations for t-matrices. Following the convention in [20], we represent a t-matrix $X_{TM} \in C^{M_1 \times M_2}$ by a complex array in $C^{I_1 \times \dots \times I_N \times M_1 \times M_2}$.

T-matrices can characterize much information. For example, an RGB image in the form of an order-three real array can be characterized by a t-matrix of order-one t-scalars. A color video in the form of an order-four real array can be represented by a t-matrix of order-two t-scalars. Even a monochrome image in the form of an order-two array can be represented by a t-matrix of order-zero t-scalars.

However, we contend, to have an effective t-matrix representation of high-order data, the complex entries of a t-scalar need to be correlated. Otherwise, the t-scalar multiplication based on circular convolution is pointless.

Hence, a convenient application arena of the t-matrix paradigm is for analyzing visual information including but not limited to images, videos, and sequential data such as time series, where for a raw data sample, there are always spatially-correlated neighborhoods available for exploitation.

5.2. T-matrix Representation of Legacy Visual Information. To reuse the legacy data representation and, on the other hand, exploit the potential of the t-matrix paradigm, one needs a consistent neighborhood strategy for t-matricizing visual information.

Figure 5.1 demonstrates a “ 3×3 -neighborhood” strategy for t-matricizing a small grey image of 16 pixels, i.e., 16 real numbers, in the form of a 4×4 real array. The t-matricization yields a t-matrix in $C^{4 \times 4}$, i.e., an order-four array in $C^{3 \times 3 \times 4 \times 4}$.

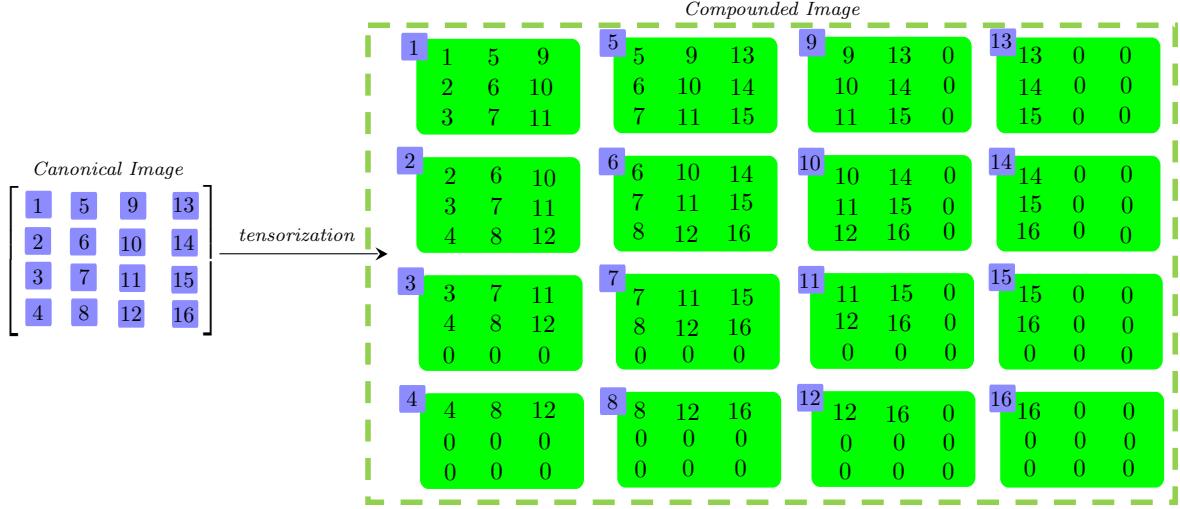


FIGURE 5.1. A t-matricization strategy of a 2D legacy image using 3×3 ‘inception’ neighborhoods —— from an input order-two array in $\mathbb{C}^{4 \times 4}$ (i.e., canonical matrix) to an order-four array in $C^{4 \times 4} \equiv \mathbb{C}^{3 \times 3 \times 4 \times 4}$ (i.e., a t-matrix).

Each small blue box corresponds to a scalar entry of the 4×4 input matrix in Figure 5.1. There are many distinct 3×3 neighborhoods available for each scalar of the input matrix. For example, one can either have a ‘central’ neighborhood set $\{1, 2, 3, 5, 6, 7, 9, 10, 11\}$ for the scalar 6 or alternatively, a so-called ‘inception’ neighborhood set $\{6, 7, 8, 10, 11, 12, 14, 15, 16\}$.

Figure 5.1 adopts the so-called ‘inception’ neighborhood of each scalar for t-matricizing the input 4×4 matrix. If a scalar is located at the image border, one can pad with 0 when necessary to have a 3×3 neighborhood. Each ‘inception’ neighborhood in the form of a 3×3 green box is highlighted by the corresponding scalar represented by a small blue box at the top-left corner of each green box.

The feasibility of the ‘neighborhood’ strategy, as demonstrated in Figure 5.1, is under the condition that input data is spatially-constrained. Hence, the t-matrix paradigm with the demonstrated neighborhood strategy is suitable to analyzing images or other visual information.⁶

The neighborhood strategy can be reused to extend input data to higher-orders. Figure 5.2 demonstrates how to extend a canonical grey image (i.e., an order-two array

⁶If the given matrix is not spatially-constrained, the spatially-correlated neighborhood strategy makes no sense. It is possible to analyze spatially-correlated data with the t-matrix paradigm. However, one needs a different t-matricization strategy to extend legacy data to higher-order.

in $\mathbb{R}^{4 \times 4}$) to a low-order compounded image (i.e., an order-four array in $\mathbb{R}^{3 \times 3 \times 4 \times 4}$) and then to a higher-order compounded image (i.e., an order-six array in $\mathbb{R}^{3 \times 3 \times 3 \times 3 \times 4 \times 4}$).

The spatial neighborhood strategy enables general visual information analysis with the higher-order t-matrix paradigm.

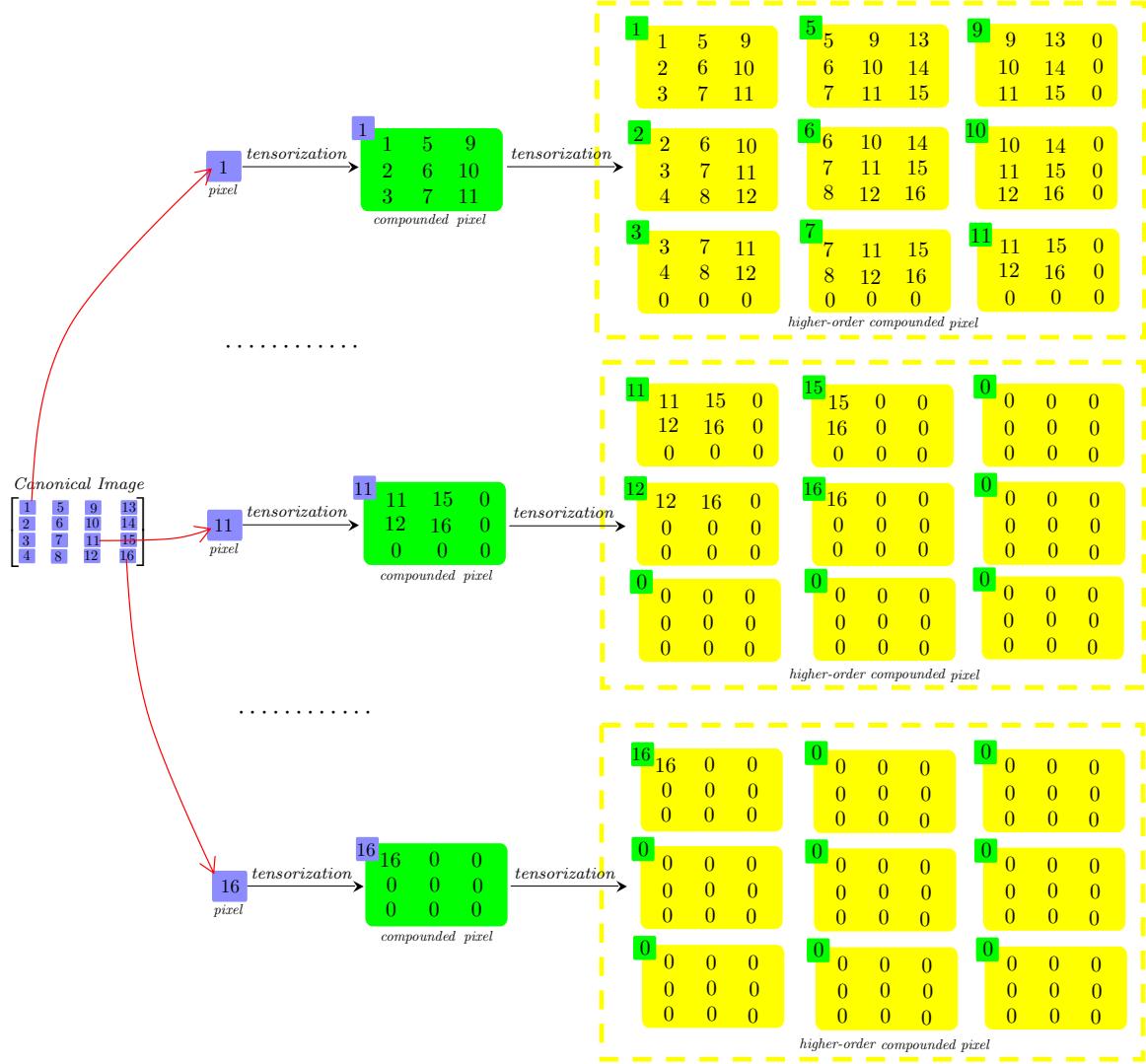


FIGURE 5.2. Reuse the neighborhood strategy to extend to a legacy grey image (i.e., an array in $\mathbb{R}^{4 \times 4}$) to a low-order compounded image (i.e., an array in $\mathbb{R}^{3 \times 3 \times 4 \times 4}$) and, then to a higher-order compounded image (i.e., an array in $\mathbb{R}^{3 \times 3 \times 3 \times 3 \times 4 \times 4}$).

5.3. Generalized Low-rank Approximation over C . With the t-matrix paradigm, many applications can be straightforwardly generalized. To this end, we discuss a high-order generalization of the Eckart-Young-Mirsky theorem, named after the authors of the theorem [7, 24].

Low-rank approximation plays an important role in modeling many applications in machine learning and data analytics. The problem is to find a low-rank optimal approximation to a given matrix. Specifically, given a matrix $X_{mat} \in \mathbb{C}^{M_1 \times M_2}$, one seeks an approximation matrix \hat{X}_{mat} to X_{TM} , satisfying

$$\begin{aligned} \|X_{mat} - \hat{X}_{mat}\|_F &= \min_{\text{rank}(Y_{mat}) \leq r} \|X_{mat} - Y_{mat}\|_F \\ \text{subject to } \text{rank}(\hat{X}_{mat}) &\leq r \leq \text{rank}(X_{mat}) . \end{aligned} \quad (5.1)$$

The optimal approximation is given by the Eckart-Young-Mirsky theorem via the SVD (Singular Value Decomposition) of X_{mat} . More specifically, given a complex matrix $X_{mat} \in \mathbb{C}^{M_1 \times M_2}$, let $M \doteq \min(M_1, M_2)$ and $X_{mat} = U_{mat} \cdot S_{mat} \cdot V_{mat}^*$ be the compact SVD of X_{mat} such that $S_{mat} \doteq \text{diag}(\lambda_1, \dots, \lambda_M)$ where $\lambda_1 \geq \dots \geq \lambda_M \geq 0$.

The Eckart-Young-Mirsky theorem gives the optimal approximation \hat{X}_{mat} via the SVD (Singular Value Decomposition) of X_{mat} as follows.

$$\hat{X}_{mat} = U_{mat} \cdot \hat{S}_{mat} \cdot V_{mat}^* \quad (5.2)$$

where $\hat{S}_{mat} \doteq \text{diag}(\lambda_1, \dots, \lambda_r, \underbrace{0, \dots, 0}_{(M-r) \text{ copies}})$.

With the canonical paradigm where a nonnegative integer defines rank, a solution to a higher-order generalization of equation (5.1) is NP-hard [14, 27, 2], and as a consequence, “naive approach to this problem is doomed to failure” [6].

However, with the t-matrix paradigm or Kilmer’s t-product model, a higher-order generalization of equation (5.1) with an analytical solution analogous to equation (5.2) is straightforward. The higher-order generalization with the t-matrix paradigm is as follows.

For a t-matrix $X_{TM} \in C^{M_1 \times M_2} \equiv \mathbb{C}^{I_1 \times \dots \times I_N \times M_1 \times M_2}$, the generalized optimization over C is to find a low-rank t-matrix $\hat{X}_{TM} \in C^{M_1 \times M_2}$ such that

$$\begin{aligned} r(X_{TM} - \hat{X}_{TM})_F &= \min_{\text{rank}(Y_{TM}) \leq H_T} r(X_{TM} - Y_{TM})_F \\ \text{subject to } \text{rank}(\hat{X}_{TM}) &\leq H_T \leq \text{rank}(X_{TM}) . \end{aligned} \quad (5.3)$$

Let $H_T = \sum_{k=1}^K r_k \cdot Q_{T,k} \geq Z_T$ be the orthogonality series of the nonnegative t-scalar H_T . Without loss of generality, let’s assume that $0 \leq r_1, \dots, r_K \leq M$ are all nonnegative integers where $M \doteq \min(M_1, M_2)$. In other words, there must exist a t-matrix in $C^{M_1 \times M_2}$ whose rank is equal to H_T .

The nonnegative t-scalar H_T can be uniquely represented by the sum of M idempotent t-scalars as follows

$$H_T = \sum_{m=1}^M \delta_{T,m} \quad (5.4)$$

where $\delta_{T,1}, \dots, \delta_{T,M} \in S^{idem}$ and $\delta_{T,1} \geq \dots \geq \delta_{T,M}$.

The sum in equation (5.4) is unique, and the m -th idempotent t-scalar $\delta_{T,m}$ is given as follows.

$$\delta_{T,m} = \sum_{k=1}^K \mathbf{1}_{m \leq r_k} \cdot Q_{T,k} \quad (5.5)$$

where $\mathbf{1}_{m \leq r_k}$ is the indicator function, which returns 1 when $m \leq r_k$ and otherwise, returns 0.

Let $X_{TM} = U_{TM} \circ S_{TM} \circ V_{TM}^*$ be the compact TSVD of X_{TM} where $S_{TM} = \text{diag}(\lambda_{T,1}, \dots, \lambda_{T,M})$ and $\lambda_{T,1} \geq \dots \geq \lambda_{T,M} \geq Z_T$. The analytical solution of equation (5.3) is given by

$$\hat{X}_{TM} = U_{TM} \circ \hat{S}_{TM} \circ V_{TM}^* \quad (5.6)$$

where $\hat{S}_{TM} \doteq \text{diag}(\lambda'_{T,1}, \dots, \lambda'_{T,M})$ and $\lambda'_{T,k} \doteq \lambda_{T,k} \circ \delta_{T,k}$ for all $k \in [M]$.

A simplified version of the above generalization is given when H_T is in the form $H_T = r \cdot E_T \equiv \sum_{k=1}^K r \cdot Q_{T,k}$ where $r \leq M$. Under this condition, equation (5.5) reduces to

$$\begin{cases} \delta_{T,1} = \dots = \delta_{T,r} = E_T \\ \delta_{T,r+1} = \dots = \delta_{T,M} = Z_T \end{cases} \quad (5.7)$$

Namely, the t-matrix \hat{S}_{TM} in equation (5.6) reduces to $\hat{S}_{TM} \doteq \text{diag}(\lambda_{T,1}, \dots, \lambda_{T,r}, \underbrace{Z_T, \dots, Z_T}_{(M-r) \text{ copies}})$.

In this case, the approximation $\hat{X}_{TM} \doteq U_{TM} \circ \hat{S}_{TM} \circ V_{TM}^* \in C^{M_1 \times M_2} \equiv \mathbb{C}^{I_1 \times \dots \times I_N \times M_1 \times M_2}$ is analogous to the canonical approximation \hat{X}_{mat} in equation (5.2) and is called the “truncated” TSVD approximation.

When $I_1 = \dots = I_N = 1$, equation (5.3) reduces to equation (5.1), and the generalized solution given by equation (5.6) reduces to the canonical solution given by equation (5.2).

In other words, equation (5.3) is a straightforward generalization of the analytical solution given by the Eckart-Young-Mirsky theorem.

5.4. Generalized Least-squares over C . By the semisimplicity of C , many canonical applications can be generalized using the t-matrix paradigm. These generalizations are completely compatible with their canonical counterparts.

For example, in [20, 25], Liao and Maybank et al. generalize the algorithms of HOSVD (Higher-Order Singular Value Decomposition), PCA (Principal Component Analysis),

2DPCA (Two Dimensional PCA), and Grassmannian Component Analysis over C , used for analyzing or classifying visual data.

To show the general principles and particularly the backward-compatibility of the t-matrix paradigm, we discuss the backward-compatible generalization of the well-known least-squares, which belongs to a special class of convex optimization. The principles demonstrated in the following discussion apply to generalize other canonical optimization applications, even those not convex.

The least-squares optimization over C is backward-compatible with the canonical least-squares optimization over \mathbb{C} and is formulated as follows

$$r(W_{TM} \circ \hat{\beta}_{TV} - A_{TV})_2 = \min_{\beta_{TV} \in C^M} r(W_{TM} \circ \beta_{TV} - A_{TV})_2 . \quad (5.8)$$

In the above equation, $r(\cdot)$ is the generalized norm defined by equation (4.22). The t-matrix $W_{TM} \in C^{D \times M}$ ($D \geq M$) and the t-vector $A_{TV} \in C^D$ are given in advance. The t-vector $\beta_{TV} \in C^M$ is optimizable, and the t-vector $\hat{\beta}_{TV} \in C^M$ is the optimal solution of β_{TV} .

The generalized least-squares has a geometric interpretation. Precisely, the column t-vectors of the t-matrix $W_{TM} \in C^{D \times M}$ form a generating set spanning a submodule $\mathcal{M} \subseteq C^D$ with a generalized dimension $\dim(\mathcal{M}) \doteq \text{rank}(W_{TM})$.

The projection A'_{TV} of the t-vector A_{TV} on the submodule \mathcal{M} is given by

$$\begin{aligned} A'_{TV} &\doteq W_{TM} \circ \hat{\beta}_{TV} \\ &= W_{TM} \circ (W_{TM}^* \circ W_{TM})^+ \circ W_{TM}^* \circ A_{TV} \\ &\doteq P_{TM} \circ A_{TV} \in \mathcal{M} \end{aligned} \quad (5.9)$$

where $P_{TM} \doteq W_{TM} \circ (W_{TM}^* \circ W_{TM})^+ \circ W_{TM}^* \in C^{D \times D}$ is called the projection t-matrix for the submodule \mathcal{M} .

The t-matrix P_{TM} is idempotent in the sense that

$$P_{TM} \circ P_{TM} = P_{TM} . \quad (5.10)$$

Also, the following equalities hold for all t-matrix $W_{TM} \in C^{D \times M}$,

$$\begin{aligned} \text{rank}(W_{TM}) &\equiv \text{rank}(P_{TM}) \\ W_{TM}^+ &\equiv (W_{TM}^* \circ W_{TM})^+ \circ W_{TM}^* . \end{aligned} \quad (5.11)$$

The generalized least-squares is equivalently defined by the generalized distance between A_{TV} and the submodule \mathcal{M} , i.e., the generalized distance between $A_{TV} \in C^D$ and $A'_{TV} \in \mathcal{M}$. More precisely,

$$r(A'_{TV} - A_{TV})_2 \equiv d(A'_{TV}, A_{TV}) \geq Z_T . \quad (5.12)$$

Note that the generalized least-squares $r(W_{TM} \circ \hat{\beta}_{TV} - A_{TV})$ is unique for all t-matrix $W_{TM} \in C^{D \times M}$ and all t-vector $A_{TV} \in C^D$. However, the t-vector $\hat{\beta}_{TV} \in C^M$ is not necessarily unique and is given as follows.

$$\hat{\beta}_{TV} \in \left\{ W_{TM}^+ \circ A_{TV} + (I_{TM} - W_{TM}^+ \circ W_{TM}) \circ \xi_{TV} \mid \xi_{TV} \in C^M \right\} \quad (5.13)$$

where $I_{TM} \doteq \text{diag}(\underbrace{E_T, \dots, E_T}_{M \text{ copies}})$ is the identity t-matrix.

The t-vector $\hat{\beta}_{TV} \in C^M$ is unique if and only if the column t-vectors of $W_{TM} \in C^{D \times M}$, where $D \geq M$, are independent over C , or in other words, the t-matrix W_{TM} is of full rank.

The condition that the column t-vectors of W_{TM} are independent over C is equivalent to one of the following conditions

- (i) $\text{rank}(W_{TM}) = M \cdot E_T$ where $M \doteq \min(M_1, M_2)$
- (ii) $W_{TM}^+ \circ W_{TM} = I_{TM} \doteq \text{diag}(\underbrace{E_T, \dots, E_T}_{M \text{ copies}})$

When the column t-vectors of W_{TM} are independent over C , the minimizer $\hat{\beta}_{TV} \in C^M$ is unique and is given by

$$\begin{aligned} \hat{\beta}_{TV} &\doteq \underset{\beta_{TV} \in C^M}{\text{argmin}} r(W_{TM} \circ \beta_{TV} - A_{TV})_2 \\ &= W_{TM}^+ \circ A_{TV} \\ &\equiv (W_{TM}^* \circ W_{TM})^+ \circ W_{TM}^* \circ A_{TV} \\ &\equiv (W_{TM}^* \circ W_{TM})^{-1} \circ W_{TM}^* \circ A_{TV} \end{aligned} \quad (5.15)$$

where $(W_{TM}^* \circ W_{TM})^{-1}$, called the inverse of the t-matrix $W_{TM}^* \circ W_{TM} \in C^{M \times M}$, is a special case of the pseudoinverse $(W_{TM}^* \circ W_{TM})^+$ when the t-matrix $(W_{TM}^* \circ W_{TM}) \in C^{M \times M}$ is of full rank.

When $I_1 = \dots = I_N = 1$, the generalized least-squares over C reduces to the canonical least-squares over \mathbb{C} .

5.5. Generalized Principal Component Analysis over C . Using generalized least-squares over C , one can generalize the well-known method of PCA (Principal Component Analysis). The generalized PCA is called TPCA (Tensorial PCA).

Precisely, given N t-vectors $X_{TV,1}, \dots, X_{TV,N} \in C^D$, the generalized component analysis of these t-vectors is to find a finite number of principal components $U_{TV,1}, \dots, U_{TV,Q} \in$

C^D such that $U_{TM}^* \circ U_{TM} = I_{TM} \doteq \text{diag}(E_T, \dots, E_T) \in C^{Q \times Q}$ where the k -th column t-vector of $U_{TM} \in C^{D \times Q}$, denoted by $(U_{TM})_{:,k}$, is the principal component $U_{TV,k}$, namely, $(U_{TM})_{:,k} \doteq U_{TV,k}, \forall k \in [Q]$.

The principal components $U_{TV,1}, \dots, U_{TV,Q} \in C^D$ capture the dominant information of the t-vectors $X_{TV,1}, \dots, X_{TV,N} \in C^D$ such that the first principal component $U_{TV,1}$ is given by

$$\begin{aligned} U_{TV,1} &\doteq \underset{r(Y_{TV})_2 = E_T}{\text{argmax}} \left\{ \sum_{k=1}^N |Y_{TV}^* \circ (X_{TV,k} - \bar{X}_{TV})|^2 \right\} \\ &\equiv \underset{r(Y_{TV})_2 = E_T}{\text{argmax}} r^2(Y_{TV}^* \circ W_{TM})_F \\ &\equiv \underset{r(Y_{TV})_2 = E_T}{\text{argmax}} Y_{TV}^* \circ W_{TM} \circ W_{TM}^* \circ Y_{TV} \end{aligned} \quad (5.16)$$

where

$$\bar{X}_{TV} \doteq (1/N) \cdot \sum_{k=1}^N X_{TV,k} \quad (5.17)$$

is the mean of the t-vectors $X_{TV,1}, \dots, X_{TV,N}$ and $W_{TM} \in C^{D \times N}$ denotes the t-matrix whose k -th column $(W_{TM})_{:,k} \in C^D$ is given by

$$(W_{TM})_{:,k} \doteq X_{TV,k} - \bar{X}_{TV}, \quad \forall k \in [N]. \quad (5.18)$$

Note that the condition $r(Y_{TV})_2 = E_T$ is equivalent to $Y_{TV}^* \circ Y_{TV} = E_T$. Equation (5.16) is to find the stationary point(s) of the following formulation with a generalized Lagrange multiplier $\lambda_T \in C$,

$$\mathcal{L}(Y_{TV}) = Y_{TV}^* \circ W_{TM} \circ W_{TM}^* \circ Y_{TV} - \lambda_T \circ (Y_{TV}^* \circ Y_{TV} - E_T). \quad (5.19)$$

The stationary point(s) of equation (5.19) can be determined by its derivative over C , equal to Z_T .⁷

$$\begin{aligned} \frac{\partial \mathcal{L}(Y_{TV})}{\partial Y_{TV}} &= 2 \cdot \left(Y_{TV}^* \circ W_{TM} \circ W_{TM}^* - \lambda_T \circ Y_{TV}^* \right) = Z_T \\ \Rightarrow W_{TM} \circ W_{TM}^* \circ Y_{TV} &= Y_{TV} \circ \lambda_T \Rightarrow Y_{TV}^* \circ W_{TM} \circ W_{TM}^* \circ Y_{TV} = \lambda_T. \end{aligned} \quad (5.20)$$

It shows that $U_{TV,1}$ is the generalized eigenvector in $C^D \equiv \mathbb{C}^{I_1 \times \dots \times I_N \times D}$ with the generalized maximum eigenvalue $\lambda_T \in S^{\text{nonneg}}$ of the Hermitian t-matrix $W_{TM} \circ W_{TM}^* \in C^{D \times D} \equiv \mathbb{C}^{I_1 \times \dots \times I_N \times D \times D}$.

The t-vector $U_{TV,1}$ is also the dominant singular t-vector with the generalized maximum singular value, i.e., a nonnegative t-scalar, of the t-matrix $W_{TM} \in C^{D \times N} \equiv \mathbb{C}^{I_1 \times \dots \times I_N \times D \times N}$.

We provide an unrigorous interpretation of the derivative as in equation (5.20) —— given a mapping $\mathcal{L} : Y_{TV} \mapsto \mathcal{L}(Y_{TV})$, the dependable $\mathcal{L}(Y_{TV}) \in C$ can be written as

⁷A rigorous investigation of differentiation over C is beyond the scope of this article.

follows

$$\begin{cases} \mathcal{L}(Y_{TV}) = \sum_{k=1}^K \mathcal{L}_k(Y_{vec,k}) \cdot Q_{T,k} \\ Y_{TV} = \sum_{k=1}^K Y_{vec,k} \bowtie Q_{T,k} \end{cases} \quad (5.21)$$

where \mathcal{L}_k is the k -th sub-mapping of \mathcal{L} for each $k \in [K]$.

Then, the derivative of $\mathcal{L}(Y_{TV})$ with respect to Y_{TV} is given by

$$\frac{\partial \mathcal{L}(Y_{TV})}{\partial Y_{TV}} \doteq \sum_{k=1}^K \frac{\partial \mathcal{L}_k(Y_{vec,k})}{\partial Y_{vec,k}} \cdot Q_{T,k} \in C \quad (5.22)$$

where the derivative of the left side of the equation is the generalized derivative on C , and the derivatives of the right side denote the canonical derivatives on complex numbers.

In a simple case as equation (5.20), the sub-mappings $\mathcal{L}_1, \dots, \mathcal{L}_K$ are the identical real-valued vector functions given by

$$\mathcal{L}_k : Y_{vec} \mapsto Y_{vec}^* \cdot W_{mat} \cdot W_{mat}^* \cdot Y_{vec} - \lambda \cdot (Y_{vec}^* \cdot Y_{vec} - 1) , \forall k \in [K] . \quad (5.23)$$

It shows that equation (5.20) is a result given by equation (5.22).

When the first q principal t-vectors $U_{TV,1}, \dots, U_{TV,q}$ are obtained, one uses the following equation to project the t-matrix W_{TM} on the orthogonal complement submodule of the submodule spanned by the principal t-vectors $U_{TV,1}, \dots, U_{TV,q} \in C^D$. More precisely,

$$W_{TM,(q+1)} = \left(I_{TM} - \sum_{i=1}^q U_{TV,i} \circ U_{TV,i}^* \right) \circ W_{TM} \in C^{D \times N} . \quad (5.24)$$

Let the t-matrices $P_{TM,q}, P_{TM,q}^\perp \in C^{D \times D}$ be given by

$$\begin{cases} P_{TM,q} \doteq \sum_{i=1}^q U_{TV,i} \circ U_{TV,i}^* \in C^{D \times D} \\ P_{TM,q}^\perp \doteq I_{TM} - P_{TM,q} \in C^{D \times D} \end{cases} . \quad (5.25)$$

Then, the following equalities hold for all $q \in [Q]$,

$$\begin{aligned} P_{TM,q} \circ P_{TM,q} &= P_{TM,q} = P_{TM,q}^* \\ P_{TM,q}^\perp \circ P_{TM,q}^\perp &= P_{TM,q}^\perp = (P_{TM,q}^\perp)^* \\ \text{rank}(P_{TM,q}^\perp) + \text{rank}(P_{TM,q}) &= D \cdot E_T \\ P_{TM,q} \circ P_{TM,q}^\perp &= P_{TM,q}^\perp \circ P_{TM,q} = Z_T \end{aligned} . \quad (5.26)$$

When the t-matrix $W_{TM, (q+1)} \in C^{D \times N}$ is obtained as in equation (5.24), the $(q+1)$ -th principal t-vector $U_{TV, (q+1)} \in C^D$ is given by

$$\begin{aligned} U_{TV, (q+1)} &\doteq \operatorname{argmax}_{r(Y_{TV})_2=E_T} r^2(Y_{TV}^* \circ W_{TM, (q+1)})_F \\ &\equiv \operatorname{argmax}_{r(Y_{TV})_2=E_T} Y_{TV}^* \circ W_{TM, (q+1)} \circ W_{TM, (q+1)}^* \circ Y_{TV}, \forall q \in [Q]. \end{aligned} \quad (5.27)$$

The t-vector $U_{TV, (q+1)} \in C^D$ is the $(q+1)$ -th dominant generalized eigenvector of $W_{TM} \circ W_{TM}^* \in C^{D \times D}$, also the $(q+1)$ -th dominant singular t-vector of $W_{TM} \in C^{D \times N}$. The maximum t-scalar $\max_{r(Y_{TV})_2=E_T} r(Y_{TV}^* \circ W_{TM, (q+1)})_F \in S^{nonneg}$ is the $(q+1)$ -th dominant singular value, i.e., a t-scalar, of $W_{TM} \in C^{D \times M}$.

Note that the t-vectors $(X_{TV, 1} - \bar{X}_{TV}), \dots, (X_{TV, N} - \bar{X}_{TV})$ are not independent on the module C^D . This leads to

$$\operatorname{rank}(W_{TM}) \leq Q_m \cdot E_T \text{ where } Q_m \doteq \min(D, N-1). \quad (5.28)$$

Let the compact TSVD of the t-matrix $W_{TM} \in C^{D \times N}$ be $W_{TM} = U_{TM} \circ S_{TM} \circ V_{TM}^*$ such that $U_{TM} \in C^{D \times Q_m}$, $V_{TM} \in C^{N \times Q_m}$, $U_{TM}^* \circ U_{TM} = V_{TM}^* \circ V_{TM} = I_{TM} \in C^{Q_m \times Q_m}$ and $S_{TM} \doteq \operatorname{diag}(\lambda_{T, 1}, \dots, \lambda_{T, Q_m})$ where $\lambda_{T, 1} \geq \dots \geq \lambda_{T, Q_m} \geq Z_T$.

If and only if $\operatorname{rank}(W_{TM}) = Q_m \cdot E_T$, the generalized singular values $\lambda_{T, 1}, \dots, \lambda_{T, Q_m}$ are all multiplicatively invertible, and all Q_m column t-vectors of U_{TM} are the principal t-vectors given by equations (5.16) and (5.27).

We call the above-mentioned generalized principal component analysis TPCA (Tensorial PCA). TPCA is backward-compatible with its canonical counterpart PCA. Analogous to its canonical counterpart PCA, TPCA applies to reduce the dimension of data.

More precisely, N t-vectors $X_{TV, 1}, \dots, X_{TV, N} \in C^D$ are given in advance and let the nonnegative t-scalar H_T be a generalized dimension subject to

$$H_T \leq \operatorname{rank}(W_{TM}) \leq Q_m \cdot E_T. \quad (5.29)$$

Without loss of generality, the generalized dimension $H_T \geq Z_T$ is uniquely represented in the form of equation (5.4), namely,

$$H_T = \sum_{q=1}^{Q_m} \delta_{T, q} \quad (5.30)$$

where $\delta_{T, q} \in S^{idem}$ for each $q \in [Q_m]$ and $\delta_{T, 1} \geq \dots \geq \delta_{T, Q_m} \geq Z_T$.

A query sample $Y_{TV}^{raw} \in C^D$ is then reduced to a t-vector $Y_{TV}^{TPCA} \in C^{Q_m}$ as follows

$$Y_{TV}^{TPCA} \doteq \hat{U}_{TM}^* \circ (Y_{TV}^{raw} - \bar{X}_{TV}) \quad (5.31)$$

where

$$\hat{U}_{TM} \doteq U_{TM} \circ \operatorname{diag}(\delta_{T, 1}, \dots, \delta_{T, Q_m}) \in C^{D \times Q_m}. \quad (5.32)$$

The projection Y_{TV}^{proj} of the t-vector $(Y_{TV}^{raw} - \bar{X}_{TV}) \in C^D$ on the submodule \mathcal{M} spanned by the columns of the t-matrix $\hat{U}_{TM} \in C^{D \times Q_m}$ is given by the following generalized least-squares problem

$$\begin{aligned} Y_{TV}^{proj} &\doteq \operatorname{argmin}_{Y_{TV} \in \mathcal{M}} r(Y_{TV}^{raw} - \bar{X}_{TV} - Y_{TV})_2 \\ &= U_{TM} \circ Y_{TV}^{TPCA} \\ &\equiv \hat{U}_{TM} \circ Y_{TV}^{TPCA} \\ &\equiv \hat{U}_{TM} \circ \hat{U}_{TM}^* \circ (Y_{TV}^{raw} - \bar{X}_{TV}) \in C^D . \end{aligned} \quad (5.33)$$

It follows that the t-matrix $P_{TM} \doteq \hat{U}_{TM} \circ \hat{U}_{TM}^* \in C^{D \times D}$ is idempotent, hermitian, and low-rank over C , namely $P_{TM} \circ P_{TM} = P_{TM} = P_{mat}^*$ and $\operatorname{rank}(P_{TM}) = H_T$.

The reconstruction Y_{TV}^{recon} of the query t-vector Y_{TV}^{raw} is then given by

$$Y_{TV}^{recon} = Y_{TV}^{proj} + \bar{X}_{TV} = P_{TM} \circ Y_{TV}^{raw} + (I_{TM} - P_{TM}) \circ \bar{X}_{TV} . \quad (5.34)$$

When the given generalized dimension is in the form $H_T = Q \cdot E_T$, where Q is a positive integer, the t-matrix \hat{U}_{TM} reduces to

$$\hat{U}_{TM} \doteq U_{TM} \circ \operatorname{diag}(\underbrace{E_T, \dots, E_T}_{Q \text{ copies}}, \underbrace{Z_T, \dots, Z_T}_{(Q_m-Q) \text{ copies}}) \in C^{D \times Q_m} \quad (5.35)$$

or, equivalently,

$$\hat{U}_{TM} = (U_{TM})_{:,1:Q} \in C^{D \times Q} \quad (5.36)$$

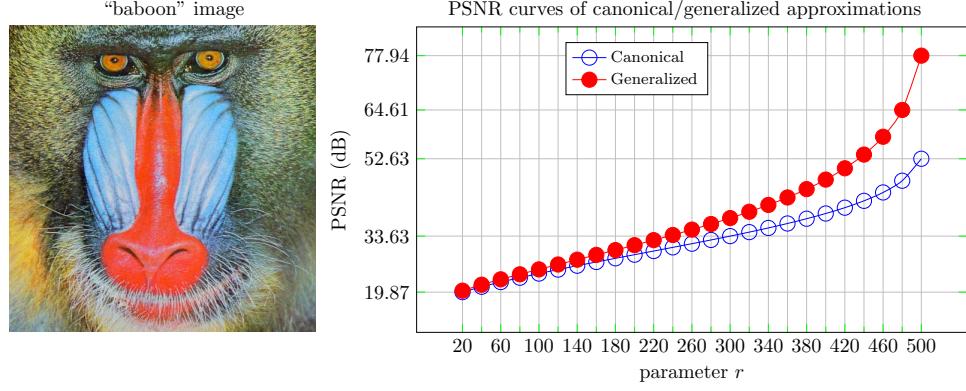
where \hat{U}_{TM} is the sub-structure containing the first Q columns of $U_{TM} \in C^{D \times Q_m}$.

In this situation, the last $(Q_m - Q)$ t-scalar entries of $Y_{TV}^{TPCA} \in C^{Q_m}$ are discarded. When $I_1 = \dots = I_N = 1$, TPCA reduces to its canonical counterpart PCA.

6. EXPERIMENTAL VERIFICATIONS

In this section, we demonstrate the t-matrix paradigm for general visual information analytics. We give some experiment results via t-matrices with quantitative comparison to their canonical counterparts.

6.1. Generalized Low-rank Approxiamtion. In the first experiment of low-rank approximation, we compare the approximation results of SVD and TSVD on the publicly available images.



PSNRs (dB) of canonical/generalized approximations with different rank parameter r .

Canonical = two-way array approximation via SVD for flattened RGB image;

Generalized* = three-way array approximation via TSVD

APPROXIMATION TYPE	parameter r												
	20	60	100	140	180	220	260	300	340	380	420	460	500
Canonical	19.87	22.35	24.43	26.37	28.18	29.97	31.77	33.63	35.64	37.90	40.61	44.35	52.63
Generalized	20.23	23.02	25.46	27.84	30.20	32.64	35.23	38.06	41.29	45.17	50.28	58.03	77.94

FIGURE 6.1. A comparison of low-rank approximation by SVD and TSVD on the “baboon” image

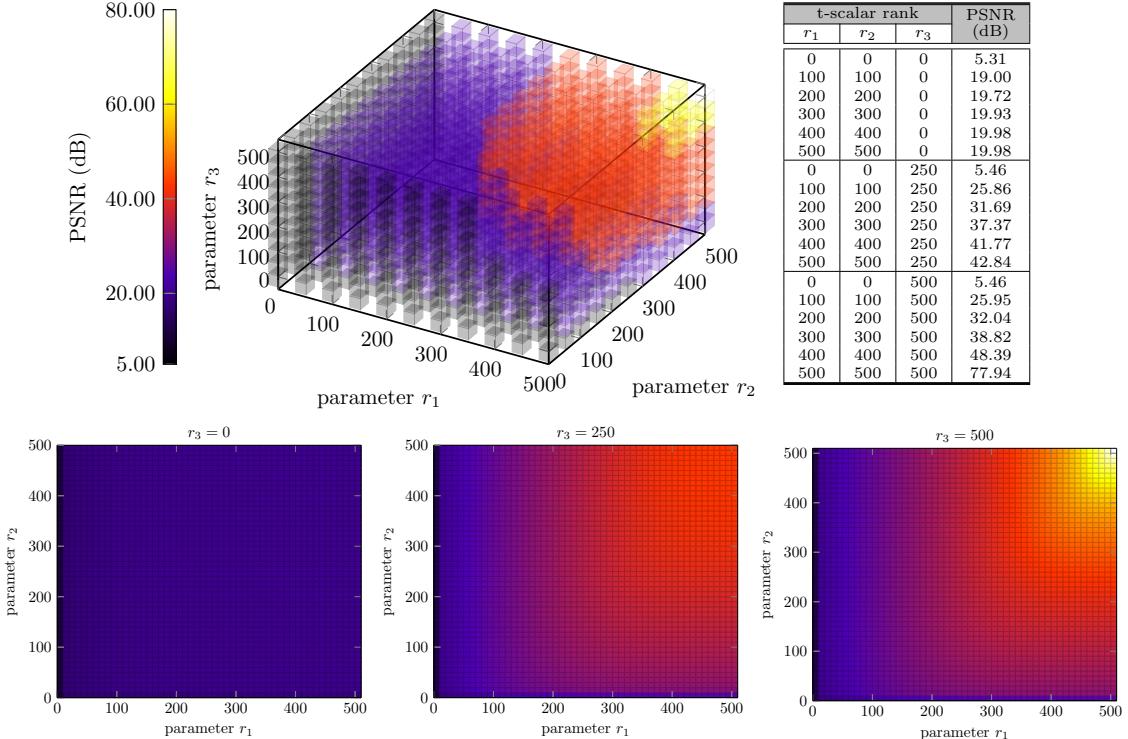


FIGURE 6.2. PSNR (dB) heat maps of TSVD approximations with different t-matrix ranks, characterized by the tuple r_1, r_2, r_3 on the RGB “baboon” image

6.1.1. “*Baboon*” image. The RGB image used in the first experiment in the $512 \times 512 \times 3$ is the “baboon” image.

Since TSVD applies to order-three arrays while SVD is only applicable to order-two arrays, to use SVD, the “baboon” image is flattened to a 512×1536 matrix by concatenating each 512×512 sub-image along the horizontal direction.

TSVD applies to order-three arrays of complex numbers, while SVD is only applicable to order-two arrays of complex numbers. Using SVD, the “baboon” image is flattened to a 512×1536 matrix by concatenating each 512×512 sub-image along the horizontal direction. Using TSVD, the “baboon” image is represented by a t-matrix in $C^{512 \times 512} \equiv \mathbb{C}^{3 \times 512 \times 512}$. The only requirement of transforming the RGB image to the underlying $3 \times 512 \times 512$ array of the t-matrix is a permutation of the indices of the raw $512 \times 512 \times 3$ array.

To make a reasonable comparison with the SVD approximation, we use a simplified TSVD approximation under the constraint $\text{rank}(\hat{X}_{TM}) \leq r \cdot E_T$ where $r \in \{0, \dots, 512\}$.

To give a quantitative comparison, when an approximation array \hat{X} of a given array X is obtained, the PSNR (Peak Signal Noise Ratio) of the approximation is given as follows.

$$\text{PSNR} = 20 \cdot \log_{10} \left(\sqrt{N^{\text{entry}}} \cdot \text{MAX} \cdot \|X - \hat{X}\|_F^{-1} \right) \quad (6.1)$$

where N^{entry} denotes the number of scalar entries of the array X , and MAX represents the maximum possible real value in X .

In the experiment of using the “baboon” image, $N^{\text{entry}} = 786432 \equiv 512 \times 512 \times 3$, $\text{MAX} = 255$. Figure 6.1 gives the PSNRs of canonical approximation (via SVD) and generalized approximation (via TSVD) with different rank parameters. It is easy to follow that TSVD consistently outperforms SVD. When $r = 500$, the PSNRs of TSVD and SVD differ by more than 25 dB.

Also note, in an approximation problem as in equation (5.3), the rank parameter H_T can be any t-scalar $Z_T \leq H_T \leq \text{rank}(X_{TM})$ rather than just in the form of $H_T \doteq r \cdot E_T$ where $r \in \{0, \dots, 512\}$.

The peak signal-noise ratios (PSNRs in dB) of TSVD approximation using different general t-matrix ranks subject to $Z_T \leq H_T \leq \text{rank}(X_{TM})$ are given in Figure 6.2.

Since the generalized rank of the t-matrix $\hat{X}_{TM} \in C^{512 \times 512} \equiv \mathbb{C}^{3 \times 512 \times 512}$ can be written in the form of $\text{rank}(\hat{X}_{TM}) = \sum_{k=1}^3 r_k \cdot Q_{T,k}$, the generalized rank of \hat{X}_{TM} is equivalently characterized by the tuple (r_1, r_2, r_3) .

A 3D heat map and three 2D heat maps with different r_1, r_2 and r_3 are shown in the figure. It meets the expectation that a better approximation is obtained with higher values r_1, r_2 and r_3 , or equivalently, a higher t-matrix rank of $\hat{X}_{TM} \in C^{512 \times 512}$.

6.1.2. “*Cameraman*” image. For a second low-rank approximation experiment, the “cameraman” image is used. The size of the grey image 256×256 . This image is easy to be approximated via SVD.

To exploit the potential of TSVD, we use the 3×3 “inception” neighborhood strategy (see Figure 5.1) to t-matricize the “cameraman” image to a t-matrix in $C^{256 \times 256} \equiv \mathbb{C}^{3 \times 3 \times 256 \times 256}$.

In the TSVD approximation experiment, the rank condition is given by $\text{rank}(\hat{X}_{TM}) \leq r \cdot E_T$, where r is a nonnegative integer. Namely, the “truncated” TSVD approximation is used.

Note that the approximation by TSVD is an array in $\mathbb{C}^{3 \times 3 \times 256 \times 256}$, while the approximation by SVD is an array in $\mathbb{C}^{256 \times 256}$. To give a more relevant comparison on PSNRs, we extract the “inception” slice of the TSVD approximation to compare with the approximation by SVD. The “inception” slice is the 256×256 matrix by only keeping the first scalar entry of each t-scalar in a t-matrix.

To compute PSNRs for this experiment, the parameters in equation (6.1) are $MAX = 255$ and $N^{entry} = 65536 \equiv 256 \times 256$.

Figure 6.3 shows that the inception slice approximations via TSVD consistently outperform the SVD approximation in PSNRs. When the parameter $r = 250$, the approximation by the inception slice via TSVD with order-two t-scalars outperforms the canonical approximation via SVD by more than 28 dB.

By reusing the neighborhood t-matricization solution demonstrated in Figure 5.2, the order of an obtained t-matrix is increased. When using order-four t-scalars, Figure 6.3 shows that an additional gain of more than 59 dB is obtained, reaching 179.08 dB.

The approximation results by TSVD shown in Figure 6.3 are computed via the inception slice of a TSVD approximation. One might be interested in the PSNR of a whole TSVD approximation rather than its slice.

To this end, another type of PSNRs is computed. More precisely, given a higher-order array $X_{TM} \doteq X \in C^{256 \times 256}$ and its approximation $\hat{X}_{TM} \doteq \hat{X} \in C^{256 \times 256}$, its PSNR is computed as in equation (6.1) with $MAX = 255$ and $N^{entry} = 256 \times 256 \times K \doteq 65536 \times K$, where $K \doteq I_1 \times \cdots \times I_N$ denotes the number of scalar-entries in a t-scalar.

Using the neighborhood strategy of data t-matricization (see Figures 5.1 and 5.2), two distinct t-scalar sizes are adopted in this experiment. More precisely, the order-two t-scalars are elements of $C \equiv \mathbb{C}^{3 \times 3}$, i.e., $K = 9$. The order-four t-scalars are elements of $C \equiv \mathbb{C}^{3 \times 3 \times 3 \times 3}$, i.e., $K = 81$.

Two distinct t-scalar sizes are adopted in approximating the “cameraman” image. More precisely, order-two t-scalars are elements of $C \equiv \mathbb{C}^{3 \times 3}$, i.e., $K = 9$, and order-four t-scalars are elements of $C \equiv \mathbb{C}^{3 \times 3 \times 3 \times 3}$, i.e., $K = 81$.

Figure 6.4 shows the PSNR curves of high-order approximations with respectively the order-two t-scalars ($K = 9$) and the order-four t-scalars ($K = 81$) and the PSNR curve obtained by SVD ($K = 1$) where the parameter $N^{\text{entry}} \doteq 65536 \times K$ is given in equation (6.1). The PSNR comparison shown in Figure 6.4 corroborates the outperformance of TSVD over SVD, shown in Figure 6.3.

6.2. Generalized Least-squares. To evaluate canonical least squares and generalized least squares, we compare their performances on approximating images. The experimental images are from the publically available ORL dataset, which contains 400 facial images from 40 subjects.⁸ Each of these images is monochrome and has 112×92 pixels. We choose the images from two classes for the experiment.

Figure 6.5 shows the chosen images from each class — for each class, the first three images of each class are chosen. These images are further t-matricized to higher-order arrays, i.e., t-matrices.

For each class, the experiment uses the last two of the chosen images, i.e., t-matrices, to approximate the first one via the generalized least-squares.

Let the three images in the form of t-matrices be $A_{TM}, B_{TM}, C_{TM} \in \mathbb{C}^{112 \times 92}$. The goal is to use a generalized linear combination of $\lambda_T \circ B_{TM} + \xi_T \circ C_{TM}$, where $\lambda_T, \xi_T \in C$, to approximate A_{TM} .

The following equation gives the optimal approximation $A_{TM}^{\text{opt}} \doteq \lambda_T^{\text{opt}} \circ B_{TM} + \xi_T^{\text{opt}} \circ C_{TM}$ of the t-matrix A_{TM} by

$$\begin{aligned} r(A_{TM}^{\text{opt}} - A_{TM})_F &\doteq r(\lambda_T^{\text{opt}} \circ B_{TM} + \xi_T^{\text{opt}} \circ C_{TM} - A_{TM})_F \\ &= \min_{\lambda_T, \xi_T \in C} r(\lambda_T \circ B_{TM} + \xi_T \circ C_{TM} - A_{TM})_F \geq Z_T . \end{aligned} \quad (6.2)$$

The problem in equation (6.2) can be recast to a generalized least-squares problem in equation (5.8) to obtain the t-matrix $A_{TM}^{\text{opt}} \in \mathbb{C}^{112 \times 92}$.

When the approximation t-matrix $A_{TM}^{\text{opt}} \in \mathbb{C}^{112 \times 92} \equiv \mathbb{C}^{I_1 \times \dots \times I_N \times 112 \times 92}$ is obtained, the PSNR of the approximation is computed with $\text{MAX} \doteq 255$, $N^{\text{entry}} \doteq 112 \times 92 \times K$ where $K = I_1 \times \dots \times I_N$.

To have a comprehensive comparison, we have the experiment images t-matricized using or reusing 3×3 neighborhood strategy to arrays of order-four, order-six, order-eight, and order-ten.

⁸<https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>.

To have a comprehensive comparison, we have the experiment images t-matricized using or reusing 3×3 neighborhood strategy to arrays of order-four (where $N = 2, I_1 = I_2 = 3$), order-six (where $N = 4, I_1 = \dots = I_4 = 3$), order-eight (where $N = 6, I_1 = \dots = I_6 = 3$), and order-ten (where $N = 8, I_1 = \dots = I_8 = 3$).

Besides the generalized least-squares, the canonical least-squares also applies to approximate higher-order images in the form of t-matrices.

More precisely, given higher-order images $A_{TM}, B_{TM}, C_{TM} \in \mathbb{C}^{I_1 \times \dots \times I_N \times M_1 \times M_2}$, the canonical least-squares use the linear combination $\alpha \cdot B_{TM} + \beta \cdot C_{TM}$ to approximate A_{TM} , where α and β are complex numbers.

The following equation gives the optimal approximation $A_{TM}^{opt} \doteq \alpha^{opt} \cdot B_{TM} + \beta^{opt} \cdot C_{TM}$ of the t-matrix A_{TM} by

$$\begin{aligned} \|A_{TM}^{opt} - A_{TM}\|_F &\doteq \|\alpha^{opt} \cdot B_{TM} + \beta^{opt} \cdot C_{TM} - A_{TM}\|_F \\ &= \min_{\alpha, \beta \in \mathbb{C}} \|\alpha \cdot B_{TM} + \beta \cdot C_{TM} - A_{TM}\|_F \geq 0 . \end{aligned} \quad (6.3)$$

Figure 6.5 gives the PSNRs by the canonical least-squares and generalized least-squares. The PSNRs by the canonical least-squares on the original ORL images is 17.26 dB and 20.12 dB. Higher-order images contribute higher PSNRs by the canonical least-squares. However, the generalized least-squares outperform the canonical least-squares on higher-order images, yielding higher quality of approximation with higher PSNRs. The highest PSNRs are yielded by the generalized least-squares, namely 18.07 dB and 20.60 dB on the chosen experiment images of each class.

6.3. Generalized Principal Component Analysis. To show its performance, we use the generalized principal component analysis (TPCA) to extract features of the public CIFAR-10 image dataset.⁹

The CIFAR-10 dataset contains thousands of color images, each image a $32 \times 32 \times 3$ array. We choose the first 36 images of the first training set of the dataset for extracting principal t-vectors or vectors, namely $N = 36$.

6.3.1. TPCA. The first 25 images from the test set of the dataset are chosen as the query images. The subfigures in the first column of Figure 6.6 show the chosen training images and query images.

Each raw CIFAR-10 image is an order-three array in $\mathbb{C}^{32 \times 32 \times 3}$. With a permutation of entry index and then an array reshaping, each order-three array can be transformed into a t-vector in $C^{1024} \equiv \mathbb{C}^{3 \times 1024} \equiv \mathbb{C}^{3 \times 1 \times \dots \times 1 \times 1024 \times 1}$. Namely, $I_1 = 3, I_2 = \dots = I_N = 1, M_1 = 1024, M_2 = 1$, and $1024 = 32 \times 32$ in the form of $\mathbb{C}^{I_1 \times \dots \times I_N \times M_1 \times M_2}$.

⁹<http://www.cs.toronto.edu/~kriz/cifar.html>

In this image approximation experiment, TPCA works on the obtained t-vectors in $C^{1024} \equiv \mathbb{C}^{3 \times 1024}$, where all t-scalars are order-one arrays, each containing three scalar-entries, namely, $K = 3$. Then, given a generalized dimension parameter $H_T \in S^{\text{nonneg}}$, it can be represented by a 3-tuple of nonnegative integers r_1, r_2 , and r_3 in the following form

$$H_T = r_1 \cdot Q_{T,1} + r_2 \cdot Q_{T,2} + r_3 \cdot Q_{T,3} \quad (6.4)$$

where $r_1, r_2, r_3 \in \{0, \dots, Q_m\} \equiv \{0, \dots, 35\}$.

With the parameter H_T or equivalently, the tuple (r_1, r_2, r_3) , each of the 25 query t-vector is reconstructed as in equation (5.34). Each reconstructed image is then obtained by transforming its counterpart of the reconstructed t-vector into the original form, namely, a $32 \times 32 \times 3$ array.

After having all the 25 raw and reconstructed query images, we arrange them in two 3072×25 arrays. Precisely, each image is reshaped to a column of length 3072 where $3072 = 32 \times 32 \times 3$. Then, the PSNR is computed as in equation (6.1) with $N^{\text{entry}} = 76800 = 3072 \times 25 = (32 \times 32 \times 3) \times 25$ and $\text{MAX} = 255$.

The PSNRs of the TPCA reconstructions are given in Figure (6.6). Figure (6.6) shows that a larger value of H_T contributes to a higher PSNR of TPCA reconstruction. When $H_T = E_T$, or equivalently $r_1 = r_2 = r_3 = 1$, the PSNR of TPCA reconstruction is 13.17 dB. When $H_T \doteq Q_m \cdot E_T = 35 \cdot E_T$ (or equivalently, $r_1 = r_2 = r_3 = Q_m \doteq N - 1 = 35$), the PSNR is 16.81 dB.

6.3.2. PCA vs. TPCA. Note that both PCA and TPCA apply to extract principal components and reconstruct RGB images in the form of higher-order arrays.

One might be interested in comparing the performance of PCA and TPCA. In this part, we compare the results of TPCA and PCA on reconstructing the CIFAR-10 images.

The underlying $32 \times 32 \times 3$ array of each RGB image is reshaped to a 3072-dimensional vector. When using PCA. Hence, with the same samples introduced in Section 6.3.1, we have 36 training vectors and can extract 35 principal component vectors. The first r principal vectors are used to reconstruct each of the 25 query vectors, where $r \leq 35$.

To have a fair and reasonable comparison to PCA, TPCA works on the same CIFAR-10 images, as already reported in Section 6.3.1, but with the generalized dimension parameter H_T in equation (6.5) constrained with the so-called “truncated” condition $r_1 = r_2 = r_3 \equiv r$. More precisely, the parameter $H_T \geq Z_T$ is rewritten as follows

$$H_T \doteq r \cdot E_T \equiv r \cdot Q_{T,1} + r \cdot Q_{T,2} + r \cdot Q_{T,3} \quad (6.5)$$

where $r \in \{0, \dots, Q_m\} \equiv \{0, \dots, 35\}$.

Note that, in this experiment, only image reshape and scalar index permutation are adopted. No data t-matricization, i.e., the proposed neighborhood strategy, is employed. In other words, PCA and TPCA use the same raw images, only reorganized in different formats.

The PSNRs of the PCA and TPCA reconstructions are computed with the same settings, as described in Section 6.3.1. These PSNRs are both tabulated and are shown as the curves of parameter r in Figure 6.7. It shows that TPCA consistently outperforms PCA on the CIFAR-10 images in terms of PSNR.

6.3.3. TPCA with Higher-order T-scalars. One might also be interested in the effect of higher-order t-scalars on the performance of a generalized application. To this end, we adopt t-scalars of different orders with TPCA to conduct image approximations.

Each experiment RGB image is a $32 \times 32 \times 3$ array of real numbers and has three monochrome subimages in the form of a 32×32 array. Using the 3×3 neighborhood strategy (see Figure 5.1) on each monochrome subimage, an experiment RGB image is t-matricized from an order-three array in $\mathbb{C}^{32 \times 32 \times 3}$ to an order-five array $\mathbb{C}^{3 \times 3 \times 32 \times 32 \times 3}$.

With simple manipulations, an obtained order-five array can be transformed into at least two versions of the t-vector. The two versions are described as follows.

Version 1: An order-five array in $\mathbb{C}^{3 \times 3 \times 32 \times 32 \times 3}$ is reshaped to an order-three array in $\mathbb{C}^{3 \times 3 \times 3072}$, where the obtained array is algebraically interpreted as a t-vector in C^{3072} and $3072 = 32 \times 32 \times 3$, namely, in this scenario, $C \equiv \mathbb{C}^{3 \times 3} \doteq \mathbb{C}^{3^2}$.

Version 2: Alternatively, an order-five array in $\mathbb{C}^{3 \times 3 \times 32 \times 32 \times 3}$ can be first permuted to an array of the same order in $\mathbb{C}^{3 \times 3 \times 3 \times 32 \times 32}$, and then reshaped to an order-four array in $\mathbb{C}^{3 \times 3 \times 3 \times 1024}$ with $1024 = 32 \times 32$, which is algebraically interpreted as a t-vector in C^{1024} , namely, in this scenario, $C \equiv \mathbb{C}^{3 \times 3 \times 3} \doteq \mathbb{C}^{3^3}$.

TPCA adopts the two versions of t-vectors in the experiment of this subsection. TPCA, with the above first version of t-vectors, is referred to as TPCA-I with $C \equiv \mathbb{C}^{3 \times 3} \doteq \mathbb{C}^{3^2}$. The second version is referred to as TPCA-1 with $C \equiv \mathbb{C}^{3 \times 3 \times 3} \doteq \mathbb{C}^{3^3}$.

On the other hand, reusing the 3×3 neighborhood strategy, as shown in Figure 5.2, makes it easy to increase the order of adopted t-scalars. In the experiment, TPCA using t-vectors in C^{3072} and $C \equiv \mathbb{C}^{3 \times 3 \times 3 \times 3} \doteq \mathbb{C}^{3^4}$ is referred to as TPCA-II. TPCA using t-vectors in C^{3072} and $C \equiv \mathbb{C}^{3 \times 3 \times 3 \times 3 \times 3} \doteq \mathbb{C}^{3^6}$ is referred to as TPCA-III.

Similarly, TPCA using t-vectors in C^{1024} with $C \equiv \mathbb{C}^{3 \times 3 \times 3 \times 3 \times 3} \doteq \mathbb{C}^{3^5}$ is referred to as TPCA-2, and TPCA using t-vectors in C^{1024} with $C \equiv \mathbb{C}^{3 \times 3 \times 3 \times 3 \times 3 \times 3} \doteq \mathbb{C}^{3^7}$ is referred to as TPCA-3.

Thus, there are six variants of TPCA using t-scalars of different higher-orders in performance comparisons. For clarity, we summarize their t-vector settings t-vectors

TABLE

6.1. T-VECTOR SETTINGS FOR THE EXPERIMENT OF TPCA WITH T-SCALARS OF DIFFERENT ORDERS

t-vector of settings	methods							
	TPCA-I	TPCA-1	TPCA-II	TPCA-2	TPCA-III	TPCA-3	TPCA	PCA
shape of t-scalars	$\underbrace{3 \times \dots \times 3}_{2 \text{ copies}}$	$\underbrace{3 \times \dots \times 3}_{3 \text{ copies}}$	$\underbrace{3 \times \dots \times 3}_{4 \text{ copies}}$	$\underbrace{3 \times \dots \times 3}_{5 \text{ copies}}$	$\underbrace{3 \times \dots \times 3}_{6 \text{ copies}}$	$\underbrace{3 \times \dots \times 3}_{7 \text{ copies}}$	3	1
orders of t-scalars	order-two	order-three	order-four	order-five	order-six	order-seven	order-one	order-zero
number of t-scalar entries	3072 $= 32 \times 32 \times 3$	1024 $= 32 \times 32$	3072 $= 32 \times 32 \times 3$	1024 $= 32 \times 32$	3072 $= 32 \times 32 \times 3$	1024 $= 32 \times 32$	1024 $= 32 \times 32$	3072 $= 32 \times 32 \times 3$ (i.e., t-scalars reduces to canonical scalars)
N^{scalar} = number of scalar entries	27648 $= 3072 \times 3^2$	27648 $= 1024 \times 3^3$	248832 $= 3072 \times 3^4$	248832 $= 1024 \times 3^5$	2239488 $= 3072 \times 3^6$	2239488 $= 1024 \times 3^7$	3072 $= 1024 \times 3$	1024 $= 1024 \times 3$

in Table 6.1. As a bottom line for performance comparison, Table 6.1 also gives the settings of TPCA with low-order t-scalars and PCA.

Figure 6.8 shows the results with the dimension parameter r by algorithms PCA, TPCA, and other variants of TPCA on the 25 CIFAR-10 images appropriately t-matricized when necessary.

Among these algorithms, PCA, TPCA-I, TPCA-II, and TPCA-III respectively use order-zero, order-two, one-four, and order-six t-scalars entries, i.e., even-number-order t-scalars.¹⁰ The top-left subfigure of Figure 6.8 shows Their PSNR curves over the dimension parameter r . On the other hand, TPCA, TPCA-1, TPCA-2, and TPCA-3 respectively use order-one, order-three, order-five, and order-seven t-scalars entries, i.e., odd-number-order t-scalars. The top-right subfigure shows their PSNR curves.

Also, a few words for computing PSNRs, let N^{scalar} be the number of scalar entries of each vector or t-vector employed by a specific algorithm. Then, given 25 test images, or their t-matricized versions, cast to 25 vectors, or t-vectors, one can arrange them and their approximation versions to two arrays of N^{entry} scalars with $N^{entry} \doteq 25 \cdot N^{scalar}$ and use equation (6.1) to compute the PSNRs yielded by an algorithm.

Two observations are apparent from the first row of Figure 6.8. (i) A higher-dimensional parameter r always corresponds to a higher quality of reconstruction in terms of PSNR. (ii) Higher-order methods outperform their lower-order counterparts in terms of reconstruction quality with the same parameter r .

Note that both PCA and TPCA are applied to the same information, cast in two different formats. Similar scenarios also occur to the pairs of TPCA-I and TPCA-1, TPCA-II and TPCA-2, as well as TPCA-III and TPCA-3, where TPCA-1, TPCA-2, and TPCA-3 are respectively with higher-order t-scalars but a smaller number of t-scalar entries than their counterparts TPCA-I, TPCA-II, TPCA-III.

The first row of Figure 6.8 shows that, even on the same information, TPCA, TPCA-1, TPCA-2, and TPCA-3, using higher-order t-scalars, respectively outperform their

¹⁰ Scalars are a special case of t-scalars, namely order-zero t-scalars.

counterparts PCA, TPCA-I, TPCA-II, and TPCA-III, using lower-order t-scalars. For example, when $r = 35$, TPCA-3 outperforms TPCA-III by 1.36 dB, i.e., $1.36 \text{ dB} = 20.23 \text{ dB} - 18.87 \text{ dB}$.

Average Pooling. To adopt the generalized outputs in t-scalars, t-vectors, or t-matrices to canonical algorithms, one needs a mechanism to down-size generalized outputs over C to canonical results over complex numbers. Average pooling is such a down-sizing mechanism, introduced as follows.

Given a t-matrix $X_{TM} \in C^{M_1 \times M_2} \equiv \mathbb{C}^{I_1 \times \dots \times I_N \times M_1 \times M_2}$, one can use average pooling to down-size all t-scalar entries of X_{TM} to have a matrix $X_{mat} \in \mathbb{C}^{M_1 \times M_2}$ given by

$$(X_{mat})_{m_1, m_2} = (1/K) \cdot (X_{T, m_1, m_2})_{i_1, \dots, i_N} \in \mathbb{C} \quad (6.6)$$

where $K \doteq I_1 \times \dots \times I_N$, $(X_{T, m_1, m_2})_{i_1, \dots, i_N} \in C$ denotes the (m_1, m_2) -th t-scalar entry of X_{TM} , $(X_{mat})_{m_1, m_2} \in \mathbb{C}$ denotes the (m_1, m_2) -th complex entry of the matrix X_{mat} , for all $m_1 \in [M_1]$ and $m_2 \in [M_2]$.

Using the average pooling in the experiment, one reduces all TPCA variants over t-scalars to their canonical counterparts over complex numbers.

As a consequence, each t-scalar is reduced to a vector with $N^{scalar} = 3072$. When computing PSNRs in equation (6.1), the parameter N^{entry} is given by $N^{entry} = 25 \cdot N^{scalar} = 76800$.

The second row of Figure 6.8 shows the PSNR curves of all the TPCA variants with average pooling. These curves corroborate the observation found from the subfigures in the first row of Figure 6.8. Furthermore, it shows from the two rows of Figure 6.8 that a PSNR curve with average pooling is even higher than the associated PSNR curve without average pooling.

To give a panoramic comparison of different algorithms with different settings, Figure 6.8 gives a 2D heat-map of PSNRs in the last row, where TPCA-3, using order-seven t-scalars, has the highest PSNRs and PCA, using order-zero t-scalars, has the lowest PSNRs.

7. CONCLUSION

A semisimple paradigm of tensorial matrices over an algebra of generalized scalars is proposed for general data analytics with visual information analysis applications. The algebraic paradigm generalizes and is backward-compatible with the canonical paradigm, combining the higher-order merits of multi-way arrays and the low-order intuition of canonical complex matrices.

In the algebraic paradigm, scalars are extended to the so-called t-scalars, which are implemented as multi-way complex arrays of a fixed-size. Under the bestowed algebraic

operations, the set of t-scalars form a semisimple associative algebra, which is unital, commutative, and a novel $*$ -algebra. Due to its semisimplicity, the semisimple algebra can be decomposed to a finite number of irreducible algebras, isomorphic to the field of complex numbers, which is also a simple algebra.

With the backward-compatible simple paradigm, many canonical algorithms and applications over complex numbers can be straightforwardly extended over the new algebra as long as the scalar entries of each t-scalars are correlated. To this end, we propose a neighborhood strategy to extend legacy visual information data to the higher-order versions. In theory, the computational cost of a higher-order generalization is a linear function of the size of a t-scalar, i.e., the number of entries of a t-scalar. To verify the semisimple paradigm's effectiveness and its backward-compatibility, we choose to generate several classical algorithms and applications to their higher-order versions and apply them to analyze legacy images. Our experiments on these publicly available images show the semisimple paradigm, generalized algorithms, and applications compare favorably with their canonical counterparts. Our experiments show that higher-order generalizations also outperform their low-order counterparts.

ACKNOWLEDGMENTS

Liang Liao would like to thank associate professor Xiuwei Zhang of the Northwestern Polytechnical University, professor Pingjun Wei, and Xinqiang Wang of the Zhongyuan University of Technology for supporting this work. The students or formal students, including Yuemei Ren, a former Ph.D. student at the Northwestern Polytechnical University, Xuechun Zhang, Shen Lin, Chengkai Yang, Haichang Ye, and Jie Yang, all postgraduate students supervised by Liang Liao at the Zhongyuan University of Technology, help a lot for the experiments.

This research of Liang Liao, by the suggestion of Professor Stephen John Maybank (Fellow of the IEEE, Member of Academia Europaea), was first launched in 2015, when Liang Liao served as a researcher at the Department of Computer Science, Birkbeck, University of London. Liang Liao is invited back to Birbeck and collaborates with Professor Stephen John Maybank to continue this research at the University of London. Liang Liao thanks Professor Yudong Zhang of the University of Leicester, Dr. Changzheng Ma, Chief engineer of MooVita Ltd., Singapore, and Hailong Zhu of the Guangdong University of Technology for helpful online discussions when Liang Liao was in London. The Henan Provincial Department of Science and Technology partially supported Liao's visit to London financially.

Liang Liao also would like to thank the Birkbeck Institute of Data Analytics for free using the high-performance computing facilities, although London's lockdown somehow affects the progress of this research.

This work was partially supported by the National Natural Science Foundation of China under the grant number No. U1404607 and the High-end Foreign Experts Program, under the grant number No. GDW20186300351, of the State Administration of Foreign Experts Affairs.

CREDITS

Liang Liao and Stephen John Maybank contribute equally to the theory of t-scalars, t-matrices, the t-algebra, and the semisimple paradigm.

OPEN SOURCE

During this research, Liang Liao has completed a workable MATLAB library for numerical experiments and application prototyping. The library is pedagogy oriented with structure programming and follows the naming protocol of many MATLAB built-in functions. After the completion of this article, we are considering to open-source the version of the MATLAB library. We hope it can help the interested readers grasp the semisimple paradigm's philology and understand the mechanism and the backward-compatibility of the t-matrix theory.

The code repository is at <https://github.com/liaoliang2020/talgebra>.

An enhanced object-oriented version of this library and a version in a second programming language is also planned. However, at present, this research is only frugally founded, and the necessity of developing a new version significantly depends on the response for the current pedagogical version of this MATLAB library. If interested, please let us know and be free to contribute spiritually or financially to this project.

CONTACT INFORMATION

All prospective fundings, supports, collaboration interests, job offers, suggestions, critics, and other responses are welcome. Be free to contact via liaolangis@126.com or liaoliang2018@gmail.com.

APPENDIX: GENERALIZATION OF SUPERVISED CLASSIFICATION AND NEURAL NETWORKS

7.1. Generalized Supervised Classification. Besides the applications mentioned above, many other applications can be improved using the t-algebra paradigm as long as the scalar entries of each t-scalar are correlated, making sense of a linear transform, not necessarily the Fourier transform, of t-scalars.

The solution via a fixed-sized small neighborhood of each scalar, e.g., the solution in Section 5.2, is one of the most convenient approaches to establish a correlation between the scalar entries of a t-scalar for spatially constrained data, including but not limited to images, videos, audios, and sequential data such as time series.

Using the solution in Section 5.2, the generalization of canonical samples to their higher-order versions yields generalized inputs to generalized classifiers for supervised classification of legacy images.

Figure 7.1 summarizes the generalized classification, over the t-algebra C , of a canonical matrix sample X_{mat}^{raw} . After t-matricizing the canonical sample X_{mat}^{raw} , using the solution in Section 5.2, to its higher-order version X_{TM}^{raw} , the t-matrix X_{TM}^{raw} is sent to a generalized feature extractor, which is represented by K canonical sub-extractors, over complex numbers, where $K \doteq I_1 \times \dots \times I_N$.

The generalized extractor's output is a t-matrix $X_{TM}^{feature}$, either sent to a canonical classifier with pooling or sent without pooling to a generalized classifier represented by K canonical sub-classifiers.

With pooling, the t-matrix $X_{TM}^{feature}$ is transformed into a canonical feature matrix $X_{mat}^{feature}$. The query matrix $X_{mat}^{feature}$ and the training matrices $Y_{mat}^{feature,(1)}, \dots, Y_{mat}^{feature,(J)}$, as the inputs to a canonical classifier, yield a class label, i.e., an output in Figure 7.1 or 7.2, for $X_{mat}^{feature}$.

If the t-matrix $X_{TM}^{feature}$ is sent to a generalized classifier without pooling, a brief interpretation is needed. To be a little concrete, let us take the following generalization of the classifier Nearest Neighbor, called TNN (Tensorial Nearest Neighbor), for example.

TNN: Generalized Nearest Neighbor. The t-matrix $X_{TM}^{feature}$ sent to a generalized classifier is representable by K complex matrices $X_{mat,1}^{feature}, \dots, X_{mat,K}^{feature}$ such that

$$X_{TM}^{feature} = \sum_{k=1}^K X_{mat,k}^{feature} \ltimes Q_{T,k} . \quad (7.1)$$

Let the raw training matrices be $Y_{mat}^{raw,(1)}, \dots, Y_{mat}^{raw,(J)}$, and let their t-matrix versions be $Y_{TM}^{raw,(1)}, \dots, Y_{TM}^{raw,(J)}$. The j -th t-matrice $Y_{TM}^{raw,(j)}$ is sent to a generalized feature extractor, yielding an output t-matrix $Y_{TM}^{feature,(j)}$ which is represented by its K constituent matrices $Y_{mat,1}^{feature,(j)}, \dots, Y_{mat,K}^{feature,(j)}$ as follows

$$Y_{TM}^{feature,(j)} = \sum_{k=1}^K Y_{mat,k}^{feature,(j)} \ltimes Q_{T,k} , \quad \forall j \in [J] . \quad (7.2)$$

The matrices $Y_{mat,k}^{feature,(j)}, \forall (k,j) \in [K] \times [J]$, computed as in equation (7.2) from K training sets as follows

$$\mathcal{S}_k^{training} \doteq \left\{ Y_{mat,k}^{feature,(1)}, \dots, Y_{mat,k}^{feature,(J)} \right\} , \quad \forall k \in [K] . \quad (7.3)$$

The k -th training set $\mathcal{S}_k^{training}$, given by equation (7.3), and the k -th feature matrix $X_{mat,k}^{feature}$ in equation (7.1), as the inputs to the k -th constituent classifier of a generalized classifier, yields the k -th constituent output for all $k \in [K]$. The ensemble of these constituent outputs gives the final output of the generalized classifier TNN.

Given J training t-matrices $Y_{TM}^{feature,(1)}, \dots, Y_{TM}^{feature,(J)}$ and a query t-matrix $X_{TM}^{feature}$, the generalized classifier TNN, in the form of K constituent canonical classifiers, yields the generalized distances $d(X_{TM}^{feature}, Y_{TM}^{feature,(1)}), \dots, d(X_{TM}^{feature}, Y_{TM}^{feature,(J)}) \geq Z_T$ as follows

$$d(X_{TM}^{feature}, Y_{TM}^{feature,(j)}) \doteq \sum_{k=1}^K d(X_{mat,k}^{feature}, Y_{mat,k}^{feature,(j)}) \cdot Q_{T,k} \in S^{nonneg}, \quad \forall j \in [J]. \quad (7.4)$$

where $d(X_{mat,k}^{feature}, Y_{mat,k}^{feature,(j)}) \geq 0$, a nonnegative real number, denotes a canonical distance between the matrices $X_{mat,k}^{feature}$ and $Y_{mat,k}^{feature,(j)}$, and the matrices $X_{mat,k}^{feature}$ and $Y_{mat,k}^{feature,(j)}$ are given by equations (7.1) and (7.2).

For example, the canonical distances $d(X_{mat,k}^{feature}, Y_{mat,k}^{feature,(j)}), \forall (k,j) \in [K] \times [J]$, in equation (7.4) can be given by the following Frobenius norm as follows

$$d(X_{mat,k}^{feature}, Y_{mat,k}^{feature,(j)}) \doteq \|X_{mat,k}^{feature} - Y_{mat,k}^{feature,(j)}\|_F \geq 0, \quad \forall (k,j) \in [K] \times [J]. \quad (7.5)$$

Let $\mathcal{P} \doteq \{d(X_{TM}^{feature}, Y_{TM}^{feature,(1)}), \dots, d(X_{TM}^{feature}, Y_{TM}^{feature,(J)})\}$ be the poset formed by the generalized distances given by equation (7.4). If $\inf \mathcal{P} \in \mathcal{P}$, in other words, if there exists the least element in the poset \mathcal{P} , let the least element be $d(X_{TM}^{feature}, Y_{TM}^{feature,(j^*)}) \doteq \inf \mathcal{P} = \min \mathcal{P}$ where $j^* \in [J]$. Then, the label of the raw query matrix X_{mat} is identified with the label of t-matrix $Y_{TM}^{feature,(j^*)}$, more precisely,

$$\text{class}(X_{mat}) \equiv \text{class}(X_{TM}^{feature}) \doteq \text{class}(Y_{TM}^{feature,(j^*)}). \quad (7.6)$$

If the poset \mathcal{P} has no least element, without loss of generality, let the t-matrices $Y_{TM}^{feature,(1)}, \dots, Y_{TM}^{feature,(J_{min})}$ be the training samples, each having a minimum generalized distance, not the least generalized distance, to the query sample $X_{TM}^{feature}$.

Then, the label of the raw query matrix X_{mat} , or equivalently, the label of t-matrix $X_{TM}^{feature}$, can be identified with the label of any t-matrix among $Y_{TM}^{feature,(1)}, \dots, Y_{TM}^{feature,(J_{min})}$. More precisely, the following identity makes sense when $\inf \mathcal{P} \notin \mathcal{P}$.

$$\text{class}(X_{mat}) \equiv \text{class}(X_{TM}^{feature}) \doteq \text{class}(Y_{TM}^{feature,(j)}) \text{ for any } j \in \{1, \dots, J_{min}\}. \quad (7.7)$$

In this scenario, one can never claim that the classification accuracy is 100%. On the other hand, if a generalized extractor and a generalized classifier are appropriately tuned, we contend, the generalized classifier should yield more favorable results than its canonical counterpart.

In theory, a generalized version's computational cost is only K times that of the canonical counterpart. In Section 6.4 of [20], a comparison of the run time of the experiments of some t-matrix manipulations corroborates the above claims on computational cost. Interested readers are referred to as the results therein for more details.

Some well-known algorithms on supervised image classification/segmentation are generalized in our early work [20, 25] and achieve favorable results compared with their canonical counterparts. Interested readers are referred to the reported experiments for more details in [20, 25].

TCNN: Generalized Convolutional Neural Networks. Besides the generalized algorithms and classifiers in [20, 25], it is possible to generalize the popular Convolutional Neural Network (CNN). The t-matrix paradigm in Figure 7.1 applies to generalize the canonical CNN model for supervised visual-pattern classification.

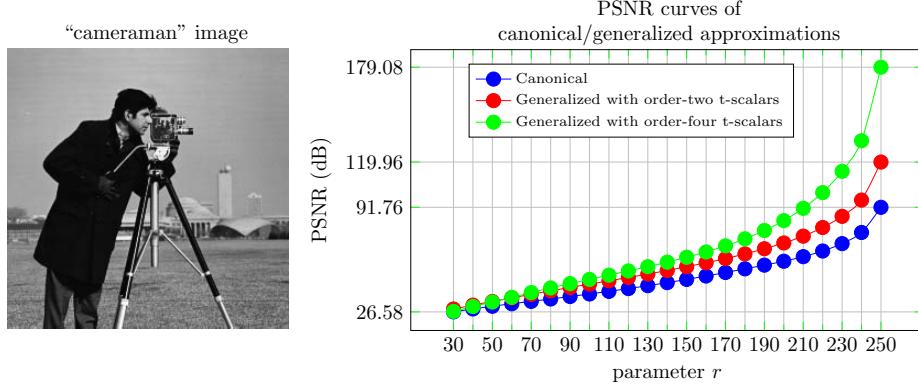
Figure 7.1 shows the diagram of a generalized CNN model over the t-algebra C . A generalized CNN over the t-algebra is represented by K (where $K \doteq I_1 \times \cdots \times I_N$) canonical CNNs trained by a set of labeled complex matrices decomposed from a generalized training set, e.g., a t-matrix set. If appropriately managed, we contend, the generalized CNN model should yield more favorable results than its canonical counterpart. We leave the verifications of this claim and the implementation of a generalized CNN for interested readers.

REFERENCES

- [1] ABLAMOWICZ, R., SOBCZYK, G., ET AL. *Lectures on Clifford (geometric) algebras and applications*. Springer, 2004.
- [2] BORALEVI, A., DRAISMA, J., HOROBET, E., AND ROBEVA, E. Orthogonal and unitary tensor decomposition from an algebraic perspective. *Israel journal of mathematics* 222, 1 (2017), 223–260.
- [3] BRACEWELL, R. N., AND BRACEWELL, R. N. *The Fourier transform and its applications*, 3rd ed. pp. 108–112. McGraw-Hill, New York, 1999.
- [4] BRAMAN, K. Third-order tensors as linear operators on a space of matrices. *Linear Algebra & Its Applications* 433, 7 (2010), 1241–1253.
- [5] CHEN, Z., WANG, B., NIU, Y., XIA, W., ZHANG, J. Q., AND HU, B. Change detection for hyperspectral images based on tensor analysis. In *Geoscience and Remote Sensing Symposium* (2015), pp. 1662–1665.
- [6] DE SILVA, V., AND LIM, L.-H. Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal on Matrix Analysis and Applications* 30, 3 (2008), 1084–1127.
- [7] ECKART, C., AND YOUNG, G. The approximation of one matrix by another of lower rank. *Psychometrika* 1, 3 (1936), 211–218.
- [8] ERDMANN, K., AND HOLM, T. *Algebras and representation theory*. Springer, 2018, ch. “Semisimple modules and semisimple algebras”, pp. 85–96.

- [9] FAN, H., LI, C., GUO, Y., KUANG, G., AND MA, J. Spatial-spectral total variation regularized low-rank tensor decomposition for hyperspectral image denoising. *IEEE Transactions on Geoscience & Remote Sensing* 56, 10 (2018), 6196–6213.
- [10] GLEICH, D. F., CHEN, G., AND VARAH, J. M. The power and arnoldi methods in an algebra of circulants. *Numerical Linear Algebra with Applications* 20, 5 (2013), 809–831.
- [11] HAMILTON, W. R. On quaternions; or on a new system of imaginaries in algebra. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 33, 219 (1848), 58–60.
- [12] HESTENES, D. Oersted medal lecture: Reforming the mathematical language of physics. *American Journal of Physics* 71, 2 (2003), 104–121.
- [13] HILGERT, J., AND NEEB, K.-H. *Structure and geometry of Lie groups*. Springer, 2011, ch. 2.3.
- [14] HILLAR, C. J., AND LIM, L.-H. Most tensor problems are NP-hard. *Journal of the ACM* 60, 6 (2013), 1–39.
- [15] HUNGERFORD, T. W. *Algebra*. Springer, 1980, ch. I.1: Semigroups, Monoids and Groups, p. 24.
- [16] HUNGERFORD, T. W. *Algebra*. Springer, 1980, ch. IX: The Structure of Rings, pp. 415–455.
- [17] KILMER, M. E., BRAMAN, K., HAO, N., AND HOOVER, R. C. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM Journal on Matrix Analysis & Applications* 34, 1 (2013), 148–172.
- [18] KILMER, M. E., AND MARTIN, C. D. Factorization strategies for third-order tensors. *Linear Algebra & Its Applications* 435, 3 (2011), 641–658.
- [19] LANG, S. *Algebra (Revised Third Edition)*. Springer, 2002, ch. I.11: Categories and Functors, p. 53.
- [20] LIAO, L., AND MAYBANK, S. J. Generalized visual information analysis via tensorial algebra. *Journal of Mathematical Imaging and Vision* (2020), 1–25.
- [21] LIAO, L., MAYBANK, S. J., ZHANG, Y., AND LIU, X. Supervised classification via constrained subspace and tensor sparse representation. In *International Joint Conference on Neural Networks* (2017), pp. 2306–2313.
- [22] LU, C., FENG, J., CHEN, Y., LIU, W., LIN, Z., AND YAN, S. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 5249–5257.
- [23] LU, C., FENG, J., CHEN, Y., LIU, W., LIN, Z., AND YAN, S. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE transactions on pattern analysis and machine intelligence* 42, 4 (2019), 925–938.
- [24] MIRSKY, L. Symmetric gauge functions and unitarily invariant norms. *The quarterly journal of mathematics* 11, 1 (1960), 50–59.
- [25] REN, Y., LIAO, L., MAYBANK, S. J., ZHANG, Y., AND LIU, X. Hyperspectral image spectral-spatial feature extraction via tensor principal component analysis. *IEEE Geoscience and Remote Sensing Letters* 14, 9 (2017), 1431–1435.
- [26] ROZENFEL'D, B. A. *The history of non-euclidean geometry: Evolution of the concept of a geometric space*. Springer, 1988, p. 385.
- [27] VANNIEUWENHOVEN, N., NICHAISE, J., VANDEBRIL, R., AND MEERBERGEN, K. On generic nonexistence of the schmidt–eckart–young decomposition for complex tensors. *SIAM Journal on Matrix Analysis and Applications* 35, 3 (2014), 886–903.
- [28] ZHANG, J., SAIBABA, A. K., KILMER, M., AND AERON, S. A randomized tensor singular value decomposition based on the t-product. *Numerical Linear Algebra with Applications* 25, 5 (2018).

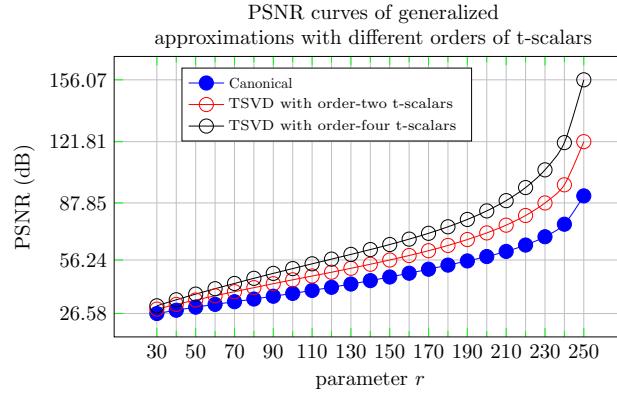
- [29] ZHANG, Z., ELY, G., AERON, S., HAO, N., AND KILMER, M. Novel methods for multilinear data completion and de-noising based on tensor-SVD. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Columbus, Ohio, USA, 2014), pp. 3842–3849.



Tabulated PSNRs (dB) of canonical/generalized low-rank approximations with different rank parameter r
 Canonical = approximation via SVD, Generalized* = inceptive slice approximation via TSVD with order-two t-scalars, Generalized† = inceptive slice approximation via TSVD with order-four t-scalars

APPROXIMATION TYPE	parameter r											
	30	50	70	90	110	130	150	170	190	210	230	250
Canonical	26.58	30.11	33.14	36.17	39.38	42.91	46.83	51.14	55.72	61.00	69.12	91.76
Generalized*	28.36	33.08	37.51	41.82	45.97	50.16	54.75	59.81	65.93	73.77	86.12	119.96
Generalized†	26.90	32.72	38.60	44.19	49.45	54.76	60.63	67.72	77.33	91.13	114.19	179.08

FIGURE 6.3. A comparison of approximations by TSVD, using inception slice and different array orders, and SVD where $N^{entry} = 65536$.



Tabulated PSNRs (dB) of canonical/generalized low-rank approximations with different rank parameter r
 Canonical = approximation via SVD, TSVD* = approximation via TSVD with order-two t-scalars, TSVD† = approximation via TSVD with order-four t-scalars

APPROXIMATION TYPE	parameter r											
	30	50	70	90	110	130	150	170	190	210	230	250
Canonical	26.58	30.11	33.14	36.17	39.38	42.91	46.83	51.14	55.72	61.00	69.12	91.76
TSVD*	29.01	34.06	38.69	43.11	47.35	51.64	56.24	61.38	67.58	75.54	87.85	121.81
TSVD†	30.85	37.34	43.24	48.83	54.13	59.34	64.81	71.03	78.71	89.13	106.19	156.07

FIGURE 6.4. A comparison of approximations by TSVD with $K \in \{9, 81\}$ and SVD with $K = 1$ where PSNRs are computed with $N^{entry} = 65536 \times K$.

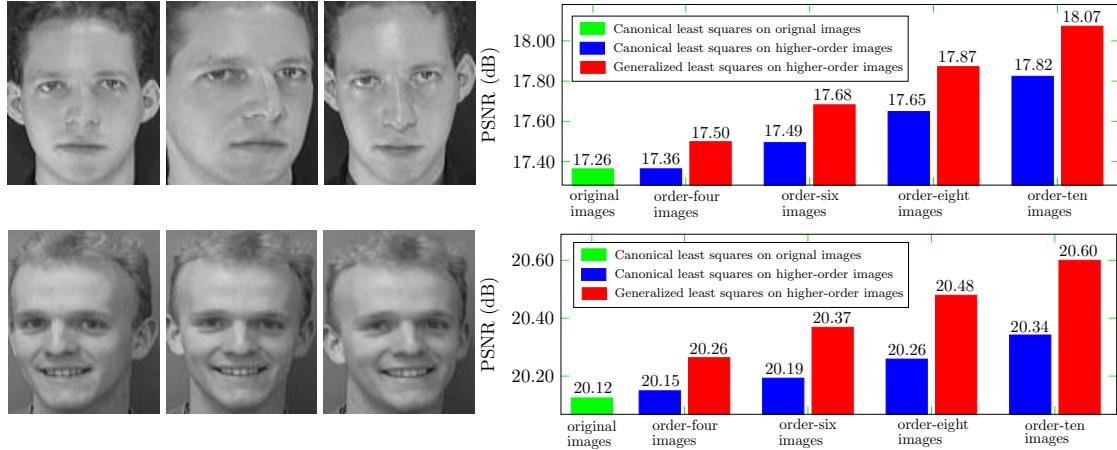


FIGURE 6.5. A quantitative comparison of approximations in PSNR by canonical least squares and generalized least squares on the ORL image dataset

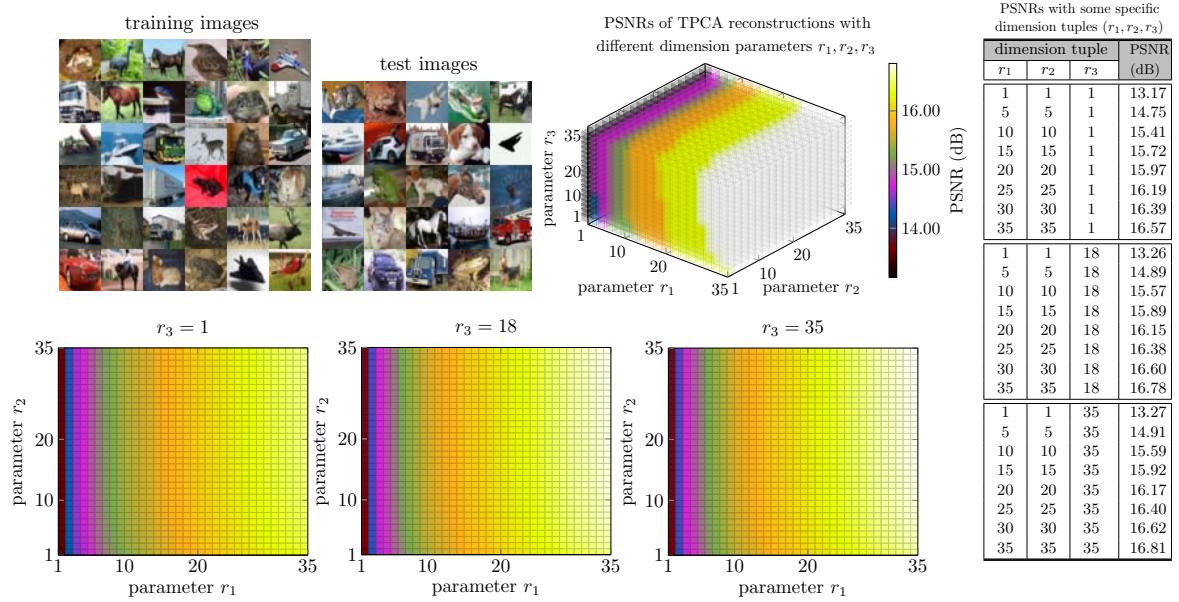


FIGURE 6.6. PSNRs by TPCA with various parameters (r_1, r_2, r_3) , or equivalently, $H_T = \sum_{k=1}^3 r_k \cdot Q_{T,k}$, on the CIFAR-10 images.

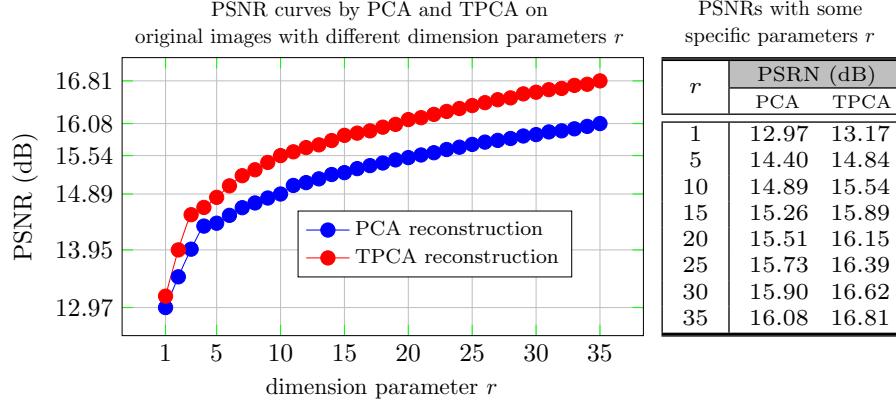


FIGURE 6.7. PSNRs by PCA and TPCA with various parameters r on the CIFAR-10 images

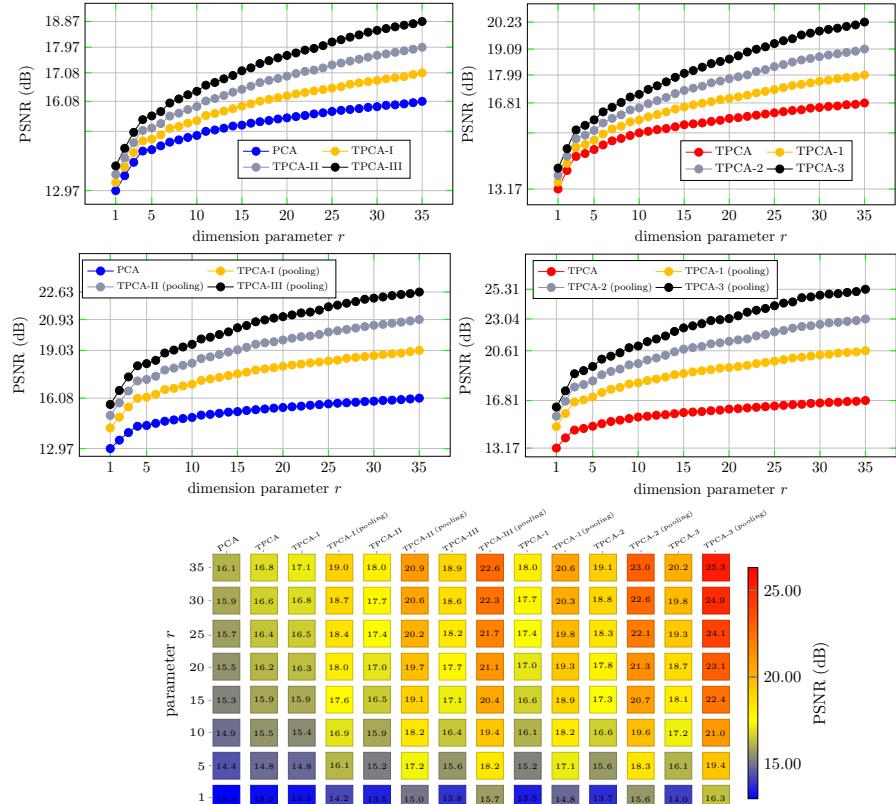


FIGURE 6.8. A comparison of PRNS by PCA and TPCA variants with various t-scalar orders on the CIFAR-10 images. TPCA-I = TPCA with order-two t-scalars, TPCA-II = TPCA with order-four t-scalars, TPCA-III = TPCA with order-six t-scalars, TPCA-1 = TPCA with order-three t-scalars, TPCA-2 = TPCA with order-five t-scalars, TPCA-3 = TPCA with order-seven t-scalars

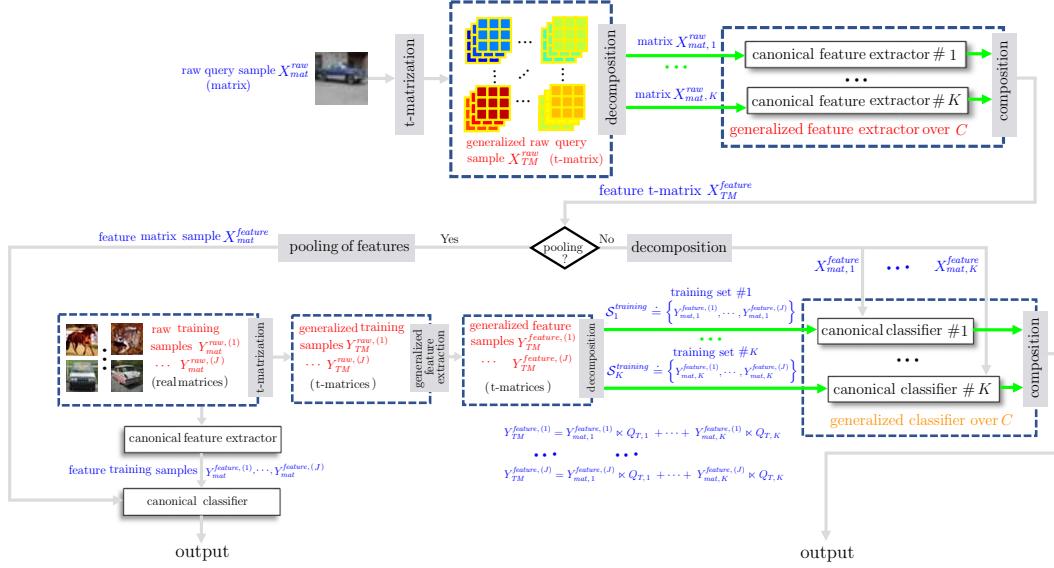


FIGURE 7.1. Generalized classification over C with or without pooling of features.

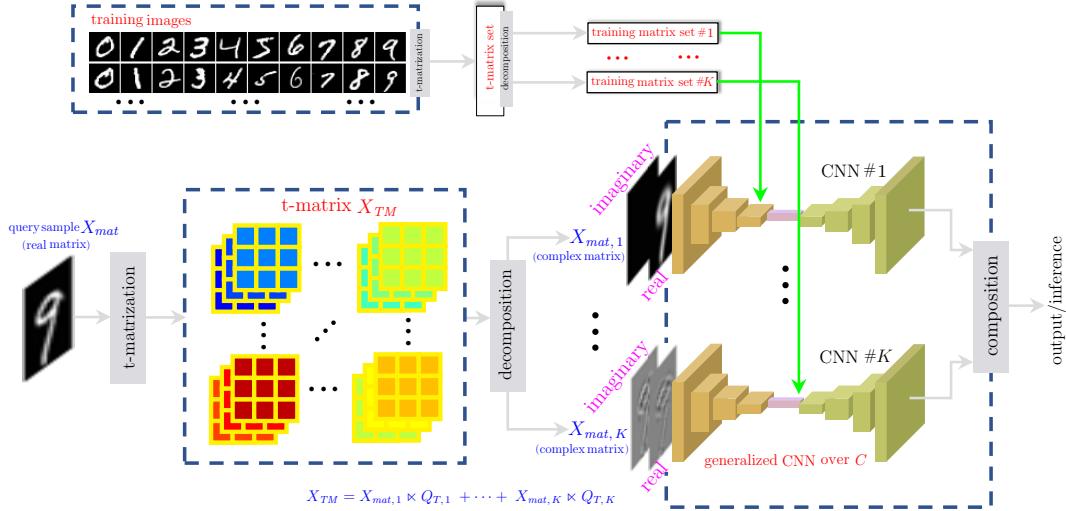


FIGURE 7.2. A diagram of generalized CNN (Convolutional Neural Network) model over C .