This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2021.3084541, IEEE Open Journal of Signal Processing

A Tensor Framework for Multi-linear Complex MMSE Estimation

Divyanshu Pandey, Harry Leib

Tensors are higher order generalization of vectors and matrices which can be used to describe signals indexed by more than two indices. This paper introduces a tensor framework for minimum mean square error (MMSE) estimation for multi-domain signals and data using the Einstein Product. The framework addresses both proper and improper complex tensors. The multi-domain nature of tensors has been harnessed to provide an augmented representation of improper complex tensors to account for covariance and pseudo-covariance. The classical notions of linear and widely linear MMSE estimators are extended to tensor case leading to the notion of multi-linear and widely multi-linear MMSE estimation. The Tucker product based n-mode Wiener filtering approach more commonly used in tensor estimation has been shown to be a special case of the proposed multi-linear MMSE estimation. An application of the tensor based estimation in a multiple antenna Orthogonal Frequency Division Multiplexing (MIMO OFDM) system is presented where the tensor formulation allows a convenient treatment of inter-carrier interference. A comparison between the tensor estimation and per sub-carrier estimation used for MIMO OFDM is presented which shows a significant performance advantage of using the tensor framework.

***Index Terms*—Einstein Product, Multi-linear MMSE, Tensors, Tensor Train format, Widely Multi-linear MMSE.**

I. Introduction

Minimum mean square error (MMSE) estimation has been extensively used in various disciplines including seismology, radio astronomy, sonar, speech and image processing, radar, medical signal processing, and communications [1]. However in many modern day applications, the data or signals to be estimated have an inherent dependency on more than two indices. For such multi-domain signals, classical MMSE estimation (based on vector and matrix operations) can not be directly employed. Such multi-domain data is often represented using tensors which can be seen as a generalization of vectors and matrices to higher orders. Tensors are multi-way arrays whose elements are indexed by more than two indices, also known as *modes*. The number of modes is called the *order* of the tensor. Hence matrices and vectors can be seen as order 2 and order 1 tensors respectively [2]. Tensors have found widespread applications in various engineering disciplines including computer vision [3], [4], signal processing [5]–[8], big data and machine learning [9]–[12], image processing [13], [14], communications [15]–[18], and multi-linear system theory [19], [20]. Our work considers the MMSE estimation problem in context of tensors, thereby extending the basic subject beyond the common vectors and matrices settings.

So far, tensor based estimation techniques have been addressed in the literature for specific applications. For instance, use of tensors for channel estimation in relay based multiple input multiple output (MIMO) systems has been considered in [21], [22]. Blind receivers based on Parallel Factors (PARAFAC) decomposition of tensors have been considered for Direct-Sequence Code Division Multiple Access (DS-CDMA) systems in [23] and for MIMO systems in [24], [25]. A tensor based MMSE channel estimation technique using compressive sensing was proposed for massive MIMO Orthogonal Frequency Division Multiplexing (OFDM) systems in [26] and using PARAFAC decomposition for Millimeter wave MIMO OFDM systems in [27]. In Image Processing, tensor based estimation has been used for de-noising by treating colored images as third order tensors and using a Tucker product based estimator [13], [28]. However, Tucker and PARAFAC operations can be represented by the Einstein product as explained in [29]. Hence, the Einstein product can be used to develop a generic framework for MMSE estimation in many applications.

The most common methods for tensor estimation found in literature are based on the Tucker or PARAFAC decompositions. The Tucker product based technique is also known as the n-mode Wiener filtering approach [30]. This method aims to find N separate factor matrices along each mode of the order N tensor to be estimated. The mode-n product between the observed noisy tensor and the factor matrices is then used to find the estimate. Such an approach has been employed in various applications including Image processing [30], speech processing [31], and communication systems [26]. However, the additional constraint assuming that the multi-linear estimator is separable across all the modes, makes the Tucker approach sub-optimal within the class of multi-linear estimators. The PARAFAC based technique [32] is also known as the Canonical Polyadic (CP) decomposition approach. The CP model decomposes a tensor into a sum of factor rank one tensors. Estimation using CP model relies on matrix unfolding of the observed tensor and uses an alternating least squares method to find the separate factor tensors [32]. Such an estimation technique assumes a finite rank decomposition of the tensor to be estimated, and is often used for tensor completion problems [33]. Further, it has applications in communication systems for joint channel and symbol estimation if the signal can be assumed to have a low rank CP structure [27], [34]. The estimation method proposed in this paper uses the Einstein product to find a multi-linear operator where no such constraints on the tensor to be estimated or the estimator are assumed.

In this paper, we present a generic tensor framework for MMSE estimation based on the Einstein product, which is concerned with estimating a signal in tensor form from a

This work was supported by grant RGPIN-2016-03647 entitled “Tensor modulation for space-time-frequency communication systems”, awarded by the Natural Sciences and Engineering Research Council of Canada (NSERC).

The authors are from the Department of Electrical and Computer Engineering, McGill University, Montreal, Quebec, H3A 0E9 Canada (email : divyanshu.pandey@mail.mcgill.ca, harry.leib@mcgill.ca)

tensor based noisy observation. Many situations in practical cases involve signals that naturally involve multiple indices such as MIMO OFDM, Generalized Frequency Division Multiplexing (GFDM), Filter-bank Multi-carrier (FBMC) systems [15]. Hence representing such signals by tensors is natural. The proposed framework provides a structured method based on a solid mathematical foundation to handle such signals. Our framework is generic as we consider estimation of complex data, both proper and improper. A majority of works that deal with complex random signal, assume a proper regime where a signal is uncorrelated with its complex conjugate. This assumption often simplifies the problem formulation but can not be justified in all cases. If the signal is improper, then covariance is no longer sufficient to characterize the second order statistics and the pseudo-covariance also needs to be accounted for. In case of vectors, augmented representation of complex data has been used for dealing with improper complex signals [35]. In this paper, we harness the multi-domain nature of tensors to provide an augmented representation of improper tensor signals.

In [36] we presented the multi-linear MMSE estimator for proper tensors. In this paper we consider the more general case of widely multi-linear estimation, that makes [36] a special case of this work. Furthermore, in addition to [36] this paper also presents the best MMSE estimate for tensors and shows that for jointly complex Gaussian tensors, the best estimator is the widely multi-linear estimator. We also present application of the tensor estimation in a MIMO OFDM system with both proper and improper inputs. Furthermore, the use of the proposed estimation technique in cases where the tensor is stored in Tensor Train (TT) decomposed format is also presented. The data in many scenarios, particularly in Big Data applications, is often stored in TT decomposed formats which restricts the freedom to convert the data into matrices or vectors. We show the application of multi-linear MMSE estimator for tensors stored in TT format. Also, we present a detailed comparison between our proposed MMSE estimator and the more commonly used estimator based on Tucker product, in terms of complexity and the mean square error performance.

The Einstein product is a form of tensor contraction which was initially used in Physics and Continuum Mechanics. It can be seen as a natural generalization of the matrix product to higher order arrays. Lately it has been considered to define notions for multi-linear algebra without reshaping the tensors into vectors and matrices [29], [37]. Since different modes of a tensor can attribute different physical meanings depending on the system model, retaining the distinction between domains is of paramount importance in order to leverage the information encapsulated in the structure of a tensor. For instance, in a communication system different modes can represent different domains such as space, time, frequency, codes, channel taps and users. Such a multi-domain approach towards communication systems using tensors can be employed in many multi-users (MU), multi-antenna and multi-carrier schemes [15].

This paper is organized as follows : a brief review of tensor algebra is presented in section II. Section III introduces the tensor MMSE estimation problem. We present the best MMSE estimator followed by a widely multi-linear MMSE estimator which can be used for improper signals. A comparison between the proposed estimator using Einstein product and the Tucker product approach, both in terms of mean square error performance and computational complexity is also presented. Section IV presents numerical examples with applications of the tensor framework for MMSE estimation of Gaussian signals, tensors stored in TT format and multi-domain communication systems. Simulation results for MIMO OFDM systems are presented to illustrate the performance of various tensor estimation techniques. The paper is concluded in Section V. Further, Appendices A, B and C present detailed proofs of some results in the paper and Appendix D presents alternate and faster approaches to implement Newton method for tensor inversion.

II. Tensor Algebra

A. Notations

Throughout this paper, deterministic vectors, matrices and tensors will be represented using lowercase underlined fonts, uppercase fonts, and uppercase calligraphic fonts respectively e.g. $\underline{\mathrm{x}}, \mathrm{X}$, and $\mathcal{X}$. Their corresponding random quantities will be denoted by bold fonts, e.g. $\underline{\mathbf{x}}, \mathbf{X}$ and $\boldsymbol{\mathcal{X}}$. Individual entries of a tensor are denoted with the indices in subscript, e.g. the (i,j,k)th element of a third order tensor $\mathcal{X}$ is denoted by $\mathcal{X}_{i,j,k}$. A colon in subscript is used to indicate all elements of a mode, e.g. $\mathcal{X}_{:,j,k}$ represents all the elements of first mode corresponding to jth second and kth third mode. All the augmented variables including vector, matrix or tensor are represented with two dots on overline as $\ddot{\underline{\mathrm{x}}}, \ddot{\mathrm{X}}, \ddot{\mathcal{X}}$. Furthermore, $()^*$ represents the complex conjugate, $\mathbb{E}[.]$ represents expectation, $\Re()$ and $\Im()$ represents the real and imaginary part of a complex entity, and $0_{\mathcal{T}}$ represents an all zero tensor.

B. Basic Definitions

In this section, we only provide some relevant tensor definitions and properties needed for this paper. More details on tensor algebra can be found in [2], [12], [15], [29], [37], [38].

Definition 1. Tensor Linear Space : The set of all tensors of size $I_1 \times \ldots \times I_K$ over $\mathbb{C}$ forms a linear space, denoted as $\mathbb{T}_{I_1,\ldots,I_K}(\mathbb{C})$. For $\mathcal{A}, \mathcal{B} \in \mathbb{T}_{I_1,\ldots,I_K}(\mathbb{C})$ and $\alpha \in \mathbb{C}$, the sum $\mathcal{A}+\mathcal{B} = \mathcal{C} \in \mathbb{T}_{I_1,\ldots,I_K}(\mathbb{C})$ where $\mathcal{C}_{i_1,\ldots,i_k} = \mathcal{A}_{i_1,\ldots,i_k} + \mathcal{B}_{i_1,\ldots,i_k}$, and scalar multiplication $\alpha \cdot \mathcal{A} = \mathcal{D} \in \mathbb{T}_{I_1,\ldots,I_K}(\mathbb{C})$ where $\mathcal{D}_{i_1,\ldots,i_k} = \alpha \mathcal{A}_{i_1,\ldots,i_k}$.

A tensor valued function $g(\mathcal{X})$ can be defined as a function whose domain and range are both tensors. Notationally, $g : \mathbb{C}^{I_1\times\ldots\times I_N} \rightarrow \mathbb{C}^{J_1\times\ldots\times J_M}$ represents a function where $\mathbb{C}^{I_1\times\ldots\times I_N}$ and $\mathbb{C}^{J_1\times\ldots\times J_M}$ are tensor linear spaces spanned by the domain and range of function g respectively.

Definition 2. Einstein product [29]: For any N, the Einstein product is defined using the operation $*_N$, where :

$$(\mathcal{A} *_N \mathcal{B})_{i_1,\ldots,i_P,j_1,\ldots,j_M} = \sum_{k_1,\ldots,k_N} \mathcal{A}_{i_1,i_2,\ldots,i_P,k_1,\ldots,k_N} \mathcal{B}_{k_1,\ldots k_N,j_1,j_2,\ldots,j_M} \quad (1)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2021.3084541, IEEE Open Journal of Signal Processing

for tensors $\mathcal{A} \in \mathbb{C}^{I_1\times\ldots\times I_P\times K_1\times\ldots\times K_N}$, $\mathcal{B} \in \mathbb{C}^{K_1\times\ldots\times K_N\times J_1\times\ldots\times J_M}$, and $\mathcal{A}*_N\mathcal{B} \in \mathbb{C}^{I_1\times\ldots\times I_P\times J_1\times\ldots\times J_M}$.

For tensors $\mathcal{X},\mathcal{Y} \in \mathbb{C}^{I_1\times I_2\times\ldots\times I_N}$ and $\mathcal{Z} \in \mathbb{C}^{J_1\times J_2\times\ldots\times J_M}$, using Einstein product we can define the inner product $\langle\mathcal{X},\mathcal{Y}\rangle$ and the outer product $\mathcal{X}\circ\mathcal{Z}$ as :

$$\langle\mathcal{X},\mathcal{Y}\rangle = \sum_{i_1=1}^{I_1}\ldots\sum_{i_N=1}^{I_N}\mathcal{X}_{i_1,i_2,\ldots,i_N}\mathcal{Y}^*_{i_1,i_2,\ldots,i_N} = \mathcal{X}*_N\mathcal{Y}^* \quad (2)$$

$$(\mathcal{X}\circ\mathcal{Z})_{i_1,\ldots,i_N,j_1,\ldots,j_M} = \mathcal{X}_{i_1,\ldots,i_N}\mathcal{Z}_{j_1,\ldots,j_M} = \mathcal{X}*_0\mathcal{Z} \quad (3)$$

$$\textit{Norm}: ||\mathcal{X}|| = \sqrt{(\mathcal{X}*_N\mathcal{X}^*)} \quad (4)$$

Tensors $\mathcal{X}$ and $\mathcal{Y}$ are said to be mutually *orthogonal* if their inner product (2) is 0 [39]. Using [29, Lemma 3.1] for order 4 tensors, which was further generalized in [37], [40] for any size tensors, several linear algebra definitions can be extended to a multi-linear setting as follows :

- A tensor $\mathcal{A} \in \mathbb{C}^{I_1\times\ldots\times I_N\times J_1\times\ldots\times J_M}$ is called a *square* tensor if $N = M$ and $I_k = J_k$ for $k = 1,\ldots,N$ [37].
- A square tensor $\mathcal{D} \in \mathbb{C}^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}$ is *pseudo-diagonal* if all its entries $\mathcal{D}_{i_1,\ldots,i_N,j_1,\ldots,j_N}$ are zero except when $i_1 = j_1, i_2 = j_2,\ldots,i_N = j_N$ [15].
- An *identity tensor*, $\mathcal{I} \in \mathbb{C}^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}$ is a square pseudo-diagonal tensor such that for any square tensor $\mathcal{A} \in \mathbb{C}^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}$, $\mathcal{A}*_N\mathcal{I} = \mathcal{I}*_N\mathcal{A} = \mathcal{A}$.
- The tensor $\mathcal{A}^{-1} \in \mathbb{C}^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}$ is an *inverse* of a square tensor of same size, $\mathcal{A} \in \mathbb{C}^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}$ if $\mathcal{A}*_N\mathcal{A}^{-1} = \mathcal{A}^{-1}*_N\mathcal{A} = \mathcal{I}$ [37].
- The *Hermitian* of a tensor $\mathcal{A} \in \mathbb{C}^{I_1\times\ldots\times I_N\times J_1\times\ldots\times J_M}$ is a tensor $\mathcal{B} \in \mathbb{C}^{J_1\times\ldots\times J_M\times I_1\times\ldots\times I_N}$ which has entries $\mathcal{B}^*_{j_1,j_2,\ldots,j_M,i_1,i_2,\ldots,i_N} = \mathcal{A}_{i_1,i_2,\ldots,i_N,j_1,j_2,\ldots,j_M}$ and is denoted as $\mathcal{A}^H$. We denote transpose as $\mathcal{A}^T$. Since tensors could have more than two modes, so there can be multiple ways to define a tensor Hermitian or transpose. To avoid overload of notation, in this paper whenever we write a tensor explicitly as $N+M$ or $2N$ order tensor, then $\mathcal{A}^H$ or $\mathcal{A}^T$ is always with respect to partition after N modes.
- A square tensor $\mathcal{X} \in \mathbb{C}^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}$ is called a *Hermitian tensor* if $\mathcal{X} = \mathcal{X}^H$, and a *unitary* tensor if $\mathcal{X}^H*_N\mathcal{X} = \mathcal{X}*_N\mathcal{X}^H = \mathcal{I}$.
- Let $\mathcal{A} \in \mathbb{C}^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}$, $\mathcal{X} \in \mathbb{C}^{I_1\times\ldots\times I_N}$, $\lambda \in \mathbb{C}$, where $\mathcal{X}$ and λ satisfy $\mathcal{A}*_N\mathcal{X} = \lambda\mathcal{X}$, then we call $\mathcal{X}$ and λ as *eigentensor* and *eigenvalue* of $\mathcal{A}$ respectively [14].
- The EVD of a Hermitian tensor $\mathcal{A} \in \mathbb{C}^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}$ is given as $\mathcal{A} = \mathcal{U}*_N\mathcal{D}*_N\mathcal{U}^H$ where $\mathcal{U} \in \mathbb{C}^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}$ is a unitary tensor and $\mathcal{D} \in \mathbb{C}^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}$ is a square pseudo-diagonal tensor with its non-zero values being the eigenvalues of $\mathcal{A}$ and $\mathcal{U}$ containing the eigentensors of $\mathcal{A}$.
- A square tensor as *positive semi-definite*, denoted by $\mathcal{A} \succeq 0$, if all its eigenvalues are non-negative.
- The *trace* of a tensor is defined as the sum of its pseudo-diagonal entries, i.e. $\mathrm{tr}(\mathcal{A}) = \sum_{i_1,\ldots,i_N}\mathcal{A}_{i_1,i_2,\ldots,i_N,i_1,i_2,\ldots,i_N}$.
- The *determinant* of a tensor is defined as the product of its eigenvalues i.e., $\det(\mathcal{A}) = \prod_{i_1,\ldots,i_N}\mathcal{D}_{i_1,i_2,\ldots,i_N,i_1,i_2,\ldots,i_N}$. It is also sometimes referred as the unfolding determinant [20], [37].

Further, the following properties can also be established:

1) For tensors $\mathcal{A} \in \mathbb{C}^{I_1\times\ldots\times I_P\times J_1\times\ldots\times J_N}$, $\mathcal{B} \in \mathbb{C}^{J_1\times\ldots\times J_N\times K_1\times\ldots\times K_M}$ and $\mathcal{C} \in \mathbb{C}^{K_1\times\ldots\times K_M\times T_1\times\ldots\times T_Q}$, we have
$$(\mathcal{A}*_N\mathcal{B})*_M\mathcal{C} = \mathcal{A}*_N(\mathcal{B}*_M\mathcal{C})$$
$$(\mathcal{A}*_N\mathcal{B})\circ\mathcal{C} = \mathcal{A}*_N(\mathcal{B}\circ\mathcal{C}) \quad (5)$$
2) For $\mathcal{A} \in \mathbb{C}^{I_1\times\ldots\times I_P\times J_1\times\ldots\times J_N}$ and $\mathcal{B} \in \mathbb{C}^{J_1\times\ldots\times J_N}$,
$$\mathcal{A}*_N\mathcal{B} = \mathcal{B}*_N\mathcal{A}^T \text{ and } (\mathcal{A}*_N\mathcal{B})^* = \mathcal{B}^**_N\mathcal{A}^H \quad (6)$$
3) For tensors $\mathcal{A} \in \mathbb{C}^{I_1\times\ldots\times I_M\times J_1\times\ldots\times J_N}$ and $\mathcal{B} \in \mathbb{C}^{J_1\times\ldots\times J_N\times K_1\times\ldots\times K_P}$:
$$(\mathcal{A}*_N\mathcal{B})^H = \mathcal{B}^H*_N\mathcal{A}^H \quad (7)$$
4) For $\mathcal{A}$ and $\mathcal{B} \in \mathbb{C}^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}$, we have :
$$(\mathcal{A}*_N\mathcal{B})^{-1} = \mathcal{B}^{-1}*_N\mathcal{A}^{-1} \quad (8)$$
5) For two tensors $\mathcal{A} \in \mathbb{C}^{I_1\times\ldots\times I_N}$ and $\mathcal{B} \in \mathbb{C}^{I_1\times\ldots\times I_N}$ of same size and order N,
$$\mathcal{A}*_N\mathcal{B} = \mathcal{B}*_N\mathcal{A} = \mathrm{tr}(\mathcal{A}\circ\mathcal{B}) = \mathrm{tr}(\mathcal{B}\circ\mathcal{A}) \quad (9)$$
6) For tensors $\mathcal{A} \in \mathbb{C}^{I_1\times\ldots\times I_N\times J_1\times\ldots\times J_M}$ and $\mathcal{B} \in \mathbb{C}^{J_1\times\ldots\times J_M\times I_1\times\ldots\times I_N}$, we have :
$$\mathrm{tr}(\mathcal{A}*_M\mathcal{B}) = \mathrm{tr}(\mathcal{B}*_N\mathcal{A}) \quad (10)$$

In this paper, to find the inverse of a tensor $\mathcal{A} \in \mathbb{C}^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}$, we use the Newton Method for tensor inversion proposed in [36]. To approximate $\mathcal{A}^{-1}$, the Newton Method requires to solve the following iterative equation:

$$\mathcal{B}^{(k+1)} = (2\mathcal{I} - \mathcal{B}^{(k)}*_N\mathcal{A})*_N\mathcal{B}^{(k)} \quad (11)$$

where the initial $\mathcal{B}^{(0)}$ can be set to $a\cdot\mathcal{A}^H$. Using similar line of arguments as in Theorem 2 in [41], it can be shown that this method converges if $0 < a < 2/\sigma^2_{max}$ where σ^2_{max} is the largest eigenvalue of $\mathcal{C} = \mathcal{A}^H*_N\mathcal{A}$.

C. Second order characteristics of complex random tensors and augmented representation

The covariance of a complex tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1\times I_2\times\ldots\times I_N}$ can be defined as a tensor of size $I_1\times I_2\times\ldots\times I_N\times I_1\times I_2\times\ldots\times I_N$ represented by $\mathcal{Q} = \mathbb{E}[(\boldsymbol{\mathcal{X}}-\mathcal{M})\circ(\boldsymbol{\mathcal{X}}-\mathcal{M})^*]$ where $\mathcal{M} = \mathbb{E}[\boldsymbol{\mathcal{X}}]$ is the mean tensor. However, a complete second-order characterization of complex tensors which accounts for correlation between the real and imaginary components of the tensor as well, requires defining pseudo-covariance, also known as complementary covariance [42]. The pseudo-covariance of tensor $\boldsymbol{\mathcal{X}}$ is given as $\tilde{\mathcal{Q}} = \mathbb{E}[(\boldsymbol{\mathcal{X}}-\mathcal{M})\circ(\boldsymbol{\mathcal{X}}-\mathcal{M})]$. Similarly, cross covariance and cross pseudo-covariance between random tensors $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ can be defined as $\mathcal{Q}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}} = \mathbb{E}[(\boldsymbol{\mathcal{X}}-\mathbb{E}[\boldsymbol{\mathcal{X}}])\circ(\boldsymbol{\mathcal{Y}}-\mathbb{E}[\boldsymbol{\mathcal{Y}}])^*]$ and $\tilde{\mathcal{Q}}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}} = \mathbb{E}[(\boldsymbol{\mathcal{X}}-\mathbb{E}[\boldsymbol{\mathcal{X}}])\circ(\boldsymbol{\mathcal{Y}}-\mathbb{E}[\boldsymbol{\mathcal{Y}}])]$ respectively. We define a complex random tensor to be *proper* if its pseudo-covariance vanishes, i.e $\tilde{\mathcal{Q}} = 0_{\mathcal{T}}$. Similarly random tensors $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ are called *cross proper* if their cross pseudo-covariance is $0_{\mathcal{T}}$. Corresponding definitions for vectors can be found in [35], [42]. For vectors, an augmented representation completely defines the second order characteristics where the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2021.3084541, IEEE Open Journal of Signal Processing

augmented vector is $\underline{\ddot{\mathbf{x}}} = \begin{bmatrix} \underline{\mathbf{x}} \\ \underline{\mathbf{x}}^* \end{bmatrix}$ [43], [44]. Next, we will exploit the multi-domain nature of tensors to develop an augmented representation of complex tensors.

Unlike a vector, a tensor has more than one mode, therefore concatenation by the conjugate tensor can be done across any mode. However given that the primary advantage of a tensor is its ability to maintain distinction between different domains, we suggest that for a complex valued tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \ldots \times I_N}$, the augmented tensor can be created by adding another domain of size 2 such that $\ddot{\mathcal{X}} \in \mathbb{C}^{I_1 \times \ldots \times I_N \times 2}$ where $\ddot{\mathcal{X}}_{i_1,\ldots,i_N,1} = \mathcal{X}_{i_1,\ldots,i_N}$ and $\ddot{\mathcal{X}}_{i_1,\ldots,i_N,2} = \mathcal{X}^*_{i_1,\ldots,i_N}$. The augmented mean tensor will be $\ddot{\mathcal{M}} = \mathbb{E}[\ddot{\mathcal{X}}]$, and the augmented covariance tensor will be given as : $\ddot{\mathcal{Q}} = \mathbb{E}[(\ddot{\mathcal{X}} - \ddot{\mathcal{M}}) \circ (\ddot{\mathcal{X}} - \ddot{\mathcal{M}})^*]$ of size $I_1 \times \ldots \times I_N \times 2 \times I_1 \times \ldots \times I_N \times 2$.

The augmented covariance tensor $\ddot{\mathcal{Q}}$ contains the covariance $\mathcal{Q}$ and the pseudo-covariance tensor $\tilde{\mathcal{Q}}$ along with their conjugates as $\ddot{\mathcal{Q}}_{\underbrace{:,\ldots,:}_{N},1,\underbrace{:,\ldots,:}_{N},1} = \mathcal{Q}$, $\ddot{\mathcal{Q}}_{:,\ldots,:,1,:,\ldots,:,2} = \tilde{\mathcal{Q}}$, $\ddot{\mathcal{Q}}_{:,\ldots,:,2,:,\ldots,:,1} = \tilde{\mathcal{Q}}^*$ and $\ddot{\mathcal{Q}}_{:,\ldots,:,2,:,\ldots,:,2} = \mathcal{Q}^*$.

Definition 3. The pdf of a general Gaussian distributed complex-valued tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ of order N is given by :

$$p_{\mathcal{X}}(\underline{\mathrm{x}}) = \frac{\exp\{-\frac{1}{2}(\underline{\ddot{\mathrm{x}}} - \underline{\ddot{\mathrm{m}}})^H \ddot{\mathbf{Q}}^{-1} (\underline{\ddot{\mathrm{x}}} - \underline{\ddot{\mathrm{m}}})\}}{(\pi)^{I_1 I_2 \ldots I_N} (\det(\ddot{\mathbf{Q}}))^{1/2}} \tag{12}$$

where $\underline{\ddot{\mathrm{x}}} = \mathrm{vec}(\ddot{\mathcal{X}})$, $\underline{\ddot{\mathrm{m}}} = \mathrm{vec}(\mathbb{E}[\ddot{\mathcal{X}}])$ and $\ddot{\mathbf{Q}}$ is the covariance matrix of the vectorised augmented tensor. We can also write this pdf as :

$$p_{\mathcal{X}}(\mathcal{X}) = \frac{\exp\{-\frac{1}{2}(\ddot{\mathcal{X}} - \ddot{\mathcal{M}})^* *_{N+1} \ddot{\mathcal{Q}}^{-1} *_{N+1} (\ddot{\mathcal{X}} - \ddot{\mathcal{M}})\}}{(\pi)^{I_1 I_2 \ldots I_N} (\det(\ddot{\mathcal{Q}}))^{1/2}} \tag{13}$$

Notationally, we write $\mathcal{X} \sim \mathcal{CN}(\mathcal{M}, \mathcal{Q}, \tilde{\mathcal{Q}})$. The equivalence of (12) and (13) can be directly established based on properties of Einstein product. For proper complex Gaussian tensor, with zero pseudo-covariance the pdf simplifies as :

$$p_{\mathcal{X}}(\mathcal{X}) = \frac{\exp\{-(\mathcal{X} - \mathcal{M})^* *_N \mathcal{Q}^{-1} *_N (\mathcal{X} - \mathcal{M})\}}{(\pi)^{I_1 I_2 \ldots I_N} \det(\mathcal{Q})} \tag{14}$$

where $\mathcal{M} = \mathbb{E}[\mathcal{X}]$ is the order N mean tensor and $\mathcal{Q} = \mathbb{E}[(\mathcal{X} - \mathcal{M}) \circ (\mathcal{X} - \mathcal{M})^*]$ is the order $2N$ covariance tensor.

III. MMSE Tensor Estimation

Consider the problem of estimating a complex tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \ldots \times I_N}$ from an observed complex tensor $\mathcal{Y} \in \mathbb{C}^{J_1 \times \ldots \times J_M}$. Throughout this section, we will assume the observed tensor and the tensor to be estimated have zero mean. We first establish an Orthogonality principle for tensors through the following theorem :

Theorem 1. *Let $g : \mathbb{C}^{J_1 \times \ldots \times J_M} \to \mathbb{C}^{I_1 \times \ldots \times I_N}$ be a tensor valued function of a tensor such that $\hat{\mathcal{X}} = g(\mathcal{Y}) \in \mathbb{C}^{I_1 \times \ldots \times I_N}$ is an estimator of tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \ldots \times I_N}$ based on the observation tensor $\mathcal{Y} \in \mathbb{C}^{J_1 \times \ldots \times J_M}$. We define the error tensor as $\mathcal{E} = \mathcal{X} - g(\mathcal{Y})$, then if :*

$$\mathbb{E}[\langle \mathcal{E}, h(\mathcal{Y})\rangle] = 0 \quad \textit{for } h : \mathbb{C}^{J_1 \times \ldots \times J_M} \to \mathbb{C}^{I_1 \times \ldots \times I_N} \tag{15}$$

then,

$$\mathbb{E}[||\mathcal{E}||^2] \leq \mathbb{E}[||\mathcal{X} - h(\mathcal{Y})||^2] \quad \textit{for } h : \mathbb{C}^{J_1 \times \ldots \times J_M} \to \mathbb{C}^{I_1 \times \ldots \times I_N} \tag{16}$$

Proof of Theorem 1 is provided in Appendix A. Using similar line of proof as for Theorem 1, the following corollaries can be established :

Corollary 1.1. Let $g, h : \mathbb{C}^{J_1 \times \ldots \times J_M} \to \mathbb{C}^{I_1 \times \ldots \times I_N}$ be tensor valued functions of tensors such that $g(\mathcal{Y}) = \mathcal{A}_1 *_M \mathcal{Y} + \mathcal{A}_2 *_M \mathcal{Y}^*$ and $h(\mathcal{Y}) = \mathcal{B}_1 *_M \mathcal{Y} + \mathcal{B}_2 *_M \mathcal{Y}^*$ where $\mathcal{A}_1, \mathcal{A}_2, \mathcal{B}_1, \mathcal{B}_2 \in \mathbb{C}^{I_1 \times \ldots I_N \times J_1 \times \ldots \times J_M}$. Let $g(\mathcal{Y})$ be an estimator of tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \ldots \times I_N}$ based on the observation tensor $\mathcal{Y} \in \mathbb{C}^{J_1 \times \ldots \times J_M}$. Then for the error tensor $\mathcal{E} = \mathcal{X} - g(\mathcal{Y})$, if,

$$\mathbb{E}[\langle \mathcal{E}, h(\mathcal{Y})\rangle] = 0 \quad \text{for } \mathcal{B}_1, \mathcal{B}_2 \in \mathbb{C}^{I_1 \times \ldots I_N \times J_1 \times \ldots \times J_M} \tag{17}$$

then,

$$\mathbb{E}[||\mathcal{E}||^2] \leq \mathbb{E}[||\mathcal{X} - h(\mathcal{Y})||^2] \quad \text{for } \mathcal{B}_1, \mathcal{B}_2 \in \mathbb{C}^{I_1 \times \ldots I_N \times J_1 \times \ldots \times J_M} \tag{18}$$

Corollary 1.2. Let $g, h : \mathbb{C}^{J_1 \times \ldots \times J_M} \to \mathbb{C}^{I_1 \times \ldots \times I_N}$ be tensor valued functions of tensors such that $g(\mathcal{Y}) = \mathcal{A} *_M \mathcal{Y}$ and $h(\mathcal{Y}) = \mathcal{B} *_M \mathcal{Y}$ where $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{I_1 \times \ldots I_N \times J_1 \times \ldots \times J_M}$. Let $g(\mathcal{Y})$ be an estimator of tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \ldots \times I_N}$ based on the observation tensor $\mathcal{Y} \in \mathbb{C}^{J_1 \times \ldots \times J_M}$. Then for the error tensor $\mathcal{E} = \mathcal{X} - g(\mathcal{Y})$, if

$$\mathbb{E}[\langle \mathcal{E}, h(\mathcal{Y})\rangle] = 0 \quad \text{for } \mathcal{B} \in \mathbb{C}^{I_1 \times \ldots I_N \times J_1 \times \ldots \times J_M} \tag{19}$$

then,

$$\mathbb{E}[||\mathcal{E}||^2] \leq \mathbb{E}[||\mathcal{X} - h(\mathcal{Y})||^2] \quad \text{for } \mathcal{B} \in \mathbb{C}^{I_1 \times \ldots I_N \times J_1 \times \ldots \times J_M} \tag{20}$$

A proof of Corollary 1.2 is also provided in [36].

A. Best MMSE Estimation of Tensors

The objective is to find the function $g(\mathcal{Y})$ which is the best estimator of $\mathcal{X}$ in mean square error sense. Let $h(\mathcal{Y})$ be any other estimator. Then we have :

$$\mathbb{E}[\langle \mathcal{X} - g(\mathcal{Y}), h(\mathcal{Y})\rangle] = \mathbb{E}_{\mathcal{Y}}\Big[\mathbb{E}[\langle \mathcal{X} - g(\mathcal{Y}), h(\mathcal{Y})\rangle \mid \mathcal{Y}]\Big] \tag{21}$$

$$= \mathbb{E}_{\mathcal{Y}}\Big[\mathbb{E}[(\mathcal{X} - g(\mathcal{Y})) *_N h(\mathcal{Y})^* \mid \mathcal{Y}]\Big] \tag{22}$$

Conditioned on $\mathcal{Y}$, we can take $h(\mathcal{Y})^*$ and $g(\mathcal{Y})$ outside of the inner expectation.

$$\mathbb{E}[\langle \mathcal{X} - g(\mathcal{Y}), h(\mathcal{Y})\rangle] = \mathbb{E}_{\mathcal{Y}}\Big[\mathbb{E}[(\mathcal{X} - g(\mathcal{Y})) \mid \mathcal{Y}] *_N h(\mathcal{Y})^*\Big] \tag{23}$$

$$= \mathbb{E}_{\mathcal{Y}}\Big[(\mathbb{E}[\mathcal{X} \mid \mathcal{Y}] - g(\mathcal{Y})) *_N h(\mathcal{Y})^*\Big] \tag{24}$$

From orthogonality principle, we know that (24) has to be 0 for any $h(\mathcal{Y})$ to minimize the mean square error. Thus $g(\mathcal{Y}) = \mathbb{E}[\mathcal{X} \mid \mathcal{Y}]$ is the best MMSE estimator of tensor $\mathcal{X}$ from $\mathcal{Y}$:

$$\hat{\mathcal{X}} = \mathbb{E}[\mathcal{X} \mid \mathcal{Y}] \tag{25}$$

Since $\mathbb{E}[\hat{\mathcal{X}}] = \mathbb{E}[\mathbb{E}[\mathcal{X} \mid \mathcal{Y}]] = \mathbb{E}[\mathcal{X}]$, the error $\mathcal{E} = \mathcal{X} - \hat{\mathcal{X}}$, is a zero mean tensor with the associated covariance as :

$$\begin{aligned}\mathcal{Q}_{\boldsymbol{\mathcal{E}}} &= \mathbb{E}[(\boldsymbol{\mathcal{X}} - \hat{\boldsymbol{\mathcal{X}}}) \circ (\boldsymbol{\mathcal{X}} - \hat{\boldsymbol{\mathcal{X}}})^*] &(26)\\ &= \mathbb{E}_{\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Y}}}[(\boldsymbol{\mathcal{X}} - \mathbb{E}[\boldsymbol{\mathcal{X}} \mid \boldsymbol{\mathcal{Y}}]) \circ (\boldsymbol{\mathcal{X}} - \mathbb{E}[\boldsymbol{\mathcal{X}} \mid \boldsymbol{\mathcal{Y}}])^*]. &(27)\end{aligned}$$

Note that (27) represents the order $2N$ covariance tensor of the error when conditional mean estimator is used to estimate an order N tensor. The trace of such a covariance tensor gives us the mean square error in estimation as :

$$\begin{aligned}\mathbb{E}[||\boldsymbol{\mathcal{E}}||^2] &= \mathbb{E}\Big[\sum_{i_1,\ldots,i_N} |\boldsymbol{\mathcal{E}}_{i_1,\ldots,i_N}|^2 \Big]\\ &= \sum_{i_1,\ldots,i_N} \mathbb{E}[|\boldsymbol{\mathcal{E}}_{i_1,\ldots,i_N}|^2] = \mathrm{tr}(\mathcal{Q}_{\boldsymbol{\mathcal{E}}}) \quad (28)\end{aligned}$$

Notice that similar to vector case [35], conditioning on $\boldsymbol{\mathcal{Y}}$ is same as conditioning on $\boldsymbol{\mathcal{Y}}^*$, i.e. $\mathbb{E}[\boldsymbol{\mathcal{X}} \mid \boldsymbol{\mathcal{Y}}] = \mathbb{E}[\boldsymbol{\mathcal{X}} \mid \boldsymbol{\mathcal{Y}}^*]$. Hence the estimator conditioned on both $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\mathcal{Y}}^*$ gives no extra information as compared to the one conditioned only on $\boldsymbol{\mathcal{Y}}$. Also the conditional mean estimator of $\boldsymbol{\mathcal{X}}^*$ would just be the complex conjugate of the conditional mean estimator of $\boldsymbol{\mathcal{X}}$.

B. Widely Multi-linear and Multi-linear MMSE Estimation of Tensors

A linear relationship between complex scalars $x = x_r + jx_i$ and $y = y_r + jy_i$, as $y = kx$ implies both the real and imaginary parts of y are related with the real and imaginary parts of x with the same coefficient k, i.e. $y_r = kx_r$ and $y_i = kx_i$. However, in a more general case a linear-conjugate-linear or widely linear relation is defined as $y = k_1x + k_2x^*$ [42] where both the real and imaginary parts of y are related with real and imaginary parts of x using different coefficients, i.e. $y_r = (k_1 + k_2)x_r$ and $y_i = (k_1 - k_2)x_i$. This notion was used to define the widely linear MMSE estimate of a complex vector in [35], [45] where the estimate of a vector depends widely linearly on the observed vector, or in other words depends linearly on both the observed vector and its conjugate through different coefficients. For complex data estimation, assuming a widely linear dependence instead of linear can significantly improve the performance since the former takes into account covariance and pseudo-covariance both, while the latter employs only the covariance [35].

In this section, we restrict ourselves to the class of estimators which depend linearly on the received tensor $\boldsymbol{\mathcal{Y}}$ and its conjugate $\boldsymbol{\mathcal{Y}}^*$. If the estimator depends linearly only on tensor $\boldsymbol{\mathcal{Y}}$, it is called a multi-linear estimate, whereas if the estimator depends linearly on both $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\mathcal{Y}}^*$ through different set of coefficients, it is called a widely multi-linear estimate. In order to estimate the tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1\times\ldots\times I_N}$ from an observed complex tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{J_1\times\ldots\times J_M}$ and its conjugate by a multi-linear structure, we are looking for the tensors $\mathcal{A}_1, \mathcal{A}_2 \in \mathbb{C}^{I_1\times\ldots I_N\times J_1\times\ldots\times J_M}$ such that the estimator:

$$\hat{\boldsymbol{\mathcal{X}}}_{WL} = \mathcal{A}_1 *_M \boldsymbol{\mathcal{Y}} + \mathcal{A}_2 *_M \boldsymbol{\mathcal{Y}}^* \quad (29)$$

satisfies

$$\begin{aligned}&\mathbb{E}[||\boldsymbol{\mathcal{X}} - (\mathcal{A}_1 *_M \boldsymbol{\mathcal{Y}} + \mathcal{A}_2 *_M \boldsymbol{\mathcal{Y}}^*)||^2] \le\\ &\qquad \mathbb{E}[||\boldsymbol{\mathcal{X}} - (\mathcal{B}_1 *_M \boldsymbol{\mathcal{Y}} + \mathcal{B}_2 *_M \boldsymbol{\mathcal{Y}}^*)||^2] \quad (30)\end{aligned}$$

for any other tensors $\mathcal{B}_1, \mathcal{B}_2 \in \mathbb{C}^{I_1\times\ldots\times I_N\times J_1\times\ldots\times J_M}$. From Corollary 1.1, we know that the optimal $\mathcal{A}_1$ and $\mathcal{A}_2$ will be such that :

$$\mathbb{E}[\langle(\boldsymbol{\mathcal{X}} - \hat{\boldsymbol{\mathcal{X}}}_{WL}), (\mathcal{B}_1 *_M \boldsymbol{\mathcal{Y}} + \mathcal{B}_2 *_M \boldsymbol{\mathcal{Y}}^*)\rangle] = 0 \quad (31)$$

for any choice of $\mathcal{B}_1, \mathcal{B}_2 \in \mathbb{C}^{I_1\times\ldots\times I_N\times J_1\times\ldots\times J_M}$. From (2) and (29) , we can write (31) as :

$$\mathbb{E}[(\boldsymbol{\mathcal{X}} - \mathcal{A}_1 *_M \boldsymbol{\mathcal{Y}} - \mathcal{A}_2 *_M \boldsymbol{\mathcal{Y}}^*) *_N (\mathcal{B}_1 *_M \boldsymbol{\mathcal{Y}} + \mathcal{B}_2 *_M \boldsymbol{\mathcal{Y}}^*)^*] = 0 \quad (32)$$

From (6), we can write $(\mathcal{B}_1 *_M \boldsymbol{\mathcal{Y}})^* = (\boldsymbol{\mathcal{Y}}^* *_M \mathcal{B}_1^H)$ and $(\mathcal{B}_2 *_M \boldsymbol{\mathcal{Y}}^*)^* = (\boldsymbol{\mathcal{Y}} *_M \mathcal{B}_2^H)$, and hence the left hand side of (32) can be written as :

$$\begin{aligned}&\mathbb{E}[(\boldsymbol{\mathcal{X}} - \mathcal{A}_1 *_M \boldsymbol{\mathcal{Y}} - \mathcal{A}_2 *_M \boldsymbol{\mathcal{Y}}^*) *_N ((\boldsymbol{\mathcal{Y}}^* *_M \mathcal{B}_1^H) + (\boldsymbol{\mathcal{Y}} *_M \mathcal{B}_2^H))]\\
&= \mathbb{E}[\,\mathrm{tr}\,\{(\boldsymbol{\mathcal{X}} - \mathcal{A}_1 *_M \boldsymbol{\mathcal{Y}} - \mathcal{A}_2 *_M \boldsymbol{\mathcal{Y}}^*)\circ\\
&\quad ((\boldsymbol{\mathcal{Y}}^* *_M \mathcal{B}_1^H) + (\boldsymbol{\mathcal{Y}} *_M \mathcal{B}_2^H))\}] \quad \text{(from (9))} &(33)\\
&= \mathbb{E}[\,\mathrm{tr}\,\{\boldsymbol{\mathcal{X}} \circ \boldsymbol{\mathcal{Y}}^* *_M \mathcal{B}_1^H - \mathcal{A}_1 *_M \boldsymbol{\mathcal{Y}} \circ \boldsymbol{\mathcal{Y}}^* *_M \mathcal{B}_1^H -\\
&\quad \mathcal{A}_2 *_M \boldsymbol{\mathcal{Y}}^* \circ \boldsymbol{\mathcal{Y}}^* *_M \mathcal{B}_1^H + \boldsymbol{\mathcal{X}} \circ \boldsymbol{\mathcal{Y}} *_M \mathcal{B}_2^H -\\
&\quad \mathcal{A}_1 *_M \boldsymbol{\mathcal{Y}} \circ \boldsymbol{\mathcal{Y}} *_M \mathcal{B}_2^H - \mathcal{A}_2 *_M \boldsymbol{\mathcal{Y}}^* \circ \boldsymbol{\mathcal{Y}} *_M \mathcal{B}_2^H\}] &(34)\\
&= \mathrm{tr}\,\{\underbrace{\mathbb{E}[\boldsymbol{\mathcal{X}} \circ \boldsymbol{\mathcal{Y}}^*]}_{\mathcal{C}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}}} *_M \mathcal{B}_1^H - \mathcal{A}_1 *_M \underbrace{\mathbb{E}[\boldsymbol{\mathcal{Y}} \circ \boldsymbol{\mathcal{Y}}^*]}_{\mathcal{C}_{\boldsymbol{\mathcal{Y}}}} *_M \mathcal{B}_1^H -\\
&\quad \mathcal{A}_2 *_M \underbrace{\mathbb{E}[\boldsymbol{\mathcal{Y}}^* \circ \boldsymbol{\mathcal{Y}}^*]}_{\tilde{\mathcal{C}}_{\boldsymbol{\mathcal{Y}}}^*} *_M \mathcal{B}_1^H + \underbrace{\mathbb{E}[\boldsymbol{\mathcal{X}} \circ \boldsymbol{\mathcal{Y}}]}_{\tilde{\mathcal{C}}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}}} *_M \mathcal{B}_2^H -\\
&\quad \mathcal{A}_1 *_M \underbrace{\mathbb{E}[\boldsymbol{\mathcal{Y}} \circ \boldsymbol{\mathcal{Y}}]}_{\tilde{\mathcal{C}}_{\boldsymbol{\mathcal{Y}}}} *_M \mathcal{B}_2^H - \mathcal{A}_2 *_M \underbrace{\mathbb{E}[\boldsymbol{\mathcal{Y}}^* \circ \boldsymbol{\mathcal{Y}}]}_{\mathcal{C}_{\boldsymbol{\mathcal{Y}}}^*} *_M \mathcal{B}_2^H\} &(35)\\
&= \mathrm{tr}\,\{\underbrace{(\mathcal{C}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}} - \mathcal{A}_1 *_M \mathcal{C}_{\boldsymbol{\mathcal{Y}}} - \mathcal{A}_2 *_M \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{Y}}}^*)}_{\bar{\mathcal{B}}_1} *_M \mathcal{B}_1^H +\\
&\quad \underbrace{(\tilde{\mathcal{C}}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}} - \mathcal{A}_1 *_M \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{Y}}} - \mathcal{A}_2 *_M \mathcal{C}_{\boldsymbol{\mathcal{Y}}}^*)}_{\bar{\mathcal{B}}_2} *_M \mathcal{B}_2^H\} &(36)\end{aligned}$$

For (36) to be 0, for any $\mathcal{B}_1$ and $\mathcal{B}_2$, we need $\bar{\mathcal{B}}_1$ and $\bar{\mathcal{B}}_2$ to be all zero tensors, which gives us the conditions for optimal $\mathcal{A}_1$ and $\mathcal{A}_2$:

$$\mathcal{C}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}} = \mathcal{A}_1 *_M \mathcal{C}_{\boldsymbol{\mathcal{Y}}} + \mathcal{A}_2 *_M \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{Y}}}^* \quad (37)$$

$$\tilde{\mathcal{C}}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}} = \mathcal{A}_1 *_M \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{Y}}} + \mathcal{A}_2 *_M \mathcal{C}_{\boldsymbol{\mathcal{Y}}}^* \quad (38)$$

Equations (37) and (38) are systems of multi-linear equations which can be solved for $\mathcal{A}_1$ and $\mathcal{A}_1$ using methods described in [29]. If the inverse of the covariance of $\boldsymbol{\mathcal{Y}}$ exists, then from (38), we get $\mathcal{A}_2 = (\tilde{\mathcal{C}}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}} - \mathcal{A}_1 *_M \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{Y}}}) *_M \mathcal{C}_{\boldsymbol{\mathcal{Y}}}^{*-1}$ which we can substitute in (37) to get :

$$\begin{aligned}\mathcal{C}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}} &= \mathcal{A}_1 *_M \mathcal{C}_{\boldsymbol{\mathcal{Y}}} + (\tilde{\mathcal{C}}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}} - \mathcal{A}_1 *_M \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{Y}}}) *_M \mathcal{C}_{\boldsymbol{\mathcal{Y}}}^{*-1} *_M \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{Y}}}^*\\
&= \mathcal{A}_1 *_M (\mathcal{C}_{\boldsymbol{\mathcal{Y}}} - \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{Y}}} *_M \mathcal{C}_{\boldsymbol{\mathcal{Y}}}^{*-1} *_M \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{Y}}}^*) + (\tilde{\mathcal{C}}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}} *_M \mathcal{C}_{\boldsymbol{\mathcal{Y}}}^{*-1} *_M \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{Y}}}^*)\\
\Rightarrow \mathcal{A}_1 &= \Big(\mathcal{C}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}} - \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}} *_M \mathcal{C}_{\boldsymbol{\mathcal{Y}}}^{*-1} *_M \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{Y}}}^*\Big) *_M \mathcal{P}_{\boldsymbol{\mathcal{Y}}}^{-1} &(39)\\
\text{where,} \quad \mathcal{P}_{\boldsymbol{\mathcal{Y}}} &= (\mathcal{C}_{\boldsymbol{\mathcal{Y}}} - \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{Y}}} *_M \mathcal{C}_{\boldsymbol{\mathcal{Y}}}^{*-1} *_M \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{Y}}}^*). &(40)\end{aligned}$$

Also from (37) we get $\mathcal{A}_1 = (\mathcal{C}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}} - \mathcal{A}_2 *_M \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{Y}}}^*) *_M \mathcal{C}_{\boldsymbol{\mathcal{Y}}}^{-1}$ which we can substitute in (38) to get :

$$\mathcal{A}_2 = \Big(\tilde{\mathcal{C}}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}} - \mathcal{C}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}} *_M \mathcal{C}_{\boldsymbol{\mathcal{Y}}}^{-1} *_M \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{Y}}}\Big) *_M \mathcal{P}_{\boldsymbol{\mathcal{Y}}}^{*-1} \quad (41)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2021.3084541, IEEE Open Journal of Signal Processing

Substituting (39) and (41) into (29) gives us the widely multi-linear estimate $\hat{\mathcal{X}}_{WL}$ of the tensor $\mathcal{X}$. The covariance of the corresponding error tensor is :

$$\begin{aligned}\mathcal{Q}_{WL} &= \mathbb{E}[(\mathcal{X}-\hat{\mathcal{X}}_{WL})\circ(\mathcal{X}-\hat{\mathcal{X}}_{WL})^*] \\ &= \mathcal{C}_{\mathcal{X}} - \mathcal{A}_1 *_M \mathcal{C}^H_{\mathcal{X}\mathcal{Y}} - \mathcal{A}_2 *_M \tilde{\mathcal{C}}^H_{\mathcal{X}\mathcal{Y}} \end{aligned} \tag{42}$$

A detailed derivation of (42) has been included in Appendix B. On substituting $\mathcal{A}_1$ and $\mathcal{A}_2$ from (39) and (41) respectively, we get :

$$\begin{aligned}\mathcal{Q}_{WL} = \mathcal{C}_{\mathcal{X}} - &\left(\mathcal{C}_{\mathcal{X}\mathcal{Y}} - \tilde{\mathcal{C}}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}^{*-1}_{\mathcal{Y}} *_M \tilde{\mathcal{C}}^*_{\mathcal{Y}}\right) *_M \mathcal{P}^{-1}_{\mathcal{Y}} *_M \mathcal{C}^H_{\mathcal{X}\mathcal{Y}} - \\ &\left(\tilde{\mathcal{C}}_{\mathcal{X}\mathcal{Y}} - \mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}^{-1}_{\mathcal{Y}} *_M \tilde{\mathcal{C}}_{\mathcal{Y}}\right) *_M \mathcal{P}^{*-1}_{\mathcal{Y}} *_M \tilde{\mathcal{C}}^H_{\mathcal{X}\mathcal{Y}}\end{aligned} \tag{43}$$

The quantity $\mathcal{A}_1 *_M \mathcal{C}^H_{\mathcal{X}\mathcal{Y}} + \mathcal{A}_2 *_M \tilde{\mathcal{C}}^H_{\mathcal{X}\mathcal{Y}}$ can be seen as the cross covariance between the widely multi-linear estimator $\hat{\mathcal{X}}_{WL}$ and the tensor $\mathcal{X}$ (based on (94)). Hence, intuitively the error covariance tensor is the difference between the covariance tensor of $\mathcal{X}$ and the cross covariance between $\hat{\mathcal{X}}_{WL}$ and $\mathcal{X}$, i.e. $\mathcal{Q}_{WL} = \mathcal{C}_{\mathcal{X}} - \mathcal{C}_{\hat{\mathcal{X}}_{WL}\mathcal{X}}$.

The corresponding mean square error is given by

$$\begin{aligned}\text{MSE}_{WL} &= \text{tr}(\mathcal{Q}_{WL}) = \text{tr}(\mathcal{C}_{\mathcal{X}} - \mathcal{A}_1 *_M \mathcal{C}^H_{\mathcal{X}\mathcal{Y}} - \mathcal{A}_2 *_M \tilde{\mathcal{C}}^H_{\mathcal{X}\mathcal{Y}}) \\ &= \text{tr}(\mathcal{C}_{\mathcal{X}} - (\mathcal{C}_{\mathcal{X}\mathcal{Y}} - \tilde{\mathcal{C}}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}^{*-1}_{\mathcal{Y}} *_M \tilde{\mathcal{C}}^*_{\mathcal{Y}}) *_M \mathcal{P}^{-1}_{\mathcal{Y}} *_M \mathcal{C}^H_{\mathcal{X}\mathcal{Y}} - \\ &\quad (\tilde{\mathcal{C}}_{\mathcal{X}\mathcal{Y}} - \mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}^{-1}_{\mathcal{Y}} *_M \tilde{\mathcal{C}}_{\mathcal{Y}}) *_M \mathcal{P}^{*-1}_{\mathcal{Y}} *_M \tilde{\mathcal{C}}^H_{\mathcal{X}\mathcal{Y}})\end{aligned} \tag{44}$$

Next we consider the problem of MMSE estimation where we assume that the estimate of $\mathcal{X}$ depends linearly only on the received tensor $\mathcal{Y}$ and not its conjugate, i.e.

$$\hat{\mathcal{X}}_L = \mathcal{A} *_M \mathcal{Y} \tag{45}$$

and we want to find the tensor $\mathcal{A} \in \mathbb{C}^{I_1\times\ldots\times I_N\times J_1\times\ldots\times J_M}$ such that $\mathbb{E}[||\mathcal{X} - \mathcal{A} *_M \mathcal{Y}||^2] \leq \mathbb{E}[||\mathcal{X} - \mathcal{B} *_M \mathcal{Y}||^2]$ for any other tensor $\mathcal{B} \in \mathbb{C}^{I_1\times\ldots\times I_N\times J_1\times\ldots\times J_M}$. Using corollary 1.2, we know that the optimal $\mathcal{A}$ will satisfy $\mathbb{E}[(\mathcal{X} - \mathcal{A} *_M \mathcal{Y}) *_N (\mathcal{B} *_M \mathcal{Y})] = 0$. Now using the same line of proof as for (36) by substituting $\mathcal{A}_2 = 0_{\mathcal{T}}$ and $\mathcal{A}_1 = \mathcal{A}$ we can get the condition for optimal $\mathcal{A}$ as :

$$\mathcal{C}_{\mathcal{X}\mathcal{Y}} = \mathcal{A} *_M \mathcal{C}_{\mathcal{Y}} \tag{46}$$

Equation (46) can be solved using methods described in [29] for $\mathcal{A}$. If the inverse of $\mathcal{C}_{\mathcal{Y}}$ does not exist, a minimum norm least square solution is suggested in [46] leading to Moore-Penrose inverse of tensors. If the inverse of $\mathcal{C}_{\mathcal{Y}}$ exists, then we have

$$\mathcal{A} = \mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}^{-1}_{\mathcal{Y}} \tag{47}$$

and the multi-linear MMSE estimate of $\mathcal{X}$ is given by :

$$\hat{\mathcal{X}}_L = (\mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}^{-1}_{\mathcal{Y}}) *_M \mathcal{Y} \tag{48}$$

The covariance of the corresponding error tensor is :

$$\mathcal{Q}_L = \mathbb{E}[(\mathcal{X}-\hat{\mathcal{X}}_L)\circ(\mathcal{X}-\hat{\mathcal{X}}_L)^*] \tag{49}$$

which we can solve by substituting $\mathcal{A}_2 = 0_{\mathcal{T}}$ and $\mathcal{A}_1 = (\mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}^{-1}_{\mathcal{Y}})$ in (44), to get:

$$\mathcal{Q}_L = \mathcal{C}_{\mathcal{X}} - \mathcal{A} *_M \mathcal{C}^H_{\mathcal{X}\mathcal{Y}} = \mathcal{C}_{\mathcal{X}} - \mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}^{-1}_{\mathcal{Y}} *_M \mathcal{C}^H_{\mathcal{X}\mathcal{Y}} \tag{50}$$

The quantity $\mathcal{A} *_M \mathcal{C}^H_{\mathcal{X}\mathcal{Y}}$ can be seen as the cross covariance between $\hat{\mathcal{X}}_L$ and $\mathcal{X}$. Hence, intuitively the error covariance tensor is the difference between the covariance tensor of $\mathcal{X}$ and the cross covariance between $\hat{\mathcal{X}}_L$ and $\mathcal{X}$, i.e. $\mathcal{Q}_L = \mathcal{C}_{\mathcal{X}} - \mathcal{C}_{\hat{\mathcal{X}}_L\mathcal{X}}$. The mean square error is given as:

$$\text{MSE}_L = \text{tr}(\mathcal{Q}_L) = \text{tr}(\mathcal{C}_{\mathcal{X}} - \mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}^{-1}_{\mathcal{Y}} *_M \mathcal{C}^H_{\mathcal{X}\mathcal{Y}}) \tag{51}$$

C. Comparing Multi-linear and Widely Multi-linear MMSE Estimation

Let $g(\mathcal{Y}) = \mathcal{A}_1 *_M \mathcal{Y} + \mathcal{A}_2 *_M \mathcal{Y}^*$ be a widely multi-linear estimator of $\mathcal{X}$ based on tensor $\mathcal{Y}$ and its conjugate $\mathcal{Y}^*$. From corollary 1.1, we know that the mean square error achieved by the widely multi-linear estimate $g(\mathcal{Y})$ with $\mathcal{A}_1$ and $\mathcal{A}_2$ given in (39) and (41) respectively will be less than or equal to the mean square error achieved by any other choice of $\mathcal{A}_1$ and $\mathcal{A}_2$. An alternate choice of $\mathcal{A}_1$ and $\mathcal{A}_2$ can be $\mathcal{A}_1 = \mathcal{A}$ (obtained from (47)) and $\mathcal{A}_2 = 0_{\mathcal{T}}$, in which case $g(\mathcal{Y})$ will represent the multi-linear estimate. Hence the mean square error achieved by a widely multi-linear estimate given by (44) is always less than or equal to the mean square error achieved by the multi-linear estimate given by (51). The performance difference between the two cases can be found by comparing the error covariance tensors. On substituting $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ from (37) into (50), we get :

$$\begin{aligned}\mathcal{Q}_L &= \mathcal{C}_{\mathcal{X}} - (\mathcal{A}_1 *_M \mathcal{C}_{\mathcal{Y}} + \mathcal{A}_2 *_M \tilde{\mathcal{C}}^*_{\mathcal{Y}}) \\ &\qquad *_M \mathcal{C}^{-1}_{\mathcal{Y}} *_M (\mathcal{A}_1 *_M \mathcal{C}_{\mathcal{Y}} + \mathcal{A}_2 *_M \tilde{\mathcal{C}}^*_{\mathcal{Y}})^H \\ &= \mathcal{C}_{\mathcal{X}} - (\mathcal{A}_1 + \mathcal{A}_2 *_M \tilde{\mathcal{C}}^*_{\mathcal{Y}} *_M \mathcal{C}^{-1}_{\mathcal{Y}}) *_M (\mathcal{C}_{\mathcal{Y}} *_M \mathcal{A}^H_1 + \tilde{\mathcal{C}}_{\mathcal{Y}} *_M \mathcal{A}^H_2) \\ &= \mathcal{C}_{\mathcal{X}} - \mathcal{A}_1 *_M \mathcal{C}_{\mathcal{Y}} *_M \mathcal{A}^H_1 - \mathcal{A}_1 *_M \tilde{\mathcal{C}}_{\mathcal{Y}} *_M \mathcal{A}^H_2 \\ &\quad - \mathcal{A}_2 *_M \tilde{\mathcal{C}}^*_{\mathcal{Y}} *_M \mathcal{A}^H_1 - \mathcal{A}_2 *_M \tilde{\mathcal{C}}^*_{\mathcal{Y}} *_M \mathcal{C}^{-1}_{\mathcal{Y}} *_M \tilde{\mathcal{C}}_{\mathcal{Y}} *_M \mathcal{A}^H_2\end{aligned} \tag{52}$$

Note that we have used the properties $\mathcal{C}_{\mathcal{Y}} = \mathcal{C}^H_{\mathcal{Y}}$ and $\tilde{\mathcal{C}}_{\mathcal{Y}} = \tilde{\mathcal{C}}^T_{\mathcal{Y}}$ in the derivation. Similarly on substituting $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ from (37) and $\tilde{\mathcal{C}}_{\mathcal{X}\mathcal{Y}}$ from (38) into (42), we get

$$\begin{aligned}\mathcal{Q}_{WL} &= \mathcal{C}_{\mathcal{X}} - \mathcal{A}_1 *_M (\mathcal{A}_1 *_M \mathcal{C}_{\mathcal{Y}} + \mathcal{A}_2 *_M \tilde{\mathcal{C}}^*_{\mathcal{Y}})^H \\ &\qquad - \mathcal{A}_2 *_M (\mathcal{A}_1 *_M \tilde{\mathcal{C}}_{\mathcal{Y}} + \mathcal{A}_2 *_M \mathcal{C}^*_{\mathcal{Y}})^H \\ &= \mathcal{C}_{\mathcal{X}} - \mathcal{A}_1 *_M \mathcal{C}_{\mathcal{Y}} *_M \mathcal{A}^H_1 - \mathcal{A}_1 *_M \tilde{\mathcal{C}}_{\mathcal{Y}} *_M \mathcal{A}^H_2 \\ &\qquad - \mathcal{A}_2 *_M \tilde{\mathcal{C}}^*_{\mathcal{Y}} *_M \mathcal{A}^H_1 - \mathcal{A}_2 *_M \mathcal{C}^*_{\mathcal{Y}} *_M \mathcal{A}^H_2\end{aligned} \tag{53}$$

Subtracting (53) from (52), we get

$$\begin{aligned}\Delta_{\mathcal{Q}} &= \mathcal{Q}_L - \mathcal{Q}_{WL} = \mathcal{A}_2 *_M (\mathcal{C}^*_{\mathcal{Y}} - \tilde{\mathcal{C}}^*_{\mathcal{Y}} *_M \mathcal{C}^{-1}_{\mathcal{Y}} *_M \tilde{\mathcal{C}}_{\mathcal{Y}}) *_M \mathcal{A}^H_2 \\ &= \mathcal{A}_2 * \mathcal{P}^*_{\mathcal{Y}} *_M \mathcal{A}^H_2 \qquad \text{(from (40))} \\ &= (\tilde{\mathcal{C}}_{\mathcal{X}\mathcal{Y}} - \mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}^{-1}_{\mathcal{Y}} *_M \tilde{\mathcal{C}}_{\mathcal{Y}}) *_M \mathcal{P}^{*-1}_{\mathcal{Y}} *_M \\ &\quad (\tilde{\mathcal{C}}_{\mathcal{X}\mathcal{Y}} - \mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}^{-1}_{\mathcal{Y}} *_M \tilde{\mathcal{C}}_{\mathcal{Y}})^H \quad \text{(based on (41))}\end{aligned} \tag{54}$$

$$\Delta_e = \text{MSE}_L - \text{MSE}_{WL} = \text{tr}(\mathcal{Q}_L - \mathcal{Q}_{WL}) = \text{tr}(\Delta_{\mathcal{Q}}) \tag{55}$$

Since $\mathcal{P}^{*-1}_{\mathcal{Y}}$ is a positive definite tensor, hence the pseudo-diagonal entries of $\Delta_{\mathcal{Q}}$ are always non negative. Hence Δ_e is always non negative. The condition when the two estimates are identical and hence provide the same mean square error, i.e. $\Delta_e = 0$, is formally established in the following Lemma.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2021.3084541, IEEE Open Journal of Signal Processing

Lemma 1. *The widely multi-linear and multi-linear MMSE estimates are identical when :*

$$\tilde{\mathcal{C}}_{\mathcal{X}\mathcal{Y}} = \mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}_{\mathcal{Y}}^{-1} *_M \tilde{\mathcal{C}}_{\mathcal{Y}} \tag{56}$$

Proof. Substituting (56) in (39), we get :

$$\begin{aligned}
\mathcal{A}_1 &= (\mathcal{C}_{\mathcal{X}\mathcal{Y}} - \mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}_{\mathcal{Y}}^{-1} *_M \tilde{\mathcal{C}}_{\mathcal{Y}} *_M \mathcal{C}_{\mathcal{Y}}^{*-1} *_M \tilde{\mathcal{C}}_{\mathcal{Y}}^{*}) *_M \\
&\quad (\mathcal{C}_{\mathcal{Y}} - \tilde{\mathcal{C}}_{\mathcal{Y}} *_M \mathcal{C}_{\mathcal{Y}}^{*-1} *_M \tilde{\mathcal{C}}_{\mathcal{Y}}^{*})^{-1} \\
&= \mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M (\mathcal{I} - \mathcal{C}_{\mathcal{Y}}^{-1} *_M \tilde{\mathcal{C}}_{\mathcal{Y}} *_M \mathcal{C}_{\mathcal{Y}}^{*-1} *_M \tilde{\mathcal{C}}_{\mathcal{Y}}^{*}) *_M \\
&\quad (\mathcal{C}_{\mathcal{Y}} *_M (\mathcal{I} - \mathcal{C}_{\mathcal{Y}}^{-1} *_M \tilde{\mathcal{C}}_{\mathcal{Y}} *_M \mathcal{C}_{\mathcal{Y}}^{*-1} *_M \tilde{\mathcal{C}}_{\mathcal{Y}}^{*}))^{-1} \\
&= \mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M (\mathcal{I} - \mathcal{C}_{\mathcal{Y}}^{-1} *_M \tilde{\mathcal{C}}_{\mathcal{Y}} *_M \mathcal{C}_{\mathcal{Y}}^{*-1} *_M \tilde{\mathcal{C}}_{\mathcal{Y}}^{*}) *_M \\
&\quad (\mathcal{I} - \mathcal{C}_{\mathcal{Y}}^{-1} *_M \tilde{\mathcal{C}}_{\mathcal{Y}} *_M \mathcal{C}_{\mathcal{Y}}^{*-1} *_M \tilde{\mathcal{C}}_{\mathcal{Y}}^{*})^{-1} *_M \mathcal{C}_{\mathcal{Y}}^{-1} \\
&= \mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}_{\mathcal{Y}}^{-1}
\end{aligned} \tag{57}$$

Also, substituting (56) in (41), we get $\mathcal{A}_2 = 0_{\mathcal{T}}$. In this case, the widely multi-linear estimate is given as $(\mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}_{\mathcal{Y}}^{-1}) *_M \mathcal{Y} + 0_{\mathcal{T}} *_M \mathcal{Y}^*$, which is the multi-linear estimate from (48). □

Lemma 1 essentially represents a condition when the error of the multi-linear estimate $(\hat{\mathcal{X}}_L - \mathcal{X})$ is uncorrelated with $\mathcal{Y}^*$, i.e. $\mathbb{E}[(\hat{\mathcal{X}}_L - \mathcal{X}) \circ (\mathcal{Y}^*)^*] = 0_{\mathcal{T}}$. Substituting (48) in this condition we get :

$$\mathbb{E}[((\mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}_{\mathcal{Y}}^{-1}) *_M \mathcal{Y} - \mathcal{X}) \circ (\mathcal{Y}^*)^*] = 0_{\mathcal{T}} \tag{58}$$

$$\Rightarrow (\mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}_{\mathcal{Y}}^{-1}) *_M \mathbb{E}[\mathcal{Y} \circ \mathcal{Y}] - \mathbb{E}[\mathcal{X} \circ \mathcal{Y}] = 0_{\mathcal{T}} \tag{59}$$

$$\Rightarrow \mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}_{\mathcal{Y}}^{-1} *_M \tilde{\mathcal{C}}_{\mathcal{Y}} - \tilde{\mathcal{C}}_{\mathcal{X}\mathcal{Y}} = 0_{\mathcal{T}} \tag{60}$$

which is same as (56). One of the trivial cases when this will be satisfied is when the tensor to be estimated and the observation are jointly proper. Tensors $\mathcal{X}$ and $\mathcal{Y}$ are called jointly proper if they are both individually proper, i.e. $\tilde{\mathcal{C}}_{\mathcal{X}} = \tilde{\mathcal{C}}_{\mathcal{Y}} = 0_{\mathcal{T}}$ and cross-proper, i.e. $\tilde{\mathcal{C}}_{\mathcal{X}\mathcal{Y}} = 0_{\mathcal{T}}$. In this case, (56) is satisfied, hence multi-linear estimate is same as widely multi-linear estimate. Notice however, that $\mathcal{X}$ and $\mathcal{Y}$ being jointly proper is not a necessary condition, as even if $\mathcal{X}$ is not proper, i.e. $\tilde{\mathcal{C}}_{\mathcal{X}} \neq 0_{\mathcal{T}}$ but $\tilde{\mathcal{C}}_{\mathcal{Y}} = 0_{\mathcal{T}}$ and $\tilde{\mathcal{C}}_{\mathcal{X}\mathcal{Y}} = 0_{\mathcal{T}}$, still (56) is satisfied and multi-linear and widely multi-linear estimate will be same.

Further, if the tensor to be estimated $\mathcal{X}$ is real, the cross pseudo-covariance is given as :

$$\tilde{\mathcal{C}}_{\mathcal{X}\mathcal{Y}} = \mathbb{E}[\mathcal{X} \circ \mathcal{Y}] = \mathbb{E}[(\mathcal{X} \circ \mathcal{Y}^*)^*] = \mathcal{C}_{\mathcal{X}\mathcal{Y}}^* \quad (\text{since } \mathcal{X} = \mathcal{X}^*) \tag{61}$$

Substituting (61) into (37) and comparing with (38) shows that $\mathcal{A}_2 = \mathcal{A}_1^*$. Thus for real $\mathcal{X}$, the widely multi-linear MMSE estimate and the associated mean square error (from (44)) are given as :

$$\begin{aligned}
\hat{\mathcal{X}}_{WL} &= \mathcal{A}_1 *_M \mathcal{Y} + \mathcal{A}_1^* *_M \mathcal{Y}^* = 2\Re(\mathcal{A}_1 *_M \mathcal{Y}) \\
&= 2\Re((\mathcal{C}_{\mathcal{X}\mathcal{Y}} - \tilde{\mathcal{C}}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}_{\mathcal{Y}}^{*-1} *_M \tilde{\mathcal{C}}_{\mathcal{Y}}^{*}) *_M \mathcal{P}_{\mathcal{Y}}^{-1} *_M \mathcal{Y})
\end{aligned} \tag{62}$$

which shows that the widely multi-linear estimate of a real signal is always real irrespective of the observation being complex. The corresponding mean square error is given as:

$$\begin{aligned}
\mathrm{MSE}_{WL} &= \mathrm{tr}(\mathcal{C}_{\mathcal{X}} - \mathcal{A}_1 *_M \mathcal{C}_{\mathcal{X}\mathcal{Y}}^{H} - \mathcal{A}_1^* *_M \mathcal{C}_{\mathcal{X}\mathcal{Y}}^{*H}) \\
&= \mathrm{tr}(\mathcal{C}_{\mathcal{X}} - 2\Re(\mathcal{A}_1 *_M \mathcal{C}_{\mathcal{X}\mathcal{Y}}^{H})) \\
&= \mathrm{tr}(\mathcal{C}_{\mathcal{X}} - 2\Re((\mathcal{C}_{\mathcal{X}\mathcal{Y}} - \tilde{\mathcal{C}}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{C}_{\mathcal{Y}}^{*-1} *_M \tilde{\mathcal{C}}_{\mathcal{Y}}^{*}) *_M \mathcal{P}_{\mathcal{Y}}^{-1} *_M \mathcal{C}_{\mathcal{X}\mathcal{Y}}^{H}))
\end{aligned} \tag{63}$$

For real or proper complex vectors, it is well known that if the signal to be estimated $\underline{\mathbf{x}}$ and the observed vector $\underline{\mathbf{y}}$ are jointly Gaussian random vectors, then the best MMSE estimate is same as the LMMSE estimate. However, to be more accurate, we can say that if $\underline{\mathbf{x}}$ and $\underline{\mathbf{y}}$ are two jointly complex Gaussian random vectors, then the best MMSE estimator $\mathbb{E}[\underline{\mathbf{x}}|\underline{\mathbf{y}}]$ is the WLMMSE estimator of $\underline{\mathbf{x}}$ from $\underline{\mathbf{y}}$ [42]. We can extend this result to widely multi-linear MMSE estimate in the form of the following theorem.

Theorem 2. *If $\mathcal{X}$ and $\mathcal{Y}$ are two jointly complex Gaussian tensors, then the best MMSE estimator $\mathbb{E}[\mathcal{X}|\mathcal{Y}]$ is the widely multi-linear estimator of $\mathcal{X}$ from $\mathcal{Y}$.*

A proof of Theorem 2 has been included in Appendix C.

D. Comparison with Tucker based Tensor MMSE filter

The n-mode Wiener filter which makes use of the Tucker product [13], [47] constitutes a commonly used signal processing technique. Consider an observed tensor $\mathcal{Y} \in \mathbb{C}^{I_1 \times \ldots \times I_N}$. The objective is to estimate $\mathcal{X} \in \mathbb{C}^{I_1 \times \ldots \times I_N}$ based on observation $\mathcal{Y}$. Note that such class of estimators assume that the signal to be estimated and observed signals have the same dimensions. The estimator of the signal tensor $\mathcal{X}$ can be represented by N n-mode filters $\mathrm{A}^{(n)} \in \mathbb{C}^{I_n \times I_n}$ using Tucker product as follows [28]:

$$\hat{\mathcal{X}}_T = \mathcal{Y} \times_1 \mathrm{A}^{(1)} \times_2 \mathrm{A}^{(2)} \times_3 \ldots \times_N \mathrm{A}^{(N)} \tag{64}$$

where the criteria for obtaining the optimal n-mode filters $\mathrm{A}^{(n)}$ is the minimization of the mean square error between $\mathcal{X}$ and $\hat{\mathcal{X}}_T$, defined as :

$$\begin{aligned}
e(\mathrm{A}^{(1)}, \mathrm{A}^{(2)}, \ldots, \mathrm{A}^{(N)}) &= \mathbb{E}[||\mathcal{X} - \hat{\mathcal{X}}_T||^2] && (65) \\
&= \mathbb{E}[||\mathcal{X} - \mathcal{Y} \times_1 \mathrm{A}^{(1)} \times_2 \mathrm{A}^{(2)} \times_3 \ldots \times_N \mathrm{A}^{(N)}||^2] && (66)
\end{aligned}$$

The optimal choice of the n-mode filters $\mathrm{A}^{(n)}$ which ensures minimum mean-square error between $\mathcal{X}$ and $\hat{\mathcal{X}}_T$ is calculated using n-mode Wiener filtering method which relies on matrix unfoldings of $\mathcal{Y}$ [47]. Initially, all the factor matrices $\mathrm{A}^{(n)}$ are set to identity matrices. For updating $\mathrm{A}^{(n)}$ for each n, it is assumed that $\mathrm{A}^{(m)}$ for $m \neq n$ are known. Hence an alternative least squares approach is employed to calculate all the optimal $\mathrm{A}^{(n)}$ where $\mathrm{A}^{(m)}$ for $m \neq n$ is fixed to find $\mathrm{A}^{(n)}$ for all n, and then we repeat for all n until a convergence criteria is met. A detailed derivation of the solution and the algorithm to calculate the optimal n-mode matrix filters is presented in [13], [28], [47]. Note that in literature the Tucker operator based estimator is presented only for the multi-linear case and not widely multi-linear. A simple extension to a widely multi-linear case using Tucker operator would require finding optimal factor matrices $\mathrm{B}^{(n)} \in \mathbb{C}^{I_n \times I_n}$ to operate on the conjugate of $\mathcal{Y}$ as well, i.e.

$$\hat{\mathcal{X}}_{WT} = \mathcal{Y} \times_1 \mathrm{A}^{(1)} \times_2 \ldots \times_N \mathrm{A}^{(N)} + \mathcal{Y}^* \times_1 \mathrm{B}^{(1)} \times_2 \ldots \times_N \mathrm{B}^{(N)} \tag{67}$$

such that the mean square error between $\mathcal{X}$ and $\hat{\mathcal{X}}_{WT}$ is minimized. Notice that both (64) and (67) can be seen as specific cases of the widely multi-linear estimator from (29). Hence the Tucker product based estimator can be seen as a specific case of the MMSE estimator presented in this paper with an additional constraint. On writing (64) element-wise,

$$\hat{\mathcal{X}}_{Ti_1,\ldots,i_N} = \sum_{j_N=1}^{I_N}\cdots\sum_{j_1=1}^{I_1} \mathcal{Y}_{j_1,\ldots,j_N} \cdot \mathrm{A}^{(1)}_{i_1,j_1}\cdot \mathrm{A}^{(2)}_{i_2,j_2}\cdots \mathrm{A}^{(N)}_{i_N,j_N} \tag{68}$$

We define a tensor $\mathcal{A} \in \mathbb{C}^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}$ such that

$$\mathcal{A}_{i_1,\ldots,i_N,j_1,\ldots,j_N} = \mathrm{A}^{(1)}_{i_1,j_1}\cdot \mathrm{A}^{(2)}_{i_2,j_2}\cdots \mathrm{A}^{(N)}_{i_N,j_N} \tag{69}$$

In this case we can re-write (68) as :

$$\hat{\mathcal{X}}_{Ti_1,\ldots,i_N} = \sum_{j_N=1}^{I_N}\cdots\sum_{j_1=1}^{I_1} \mathcal{Y}_{j_1,\ldots,j_N} \cdot \mathcal{A}_{i_1,\ldots,i_N,j_1,\ldots,j_N}$$
$$\Rightarrow \hat{\mathcal{X}}_T = \mathcal{A} *_N \mathcal{Y} \tag{70}$$

The solution for the optimal tensor $\mathcal{A}$ which minimizes the mean square-error between $\mathcal{X}$ and $\hat{\mathcal{X}}_T$ in (70) is the multi-linear MMSE estimator as given by (47). Similarly (67) can be equivalently written as (29) with constraints $\mathcal{A}_{1i_1,\ldots,i_N,j_1,\ldots,j_N} = \mathrm{A}^{(1)}_{i_1,j_1}\cdots \mathrm{A}^{(N)}_{i_N,j_N}$ and $\mathcal{A}_{2i_1,\ldots,i_N,j_1,\ldots,j_N} = \mathrm{B}^{(1)}_{i_1,j_1}\cdots \mathrm{B}^{(N)}_{i_N,j_N}$. Hence the Tucker multi-linear MMSE estimator can be seen as a special case of the Einstein product based multi-linear MMSE estimator with an additional constraint that the tensor $\mathcal{A}$ can be written as in (69). Similarly, the Tucker widely multi-linear MMSE estimator can be seen as a constrained case of the widely multi-linear MMSE estimator from (29). The constraint (69) expresses the tensor $\mathcal{A}$ as a rearranged outer product of N factor matrices $\mathrm{A}^{(n)}$. This extra constraint makes the performance of Tucker operator based estimators sub-optimal within the class of multi-linear estimators. The Tucker or the n-mode filtering approach aims to find factor matrices along each mode separately, thereby assumes that the optimal multi-linear estimator can be written in terms of product of such factor matrix elements. But the proposed estimator based on Einstein product finds the best multi-linear estimator with no constraints or assumptions on separability of the estimator across different modes, thereby providing better mean square error performance.

Complexity Analysis: Even though Tucker operator based estimator is sub-optimal, it has a computational advantage over the more general multi-linear estimator using the Einstein product proposed in our work. Finding the estimator for multi-linear MMSE estimation of order N tensor requires inverting the covariance tensor of size $I_1\times I_2\times\ldots I_N\times I_1\times I_2\times\ldots\times I_N$, which requires $\mathcal{O}((I_1\cdots I_N)^3)$ operations [36]. The Tucker operator based approach while admitting sub-optimality breaks the tensor estimation problem into N smaller linear estimation problems which requires inverting N matrices of size $I_n \times I_n$ for $n = 1,\ldots,N$ where complexity of each matrix inversion is $\mathcal{O}(I_n^3)$ [26]. A comparison of the computational cost for the two methods is presented in Table I. Assuming $I_n = L$ for all n, it can be seen from Table I that the complexity of the Einstein product approach is $\mathcal{O}(L^{3N})$, i.e. exponential in the number of domains N as opposed to the Tucker approach which has a complexity of $\mathcal{O}(NL^3)$. So as N increases, the Tucker method provides a low complexity solution but with sub-optimal performance. In section IV-B we present numerical examples illustrating the loss of performance due to the sub-optimality of Tucker approach as N grows, and compare it with the Einstein product based approach.

TABLE I: Complexity comparison for tensor estimation

Method	Complexity
Using Tucker product	$\mathcal{O}(I_1^3 + I_2^3 + \cdots + I_N^3)$
Using Einstein product	$\mathcal{O}((I_1 \cdot I_2 \cdots I_N)^3)$

The higher complexity for Einstein product approach is primarily because of the tensor inversion operation. Various algorithms, both direct and iterative are considered in literature for finding the inverse of a tensor such as higher order biconjugate gradient method [29], Newton method [36], elimination method using tensor triangular decomposition [37], which all rely on the Einstein product. Several approaches can be used to reduce the complexity of such operations. For instance, since the entity to be inverted is covariance, which is a Hermitian tensor, the symmetry in its elements can be exploited to reduce the complexity in half by utilizing appropriate triangular decompositions. For such purposes, tensor triangular decompositions are considered in [37] which also presents a Gauss elimination method for solving tensor inversion with complexity reduced by a constant factor but still remaining $\mathcal{O}(L^{3N})$. Iterative approaches such as the Newton method (NM) can provide fast convergence as shown in [36], however the complexity of each iteration is itself cubic in tensor size as it performs an Einstein product between two tensors of order $2N$. But this can be reduced to a complexity which is square in tensor size. Using the approach in Appendix D-1, the per iteration complexity of tensor inversion using NM can be brought to $\mathcal{O}(L^{2N})$. Further, use of parallel processing can bring down the time complexity of such tensor operations significantly. It was shown in [48] that the NM requires $\mathcal{O}(\log^2 n)$ operations for large n with parallel processors to reach a fixed error bound while computing the inverse of an $n \times n$ matrix. In case of tensor inversion, developing such parallel approach for NM can lead to a per-iteration complexity which is linear in the number of domains and not exponential. Using parallel processing, the time complexity of each iteration in NM is $\mathcal{O}(\log L^N)$ as shown in Appendix D-2. Further using similar line of arguments as in [48, Theorem 2.1], it can be shown that the NM requires $\mathcal{O}(\log L^N)$ iterations to converge to a fixed error bound. Some numerical results on number of iterations showing the fast convergence of NM are also presented in [36]. Hence using parallel processing, the time complexity of NM for tensor inversion is $\mathcal{O}(\log^2 L^N)$. Further exploring different tensor algorithms and their efficient implementation by exploiting their structural properties remain a topic for future investigation. Developing such algorithms for various tensor computations, particularly solving systems of multi-linear equations using the Einstein product is an active area of research in the field of numerical tensor algebra [49]–[53].

IV. Applications of Tensor MMSE Estimation

In this section, we explore the application of tensor MMSE estimation in multi-domain communication systems. In most

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2021.3084541, IEEE Open Journal of Signal Processing

modern communication systems, multiple domains of transmission and reception are utilized to fully exploit the diversity and multiplexing benefits associated with all the available resources. For instance, exploiting the space domain through use of multiple antennas (MIMO) is a common practice in most wireless communication systems. Further, exploiting the frequency domain through multi-carrier techniques such as OFDM is also widespread. In fact, the number of domains in modern communication systems go well beyond space and frequency, and can include time slots, users, transmissions devices, multipath, spreading sequence, and code, depending on specific system. A few examples of such communication systems along with the considered signal domains are presented in Table II.

TABLE II: Examples of multiple domains in Communication Systems.

Systems	Input/Output Domains
MIMO OFDM	Antenna, and sub-carrier [36].
Wideband MIMO	Antenna, and delay [54].
MIMO FBMC	Antenna, and sub-carrier [15].
MIMO OFDM-CDMA	Antenna, data stream, sub-carrier, time blocks, and chips [17].
MU-MIMO GFDM	Users, data streams, sub-carriers, and sub-symbols [15].

Tensors are a natural choice to represent such signals which are indexed by multiple indices. Subsequently, in a generic system model of a multi-domain communication system, the input and output signals can be defined using tensors of order N and M respectively where N denotes the number of transmit domains and M denotes the number of receive domains. A multi-linear channel between the input and output can be defined as an order $N+M$ tensor such that the output is given using the Einstein product between the input and channel. Let us consider a general case with input $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1\times\ldots\times I_N}$ and output $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{J_1\times\ldots\times J_M}$. The channel is denoted using $\mathcal{H} \in \mathbb{C}^{J_1\times\ldots\times J_M\times I_1\times\ldots\times I_N}$. Each I_n denotes the dimension of the nth transmit domain, and J_m denotes the dimension of the mth receive domain. The system model for a multi-domain communication system can be written as :

$$\boldsymbol{\mathcal{Y}} = \mathcal{H} *_N \boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{N}} \tag{71}$$

where $\boldsymbol{\mathcal{N}} \in \mathbb{C}^{J_1\times\ldots\times J_M}$ is the received noise tensor. The model presented in (71) is generic where the values of N, M and the individual dimensions of each domain would depend on the specific system being considered. For instance, in a MIMO GFDM system, S streams of data are transmitted using K sub-carriers and P timeslots called sub-symbols. Hence the transmit data symbols corresponding to kth sub-carrier, pth sub-symbol and sth data stream denoted as $d_{s,k,p}$ can be represented using elements of a third order tensor $\boldsymbol{\mathcal{D}} \in \mathbb{C}^{S\times K\times P}$. Similarly the received signal and noise tensors can be represented using third order tensors $\tilde{\boldsymbol{\mathcal{D}}} \in \mathbb{C}^{S\times K\times P}$ and $\boldsymbol{\mathcal{N}} \in \mathbb{C}^{S\times K\times P}$ respectively. Subsequently the channel that couples the input $\boldsymbol{\mathcal{D}}$ with the output $\tilde{\boldsymbol{\mathcal{D}}}$ can be seen as a sixth order tensor $\mathcal{H} \in \mathbb{C}^{S\times K\times P\times S\times K\times P}$ which contracts with input over 3 modes ($*_3$) to generate the output. Hence the system model can be specified using (71) with $N = M = 3$ and $I_1 = J_1 = S$ (number of data streams), $I_2 = J_2 = K$ (number of sub-carriers) and $I_3 = J_3 = P$ (number of sub-symbols), as [55]:

$$\tilde{\boldsymbol{\mathcal{D}}} = \mathcal{H} *_3 \boldsymbol{\mathcal{D}} + \boldsymbol{\mathcal{N}} \tag{72}$$

The tensor channel considered here for MIMO GFDM is the equivalent channel obtained from the cascading of the transmit filter, physical channel and the receive filter. A detailed derivation of this system model is presented in [55]. Similarly, in section IV-D, we present the system model for MIMO OFDM using (71) where the channel is modelled as a fourth order tensor.

Note that the system model in (71) is not obtained from any mapping of the MIMO matrix channel corresponding to multiple antennas at transmitter and receiver into a tensor channel. The dependence of input, output and channel on multitude of indices beyond just antenna in a multi-domain communication system naturally gives rise to the system model in (71). In fact, the standard MIMO matrix channel model is also a special case of (71) where input, output and noise are order 1 tensors (vectors), the channel is an order 2 tensor (matrix), and the Einstein product reduces to standard matrix multiplication. The advantage of defining a multi-domain communication system model using (71) is that it allows a consolidated representation of all the signalling domains while retaining the distinction between them. Further it enables characterizing all the inter-domain interferences such as inter-carrier interference, inter-antenna interference, inter-user interference in a single framework.

A standard task for a receiver in a communication system is to estimate the transmitted signal based on the noisy observation received through the channel. Hence for a system where the input and output signals are modelled using tensors as in (71), the tensor based MMSE estimation techniques developed in this paper can be employed at the receiver.

For a known channel $\mathcal{H}$, assuming $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{N}}$ to be independent and zero mean, using (71) we can write the received covariance, cross covariance, received pseudo-covariance and cross pseudo-covariance tensors as :

$$\mathcal{C}_{\boldsymbol{\mathcal{Y}}} = \mathcal{H} *_N \mathcal{C}_{\boldsymbol{\mathcal{X}}} *_N \mathcal{H}^H + \mathcal{C}_{\boldsymbol{\mathcal{N}}} \tag{73}$$

$$\mathcal{C}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}} = \mathcal{C}_{\boldsymbol{\mathcal{X}}} *_N \mathcal{H}^H \tag{74}$$

$$\tilde{\mathcal{C}}_{\boldsymbol{\mathcal{Y}}} = \mathcal{H} *_N \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{X}}} *_N \mathcal{H}^T + \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{N}}} \tag{75}$$

$$\tilde{\mathcal{C}}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}} = \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{X}}} *_N \mathcal{H}^T \tag{76}$$

respectively. Here $\mathcal{C}_{\boldsymbol{\mathcal{X}}}, \tilde{\mathcal{C}}_{\boldsymbol{\mathcal{X}}}, \mathcal{C}_{\boldsymbol{\mathcal{N}}}$, and $\tilde{\mathcal{C}}_{\boldsymbol{\mathcal{N}}}$ denote the input covariance, input pseudo-covariance, noise covariance and noise pseudo-covariance tensors respectively. Substituting (73) and (74) into (48) and (51) gives the receiver structure based on multi-linear MMSE estimation and the associated mean square error respectively. Similarly substituting (73), (74), (75) and (76) into (39) and (41) gives us tensors $\mathcal{A}_1$ and $\mathcal{A}_2$ which on substituting into (29) and (44) yields the receiver structure based on widely multi-linear MMSE estimation and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2021.3084541, IEEE Open Journal of Signal Processing

the associated mean square error respectively.

Let us assume that the transmitted tensor contains independent elements normalised to unit power such that $\mathcal{C}_{\mathbf{\mathfrak{X}}} = \mathcal{I}_N$ which is an identity tensor of size $I_1 \times \ldots \times I_N \times I_1 \times \ldots \times I_N$. Let the noise be additive circular Gaussian noise with mean zero and variance σ_n^2, such that $\mathcal{C}_{\mathbf{\mathfrak{N}}} = \sigma_n^2 \mathcal{I}_M$ and $\tilde{\mathcal{C}}_{\mathbf{\mathfrak{N}}} = 0_{\mathcal{T}}$. Hence the mean square error from widely multi-linear estimation (based on (44)) can be given as :

$$\begin{aligned}\text{MSE} = \text{tr}(\mathcal{I}_N) - \text{tr}\Big(& (\mathcal{H}^H - \tilde{\mathcal{C}}_{\mathbf{\mathfrak{X}}} *_N \mathcal{H}^T *_M (\mathcal{H} *_N \mathcal{H}^H + \\ & \sigma_n^2 \mathcal{I}_M)^{*-1} *_M (\mathcal{H} *_N \tilde{\mathcal{C}}_{\mathbf{\mathfrak{X}}} *_N \mathcal{H}^T)^*) *_M \mathcal{P}_{\mathbf{\mathfrak{Y}}}^{-1} *_M \mathcal{H}\Big) - \\ & \text{tr}\Big((\tilde{\mathcal{C}}_{\mathbf{\mathfrak{X}}} *_N \mathcal{H}^T - \mathcal{H}^H *_M (\mathcal{H} *_N \mathcal{H}^H + \sigma_n^2 \mathcal{I}_M)^{-1} *_M \\ & \mathcal{H} *_N \tilde{\mathcal{C}}_{\mathbf{\mathfrak{X}}} *_N \mathcal{H}^T) *_M \mathcal{P}_{\mathbf{\mathfrak{Y}}}^{*-1} *_M \mathcal{H}^* *_N \tilde{\mathcal{C}}_{\mathbf{\mathfrak{X}}}^H \Big)\end{aligned} \tag{77}$$

where $\mathcal{P}_{\mathbf{\mathfrak{Y}}}$ is given by (40). The MSE performance difference between receivers employing widely multi-linear and multi-linear estimators in such multi-domain communication systems can be found using (55) as :

$$\begin{aligned}\Delta_e = \text{tr}\Big[& \Big(\tilde{\mathcal{C}}_{\mathbf{\mathfrak{X}}} *_N \mathcal{H}^T - \mathcal{H}^H *_M (\mathcal{H} *_N \mathcal{H}^H + \sigma_n^2 \mathcal{I}_M)^{-1} *_M \\ & \mathcal{H} *_N \tilde{\mathcal{C}}_{\mathbf{\mathfrak{X}}} *_N \mathcal{H}^T \Big) *_M \mathcal{P}_{\mathbf{\mathfrak{Y}}}^{*-1} *_M \Big(\tilde{\mathcal{C}}_{\mathbf{\mathfrak{X}}} *_N \mathcal{H}^T - \mathcal{H}^H *_M \\ & (\mathcal{H} *_N \mathcal{H}^H + \sigma_n^2 \mathcal{I}_M)^{-1} *_M \mathcal{H} *_N \tilde{\mathcal{C}}_{\mathbf{\mathfrak{X}}} *_N \mathcal{H}^T \Big)^H \Big]\end{aligned} \tag{78}$$

Notice that if $\tilde{\mathcal{C}}_{\mathbf{\mathfrak{X}}} = 0_{\mathcal{T}}$ which is when $\mathbf{\mathfrak{X}}$ is proper, then Δ_e is also zero as widely multi-linear estimator reduces to a multi-linear estimator in this case. The MSE expression from (77) simplifies to :

$$\text{MSE} = \text{tr}\left(\mathcal{I}_N - \mathcal{H}^H *_M (\mathcal{H} *_N \mathcal{H}^H + \sigma_n^2 \mathcal{I}_M)^{-1} *_M \mathcal{H}\right) \tag{79}$$

which is same as the MSE from multi-linear estimation (from (51)) with $\mathcal{C}_{\mathbf{\mathfrak{X}}\mathbf{\mathfrak{Y}}} = \mathcal{H}^H$ and $\mathcal{C}_{\mathbf{\mathfrak{Y}}} = (\mathcal{H} *_N \mathcal{H}^H + \sigma_n^2 \mathcal{I}_M)$.

In the subsequent sub-sections, we present numerical examples to illustrate the concept of tensor multi-linear (dubbed as TL henceforth) and tensor widely multi-linear (dubbed as TWL henceforth) MMSE estimators in the context of multi-domain communication systems. We use (29) at the receiver for TWL estimation with $\mathcal{A}_1$ and $\mathcal{A}_2$ calculated from (39) and (41). Also, (48) is used at the receiver for TL estimation. The received covariance, cross covariance, pseudo-covariance and cross pseudo-covariance are calculated based on (73)-(76). The elements of input tensor are generated with zero mean variance σ_s^2 and noise tensor with zero mean variance σ_n^2. The SNR is defined as σ_s^2/σ_n^2. For all the examples, we keep $\sigma_s^2 = 1$ and vary σ_n^2 to attain different SNR values. We use Monte Carlo simulations where input, channel and noise are randomly generated with channel realizations known at the receiver. The results are averaged over $N_{ch} = 500$ channel realizations for each SNR, with $N_{in} = 100$ noise and input realizations for each channel realization. The performance is evaluated in terms of mean square error normalized with respect to the number of transmit tensor elements, i.e.

$$\text{MSE} = \frac{1}{N_{ch}} \frac{1}{N_{in}} \sum_{k=1}^{N_{ch}} \sum_{l=1}^{N_{in}} \frac{||\mathbf{\mathfrak{X}}^{(k,l)} - \hat{\mathbf{\mathfrak{X}}}^{(k,l)}||^2}{\text{numel}[\mathbf{\mathfrak{X}}^{(k,l)}]} \tag{80}$$

where $\mathbf{\mathfrak{X}}^{(k,l)}$ and $\hat{\mathbf{\mathfrak{X}}}^{(k,l)}$ denote the actual and the estimated tensors at the kth channel and lth input run for a fixed SNR. The number of transmit elements are denoted by numel$[\mathbf{\mathfrak{X}}^{(k,l)}]$.

A. Example with Gaussian input signals

In (71), let the input tensor, $\mathbf{\mathfrak{X}} \in \mathbb{C}^{4\times4\times4}$ have independent zero mean unit variance improper complex Gaussian entries. Let each element of $\mathbf{\mathfrak{X}}$ be represented by $x = a + ib$ where a and b are real scalar random variables. To generate a and b in simulation such that they are correlated with coefficient ρ, we consider the vector $\underline{\mathbf{p}} = [a, b]^T$ and its correlation matrix $\mathrm{R} = [\text{var}(a), \text{cov}(a, b); \text{cov}(b, a), \text{var}(b)]$, where $\text{cov}(a, b) = \rho\sqrt{\text{var}(a)\text{var}(b)}$. The Cholesky decomposition of R is given as $\mathrm{R} = \mathrm{LL}^T$. We generate vector $\underline{\mathbf{q}} = [c, d]^T$ where c and d are uncorrelated zero mean and variance $1/2$ Gaussian scalars. Then $\underline{\mathbf{p}} = \mathrm{L}\underline{\mathbf{q}}$ generates vector with entries a and b such that $\text{var}(a) = \text{var}(b) = 1/2$ and they are correlated with coefficient ρ. All the entries of $\mathbf{\mathfrak{X}}$ designated as $x = a + ib$ are generated independently using this Cholesky Decomposition method. Hence input pseudo-covariance $\tilde{\mathcal{C}}_{\mathbf{\mathfrak{X}}}$ is a pseudo-diagonal tensor where all its non-zero entries are $\mathbb{E}[x^2] = \text{var}(a) - \text{var}(b) + i \cdot 2\text{cov}(a, b)$. Thus for different ρ, we get different $\tilde{\mathcal{C}}_{\mathbf{X}}$. The input covariance is $\mathcal{C}_{\mathbf{\mathfrak{X}}} = \sigma_s^2 \mathcal{I}$ where $\mathcal{I}$ is an order 6 identity tensor and $\sigma_s^2 = \text{var}(a) + \text{var}(b) = 1$. Furthermore $\mathbf{\mathfrak{N}} \in \mathbb{C}^{4\times4\times4}$ is an order three received noise tensor with zero mean circular complex Gaussian entries and independent of input signal with covariance $\mathcal{C}_{\mathbf{\mathfrak{N}}} = \sigma_n^2 \mathcal{I}$ and pseudo-covariance $\tilde{\mathcal{C}}_{\mathbf{\mathfrak{N}}} = 0_{\mathcal{T}}$. The channel $\mathcal{H} \in \mathbb{C}^{4\times4\times4\times4\times4\times4}$ contains i.i.d. zero mean unit variance circular complex Gaussian entries. The objective is to estimate $\mathbf{\mathfrak{X}}$ based on the observation $\mathbf{\mathfrak{Y}}$.

Figure 1 represents the MSE vs ρ at 10 dB SNR . We can see that for a low magnitude of ρ, i.e when the signal is almost proper, both TL and TWL estimation results in almost same mean square error, but as the magnitude of ρ increases the TWL estimator performs much better than TL estimator. The mean square error essentially remains flat for TL estimation when changing ρ as it does not take into account the pseudo-covariance. As seen in Figure 1, the mean square error follows quite well the theoretical mean square error calculated from (51) and (44), which validates our simulation set up. Further, Figure 2 presents the mean square error against SNR for specific values of ρ. With increase in SNR, mean square error reduces but the TWL estimator performs much better than the TL estimator for higher values of ρ. The TL estimator performance does not change with ρ as it does not depend on pseudo-covariance. However, for a given SNR the TWL estimator performance improves as ρ increases as it uses the correlation between real and imaginary components of the tensor for estimation.

B. Example of Tucker based MMSE Estimation

We compare the Tucker operator based MMSE estimator with the Einstein product based multi-linear MMSE estimator presented in this paper. In a multi-domain communication system, consider input $\mathbf{\mathfrak{X}}$, output $\mathbf{\mathfrak{Y}}$ and noise $\mathbf{\mathfrak{N}}$ are order N tensors where dimension of each individual domain is 3. Hence

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2021.3084541, IEEE Open Journal of Signal Processing

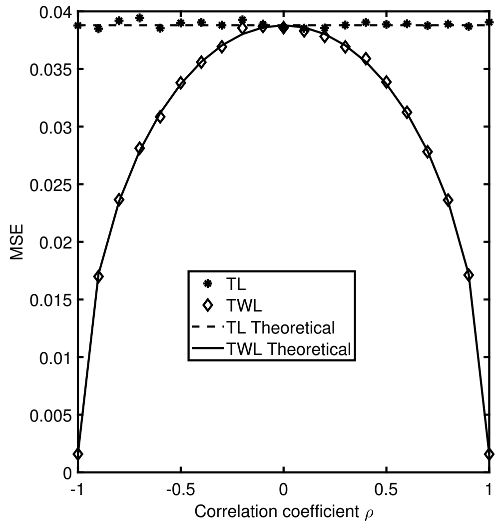

Fig. 1: MSE vs correlation coefficient between real and imaginary parts of Gaussian input at 10 dB SNR

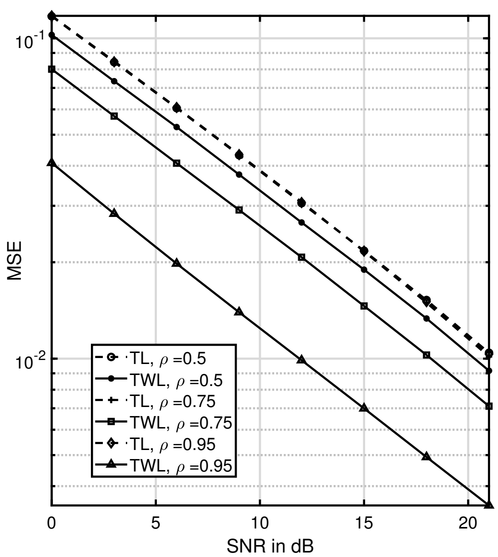

Fig. 2: MSE vs SNR with Gaussian input for different ρ

the channel $\mathcal{H}$ is an order $2N$ tensor with all the dimensions being 3. We assume that the channel can be written in terms of N factor matrices $\mathrm{H}^{(1)}, \mathrm{H}^{(2)}, \ldots, \mathrm{H}^{(N)}$ of size 3×3 each as in (69). Hence the system model can be written in two equivalent forms as :

$$\mathcal{Y} = \mathcal{X} \times_1 \mathrm{H}^{(1)} \times_2 \mathrm{H}^{(2)} \times_3 \ldots \times_N \mathrm{H}^{(N)} + \mathcal{N} \tag{81}$$

$$\mathcal{Y} = \mathcal{H} *_N \mathcal{X} + \mathcal{N} \tag{82}$$

where $\mathcal{H}_{i_1,i_2,\ldots,i_N,j_1,j_2,\ldots,j_N} = \mathrm{H}^{(1)}_{i_1,j_1} \cdot \mathrm{H}^{(2)}_{i_2,j_2} \cdots \mathrm{H}^{(N)}_{i_N,j_N}$. The equivalence of these system models can be established by writing the equations element-wise as shown in (68)-(70). For simulations, $\mathcal{X}$ and $\mathcal{N}$ are generated using circular symmetric complex Gaussian distribution with covariance as scaled identity tensor $\sigma_s^2 \mathcal{I}$ and $\sigma_n^2 \mathcal{I}$ respectively, with $\sigma_s^2 = 1$. The components of $\mathrm{H}^{(1)}, \mathrm{H}^{(2)}, \ldots, \mathrm{H}^{(N)}$ are i.i.d. drawn from circular complex Gaussian distribution with zero mean and unit

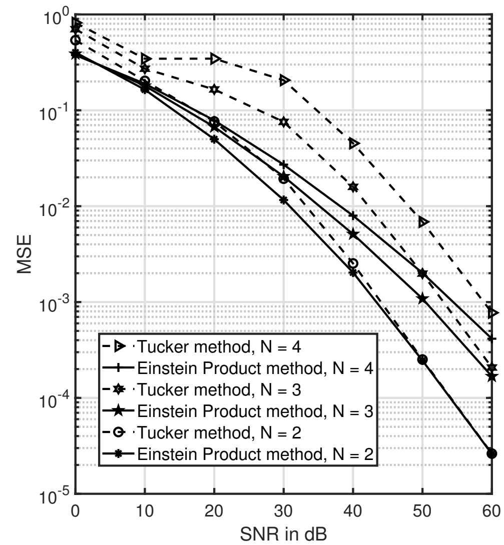

Fig. 3: MSE vs SNR for different estimation techniques

variance. The objective is to estimate $\mathcal{X}$ based on observing $\mathcal{Y}$. The Tucker operator based estimator finds factor matrices $\mathrm{A}^{(1)}, \mathrm{A}^{(2)}, \ldots, \mathrm{A}^{(N)}$ such that the estimate is given by (64). For the channel model (81) the factor matrices $\mathrm{A}^{(n)}$ are given as [26] :

$$\mathrm{A}^{(n)} = \mathrm{R}^{(n)}_{\mathcal{X}} \mathrm{H}^{(n)H} (\mathrm{H}^{(n)} \mathrm{R}^{(n)}_{\mathcal{X}} \mathrm{H}^{(n)H} + \mathrm{R}^{(n)}_{\mathcal{N}})^{-1} \tag{83}$$

for $n = 1, 2, \ldots, N$ where $\mathrm{R}^{(n)}_{\mathcal{X}} = \mathbb{E}[\mathbf{X}_{(n)} \mathbf{X}^H_{(n)}]$ and $\mathrm{R}^{(n)}_{\mathcal{N}} = \mathbb{E}[\mathbf{N}_{(n)} \mathbf{N}^H_{(n)}]$. The quantities $\mathbf{X}_{(n)}$ and $\mathbf{N}_{(n)}$ are the n-mode matrix unfoldings of $\mathcal{X}$ and $\mathcal{N}$ respectively. The mean square error achieved by this estimator is compared with the mean square error achieved by the Einstein product based multi-linear MMSE estimator of (45) where $\mathcal{A}$ is given by (47). The MSE plot is presented in Figure 3 for three different values of $N = 2, 3, 4$. The solid line corresponds to MSE when Einstein product method is used, and the dashed line corresponds to MSE when the Tucker method is used. It can be observed that in all the three cases the multi-linear MMSE estimator based on Einstein product achieves lower mean square error than the Tucker operator based estimator, and they perform similar at high SNR. Further, it can be observed that as N increases, the difference between the performance of Tucker method and Einstein product method also widens. For $N = 2$, the performance gap between the two cases is small especially at high SNR, but for $N = 4$ the gap is significant. This shows that for higher order tensors, assuming the multi-linear estimator to be separable across all the domains (as in the Tucker approach) makes the performance more sub-optimal. This is further apparent in Figure 4 where the MSE performance for the two methods is plotted against N for a fixed SNR of 30 dB. For $N = 1$, both the methods reduce to standard vector based LMMSE solution, hence they perform exactly same. As can be seen in Figure 4, with increasing N the performance difference of the Tucker approach can be significantly worse than the Einstein product method. However, it is to be noted that the complexity of Tucker approach is less than the Einstein product method as discussed in section III-D. Hence there

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2021.3084541, IEEE Open Journal of Signal Processing

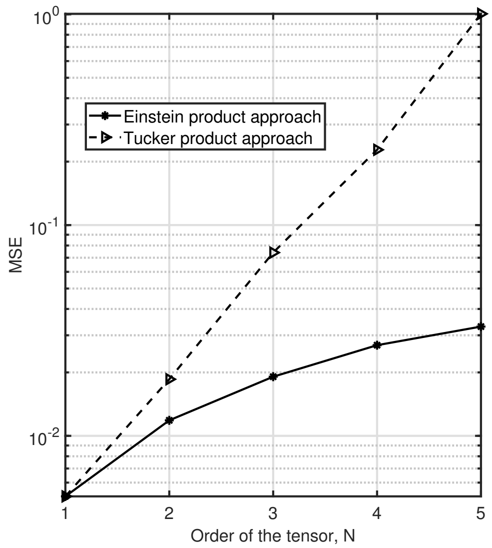

Fig. 4: MSE vs tensor order for different estimation techniques at 30 dB SNR

is an inherent trade-off between the Einstein product method and Tucker approach where the former provides much better performance, but the latter has lower complexity.

C. Estimation of Tensors in TT format

The size of a tensor grows exponentially with the number of domains. In addition, the dimensionality of each domain may be large as well. Hence tensor-based systems generate large data often which poses challenges on their storage complexity. For instance, in modern multi-domain communication systems, dimensionality of domains such as space (antenna) or frequency (sub-carriers) can be significantly large in a massive MIMO or multi-carrier scheme. However, with the help of tensor tools such as Tensor Train (TT) decomposition, effective storage mechanisms have been proposed in literature for large data. The approach using TT decomposition is not only used in cases where the data has a natural multi-way structure, but also where the data to be stored is a large vector or matrix. In such cases, the large matrix or vector is converted to a tensor for storage efficiency. This process is called Tensorization [56].

The Tensor Train (TT) decomposition represents a higher order tensor by a set of sparsely connected lower order tensors called cores or components. For a tensor $\mathcal{T}$ of size $I_1 \times I_2 \times \ldots \times I_N$, the TT decomposition is written as [57] :

$$\mathcal{T}_{i_1,\ldots,i_N} = \sum_{r_0,r_1,\ldots,r_N} \mathcal{T}^{(1)}_{r_0,i_1,r_1} \cdot \mathcal{T}^{(2)}_{r_1,i_2,r_2} \cdots \mathcal{T}^{(N)}_{r_{N-1},i_N,r_N} \quad (84)$$

where each $\mathcal{T}^{(i)}$ is a third order tensor of size $R_{i-1} \times I_i \times R_i$ and R_i denote the TT ranks with $R_0 = R_N = 1$ and $R_i \geq 1$ for $i = 1, \ldots, N-1$ [57]. In such a decomposition, the low rank structure of the core tensors is exploited for reducing the storage complexity. Rather than storing the entire tensor, only the tensor cores $\mathcal{T}^{(i)}$ are stored. Different cores can also be stored in distributed storage systems. This imposes a constraint that any mathematical operation to be performed on the tensor should not require full reconstruction or rearranging of the data but should be able to act on the cores themselves. The mechanism with which tensor operations act on the cores can be best understood using a graphical representation of the TT decomposition through a Tensor Network (TN) containing nodes and edges. A Tensor Network is a graphical illustration of tensors where each node represents a core. An edge connecting two nodes represents contraction between the two cores. The free edges correspond to the modes which are not contracted. The total number of free edges represent the order of the tensor. Hence a vector can be represented in a TN using a node with a single edge, a matrix with a node and two edges, and a tensor of order N with a node and N edges. Subsequently, the TT decomposition from (84) can be represented using a TN as shown in Figure 5. The TT ranks of a tensor determine the storage consumption of the tensor train. In many cases, the exact TT decomposition of a tensor may lead to high TT ranks. Hence often an approximation of the TT decomposition of a tensor is computed with a given accuracy ϵ to fit a desired set of low TT ranks for reduced storage [57]. We denote the TT approximation of a tensor $\mathcal{A}$ as $\bar{\mathcal{A}}$. The computed approximation satisfies $||\mathcal{A} - \bar{\mathcal{A}}||_F \leq \epsilon ||\mathcal{A}||_F$. A sequential SVD based algorithm to compute such TT decomposition with given accuracy ϵ is presented in [57].

The use of TT decomposition in signal processing applications is widespread. For instance, [58] presents the application of TT decomposition in Industrial Internet of Things (IIoT) as the TT format enables distributed processing of large scale multi-attribute data by decomposing it into different cores. A tensor framework for efficient big data processing, storage and streaming based on TT decomposition has been presented in [59]. A TT based deep neural network has been considered in [60] for channel estimation in time varying MIMO systems. Furthermore, the use of TT format in wireless communication systems has been considered in [61], [62]. In [61], the channel for a dual polarized MIMO antenna system is expressed as a fifth order tensor by exploiting the azimuth and elevation diversities at both the transmit and receive ends along with multipath. Such a fifth order tensor channel is represented using the TT format to allow low computation channel parameter estimation. In [62], a joint channel and symbol estimation approach is presented using the TT decomposition, where the received signal in a MIMO OFDM relaying system is expressed in TT format.

The tensor-based technique presented in this paper offers itself as an effective mechanism for data estimation when tensors are expressed in TT format. This is because the MMSE estimation methods proposed in our work do not rely on any tensor to vector/matrix transformation, but makes use of the Einstein Product. The Einstein product between two different tensors, $\mathcal{Q} *_K \mathcal{P}$ stored in TT format can be represented using TN as shown in Figure 6. The gray nodes represent the cores in the tensor train of $\mathcal{P} \in \mathbb{C}^{L_1 \times \ldots \times L_K \times J_1 \times \ldots \times J_M}$ and the white nodes represent the cores in the tensor train of $\mathcal{Q} \in \mathbb{C}^{I_1 \times \ldots \times I_N \times L_1 \times \ldots \times L_K}$. An algorithm to compute the Einstein Product for tensors in TT format without reconstructing the whole tensor is presented in [63]. Next we will show through an example the application of tensor MMSE estimator for tensor stored in TT format.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2021.3084541, IEEE Open Journal of Signal Processing

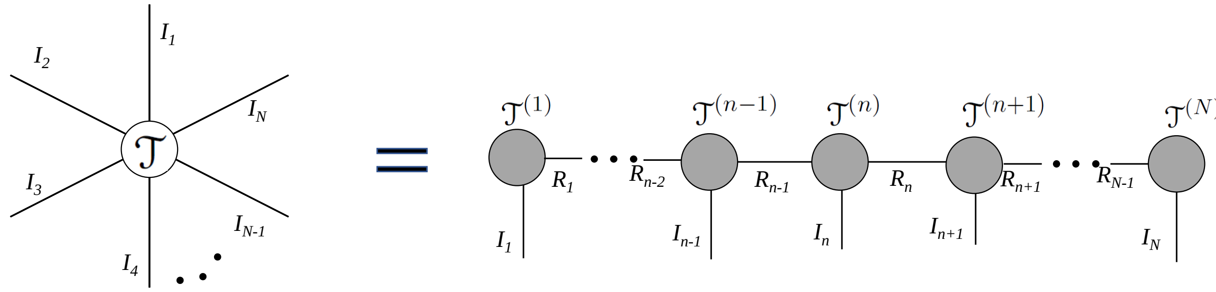

Fig. 5: TN representation of Nth order tensor $\mathcal{T} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ in TT format.

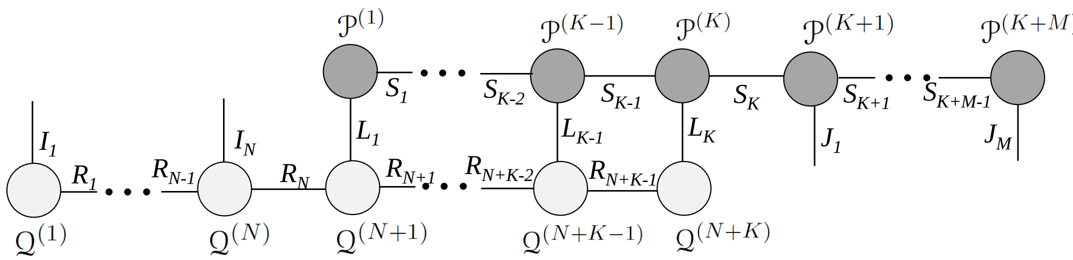

Fig. 6: TN representation of Einstein Product $\mathcal{Q} *_K \mathcal{P}$ for tensors $\mathcal{Q} \in \mathbb{C}^{I_1 \times \ldots \times I_N \times L_1 \times \ldots \times L_K}$ and $\mathcal{P} \in \mathbb{C}^{L_1 \times \ldots \times L_K \times J_1 \times \ldots \times J_M}$ in TT format where R_i and S_j represent TT ranks for $\mathcal{Q}$ and $\mathcal{P}$ respectively.

Consider the system model from (71) where $\mathbf{\mathcal{Y}}, \mathbf{\mathcal{X}}$ and $\mathbf{\mathcal{N}}$ are order 3 tensors of size $5 \times 5 \times 5$ each and $\mathcal{H}$ is an order 6 tensor of size $5 \times 5 \times 5 \times 5 \times 5 \times 5$. We use the multi-linear MMSE tensor $\mathcal{A}$ from (47) to find the estimate $\hat{\mathbf{\mathcal{X}}}$ at the receiver. For the numerical example, $\mathbf{\mathcal{X}}$ and $\mathbf{\mathcal{N}}$ are generated using i.i.d. zero mean proper complex Gaussian distribution with covariance as $\sigma_s^2 \mathcal{I}$ and $\sigma_n^2 \mathcal{I}$ respectively, with $\sigma_s^2 = 1$. The channel contains i.i.d. zero mean unit variance complex Gaussian entries and is normalized to provide unit power gain at the receiver. We show the MSE results when the observed tensor $\mathbf{\mathcal{Y}}$ and the multi-linear MMSE tensor $\mathcal{A}$ are instead stored in their TT formats $\bar{\mathbf{\mathcal{Y}}}$ and $\bar{\mathcal{A}}$ at the receiver with accuracy ϵ. Algorithm 1 from [57] is employed to calculate the TT decompositions. The estimation in tensor's original format is given as $\hat{\mathbf{\mathcal{X}}} = \mathcal{A} *_3 \mathbf{\mathcal{Y}}$. However, since we assume $\mathbf{\mathcal{Y}}$ and $\mathcal{A}$ are stored in TT formats, we find the estimate $\hat{\mathbf{\mathcal{X}}}$ also in TT format denoted as $\bar{\hat{\mathbf{\mathcal{X}}}}$ using $\bar{\mathbf{\mathcal{Y}}}$ and $\bar{\mathcal{A}}$ as :

$$\bar{\hat{\mathbf{\mathcal{X}}}} = \bar{\mathcal{A}} *_3 \bar{\mathbf{\mathcal{Y}}} \tag{85}$$

which can be written in terms of their cores as :

$$\bar{\hat{\mathbf{\mathcal{X}}}}_{i,j,k} = \sum_{r_0,r_1,r_2,r3} \hat{\mathbf{\mathcal{X}}}^{(1)}_{r_0,i,r_1} \cdot \hat{\mathbf{\mathcal{X}}}^{(2)}_{r_1,j,r_2} \cdot \hat{\mathbf{\mathcal{X}}}^{(3)}_{r_2,k,r_3} \tag{86}$$

$$= (\bar{\mathcal{A}} *_3 \bar{\mathbf{\mathcal{Y}}})_{i,j,k} = \sum_{l,m,n} \bar{\mathcal{A}}_{i,j,k,l,m,n} \cdot \bar{\mathbf{\mathcal{Y}}}_{l,m,n} \tag{87}$$

$$= \sum_{l,m,n} \Big(\sum_{p_0,p_1,\ldots,p_6} \mathcal{A}^{(1)}_{p_0,i,p_1} \cdot \mathcal{A}^{(2)}_{p_1,j,p_2} \cdot \mathcal{A}^{(3)}_{p_2,k,p_3} \cdots \mathcal{A}^{(6)}_{p_5,n,p_6} \Big) \cdot \Big(\sum_{s_0,s_1,s_2,s3} \mathbf{\mathcal{Y}}^{(1)}_{s_0,l,s_1} \cdot \mathbf{\mathcal{Y}}^{(2)}_{s_1,m,s_2} \cdot \mathbf{\mathcal{Y}}^{(3)}_{s_2,n,s_3} \Big) \tag{88}$$

The objective here is to estimate the TT cores $\hat{\mathbf{\mathcal{X}}}^{(1)}, \hat{\mathbf{\mathcal{X}}}^{(2)}$ and $\hat{\mathbf{\mathcal{X}}}^{(3)}$ using the TT cores of $\bar{\mathbf{\mathcal{Y}}}$ and $\bar{\mathcal{A}}$ without explicitly constructing the whole tensor at any stage. We use Algorithm 1 from [63] for this purpose which finds the TT cores of $\bar{\hat{\mathbf{\mathcal{X}}}}$ using the TT cores of $\bar{\mathbf{\mathcal{Y}}}$ and $\bar{\mathcal{A}}$. From the estimated cores we reconstruct $\bar{\hat{\mathbf{\mathcal{X}}}}$ and compare it with the transmitted tensor $\mathbf{\mathcal{X}}$ to calculate the mean square error in estimation. Figure 7 presents the MSE against SNR in dB for different accuracy ϵ with which TT decomposition is calculated. We compare this result with the MSE achieved when all the tensors are taken in their original non-decomposed formats, referred as 'original case' in the figure. We can see that for lower value of ϵ such as $\epsilon = 0.01$, the MSE achieved by the multi-linear MMSE estimation performed on tensors in TT format is indistinguishable from the case where tensors were used in original format. The performance degradation in MSE is observed at high SNR as ϵ increases. A small value of ϵ indicates an almost exact TT decomposition whereas a larger value of ϵ indicates a larger approximation error in the TT decomposed format. Further, Figure 8 presents the MSE against the accuracy ϵ for a fixed SNR at 20 dB. The MSE for the original case is unaffected by ϵ. For estimation in TT format, the MSE is almost same as the original case for small ϵ but after a certain value (> 0.03 in this case), the MSE increases. This increase in MSE is due to the approximation tolerance set in the computation of TT format by fixing ϵ. Hence multi-linear MMSE estimation can be used for tensors stored in TT format, and its MSE performance remains the same as that of the original case if the TT decomposition is computed with high accuracy, i.e. low ϵ. Also, since the Einstein product can be implemented directly on the cores, the proposed estimation technique does not require reconstructing the original tensors.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2021.3084541, IEEE Open Journal of Signal Processing

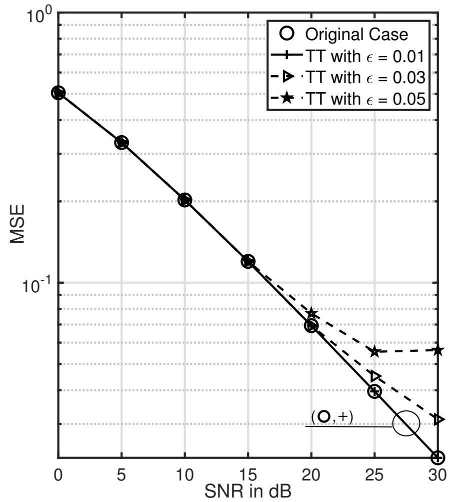

Fig. 7: MSE vs SNR for estimation of tensor in TT format

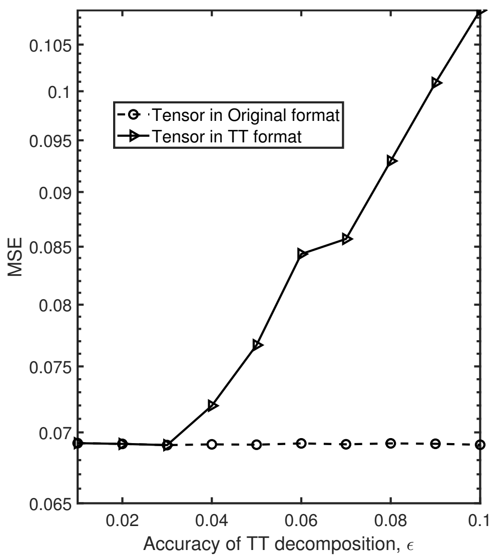

Fig. 8: MSE vs accuracy of the tensor TT format

D. Tensor Estimation for MIMO OFDM System

OFDM has been one of the most popular multi-carrier scheme which has been used in conjunction with MIMO for 4G, Wi-Fi and now also for 5G [64]. A conventional system model for MIMO OFDM in the frequency domain can be written as [65] :

$$\underline{\mathbf{y}}^{(p)} = \mathrm{H}^{(p,p)}\underline{\mathbf{x}}^{(p)} + \sum_{q=1,q\neq p}^{N_{sc}} \mathrm{H}^{(p,q)}\underline{\mathbf{x}}^{(q)} + \underline{\mathbf{n}}^{(p)}, \quad p = 1, \ldots, N_{sc} \tag{89}$$

where $\underline{\mathbf{y}}^{(p)} \in \mathbb{C}^{N_R}, \underline{\mathbf{x}}^{(p)} \in \mathbb{C}^{N_T}$ and $\underline{\mathbf{n}}^{(p)} \in \mathbb{C}^{N_R}$ are the frequency domain received, transmitted and noise symbol vectors at sub-carrier p, and N_R, N_T and N_{sc} denote the number of receive antennas, transmit antennas and sub-carriers respectively. The frequency domain channel matrix of size $N_R \times N_T$ between transmit sub-carrier q and receive sub-carrier p is $\mathrm{H}^{(p,q)}$.

One common approach used in MIMO OFDM receiver design is to assume that there is no inter-carrier interference, i.e. $\mathrm{H}^{(p,q)} = 0$ if $p \neq q$, and perform linear MMSE estimation on a per sub-carrier basis [66]–[68]. Assuming that the input $\underline{\mathbf{x}}^{(p)}$ have independent zero mean unit variance entries such that input covariance is an identity matrix, I for each p, and is independent of noise $\underline{\mathbf{n}}^{(p)}$, the receiver structure with per sub-carrier estimation is given as [66]:

$$\hat{\underline{\mathbf{x}}}^{(p)} = \mathrm{H}^{(p,p)H} \cdot (\mathrm{H}^{(p,p)} \cdot \mathrm{H}^{(p,p)H} + \mathrm{C}_N)^{-1} \cdot \underline{\mathbf{y}}^{(p)} \tag{90}$$

for $p = 1, \ldots, N_{sc}$, and $\mathrm{C}_N \in \mathbb{C}^{N_R \times N_R}$ is the noise covariance matrix. The per-sub-carrier estimation in (90) is based on the standard LMMSE filter used in matrix based systems. If we ignore the inter-carrier interference completely and assume noise to be circular symmetric white Gaussian with σ_n^2 variance entries, then $\mathrm{C}_N = \sigma_n^2 \cdot \mathrm{I}$. Alternately, one can treat the inter-carrier interference term in (89) also as noise, in which case $\mathrm{C}_N = (\sigma_n^2 \cdot \mathrm{I} + \sum_{q\neq p} \mathrm{H}^{(p,q)} \cdot \mathrm{H}^{(p,q)H})$. This approach however does not make good use of the inter-carrier interference (ICI) terms to extract signal information. In many cases such as in high mobility systems, the channel will be doubly selective leading to significant inter-carrier interference in which case ignoring the interference terms or treating them as noise would lead to sub-optimal performance. The system model suggested in (71) can be used in such a scenario where the input, output and noise can be treated as order two tensors with the two domains corresponding to antenna and sub-carriers and the channel as an order four tensor accounting for interference between antennas and between sub-carriers as well. The system model can be represented as

$$\mathbf{Y} = \mathcal{H} *_2 \mathbf{X} + \mathbf{N}, \tag{91}$$

where each vector $\underline{\mathbf{y}}^{(p)}$ and $\underline{\mathbf{n}}^{(p)}$ from (89) for $p = 1, \ldots, N_{sc}$ form the columns of received matrix $\mathbf{Y}$ and noise matrix $\mathbf{N}$ of size $N_R \times N_{sc}$, each vector $\underline{\mathbf{x}}^{(p)}$ from (89) form the columns of transmit matrix $\mathbf{X} \in \mathbb{C}^{N_T \times N_{sc}}$. The channel is represented as a fourth order tensor $\mathcal{H} \in \mathbb{C}^{N_R \times N_{sc} \times N_T \times N_{sc}}$ such that $\mathcal{H}_{:,p,:,q} = \mathrm{H}^{(p,q)}$. The ICI is reflected in the elements of $\mathcal{H}_{n_r,p,n_t,q}$ when $p \neq q$. With such a system model in place, one can use the tensor based receiver structure from (29) or (48) as the tensor formulation gives an easy mechanism to take into account the information provided by the interfering terms across all the domains.

In literature, a doubly selective channel for MIMO OFDM is handled by concatenating the transmit and receive vectors for each sub-carrier into a long vector and thereby representing the channel by a large matrix [69]–[71]. However, the tensor framework is more intuitive as it retains the distinction between the domains. Besides, for future communication systems where more than 1024 sub-carriers or hundreds of antennas in case of massive MIMO are envisioned, the number of sub-carriers and antennas can be very large. Hence concatenating the input and output as vectors will result into a large matrix channel and vector output which might be tensorized for storage efficiency through TT decomposition. Additionally, the tensor framework is intuitive and easy to extend incorporating more domains in the system model. For instance, the system

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2021.3084541, IEEE Open Journal of Signal Processing

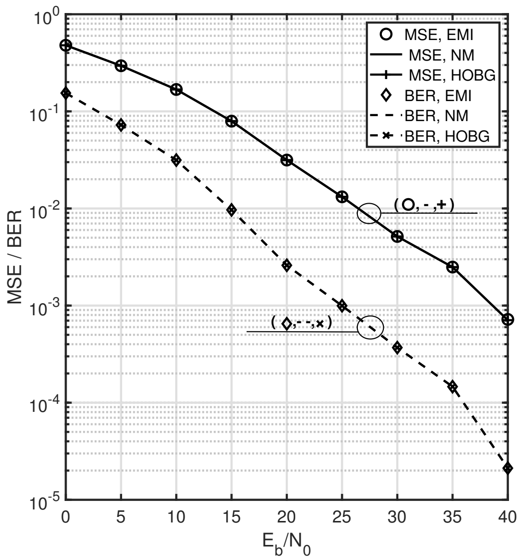

Fig. 9: MSE and BER vs E_b/N_0 for 2×2 MIMO OFDM system with 4QAM modulation

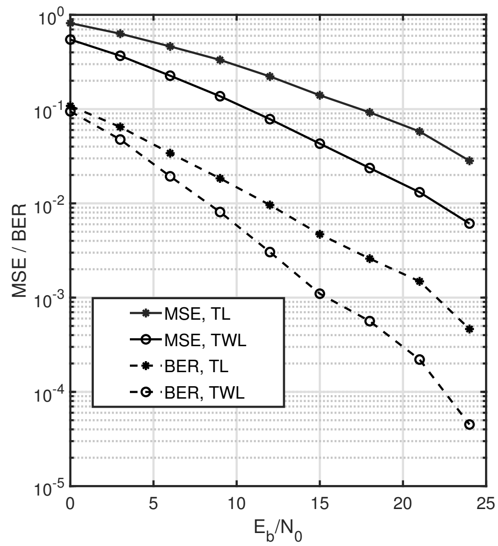

Fig. 10: MSE and BER vs E_b/N_0 for 2×2 MIMO OFDM system with BPSK modulation

model in (89) can be easily extended to multi user MIMO OFDM system incorporating users also as a domain in the system model accounting for inter-user interference as well.

We present simulation results using tensor based MMSE estimation for a MIMO OFDM system with $N_{sc} = 64$ sub-carriers, 2 transmit and 2 receive antennas. The channel between each transmit and receive pair of antennas was generated as in [72], [73]. The channel impulse response matrix between n_tth transmit and n_rth receive antenna denoted as $\bar{\mathrm{H}}^{(n_r,n_t)} \in \mathbb{C}^{N_{sc} \times N_{sc}}$ is generated by employing a two tap multipath $(L = 2)$ fading channel following Jakes' model [74] using exponential power profile, $\sigma_l^2 = \frac{\exp(-l/L)}{\sum_{l=0}^{L-1} \exp(-l/L)}$ where σ_l^2 represents the variance of the lth channel tap. This matrix is further converted to frequency domain using the DFT matrix $\mathrm{W} \in \mathbb{C}^{N_{sc} \times N_{sc}}$ with elements $\mathrm{W}_{m,n} = 1/\sqrt{N_{sc}} \exp(-j2\pi mn/N_{sc})$, which then forms the sub-tensor of the frequency domain channel tensor as $\mathcal{H}_{n_r,:,n_t,:} = \mathrm{W}\bar{\mathrm{H}}^{(n_r,n_t)}\mathrm{W}^H$. The channels are generated for different values of Doppler d, normalized to the OFDM symbol rate, to induce inter-carrier interference. Unless otherwise stated, we take $d = 0.2$. All the results presented were calculated using Monte Carlo simulations with averaging over 500 channel realizations, where at least 100 bit errors were collected for each channel realization to calculate Bit-error Rate (BER). The MSE / BER results are plotted against received SNR per bit.

1) Comparing different tensor inversion algorithms

In Figure 9 we present the MSE and BER performance for 4QAM where receiver employs TL estimator from (48) with different algorithms for tensor inversion. The output of the estimator is passed through a QAM demodulator to determine the transmitted symbols and find BER. We compare the Newton Method (NM) from (11) and Higher order Biconjugate Method (HOBG) from [29], and see that they provide similar performance. For validation of results, we also present in Figure 9 the case where tensor inversion was done via matrix transformation and using MATLAB inverse function to compute exact matrix inverse (EMI). A detailed comparison between the convergence of NM and HOBG method is presented in [36]. It is shown in [36] that as the size of tensor increases, the number of iterations required by HOBG increases significantly. The number of required iterations by the NM depends less on the tensor size, and it is in general much lower than in the HOBG technique. The rest of the results in this paper use NM for tensor inversion.

2) Improper inputs

In the previous example, a 4QAM input constellation was employed. Since 4QAM generates proper signals, hence both widely multi-linear and multi-linear receivers would give exactly same performance. Figure 10 presents the performance of TL and TWL estimators with input drawn from a BPSK constellation which makes the input improper. Hence the pseudo-covariance will be non-zero in this case. For the simulation, all the elements in the input tensor are uncorrelated and drawn from a BPSK constellation with unit energy, thereby making the input covariance $\mathcal{C}_{\mathcal{X}}$ and pseudo-covariance $\tilde{\mathcal{C}}_{\mathcal{X}}$ as identity tensors. To ensure that the input symbols are uncorrelated, one can use symbol interleaving which is commonly used in practice [75]. The received covariance, cross covariance, pseudo-covariance and cross pseudo-covariance are calculated based on (73)-(76).

For TWL estimation, (29) is used at the receiver where $\mathcal{A}_1$ and $\mathcal{A}_2$ are given by (39) and (41). For TL estimation, (48) is used at the receiver. The output of the estimator in both cases is passed through a BPSK demodulator to determine the transmitted symbols and calculate BER. Since BPSK generates improper signals, we can clearly see in Figure 10 that the widely multi-linear MMSE estimator from (29) outperforms multi-linear MMSE estimator from (48).

3) Tensor estimation against per sub-carrier estimation

In this section, we compare estimation in MIMO OFDM with 4QAM using tensor and per sub-carrier estimation. We

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2021.3084541, IEEE Open Journal of Signal Processing

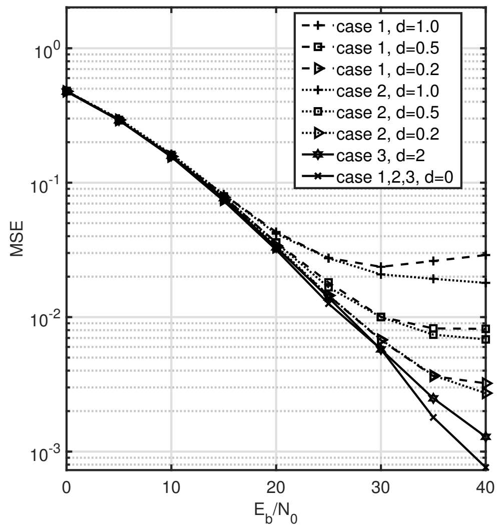

Fig. 11: MSE vs E_b/N_0 for 2×2 MIMO OFDM system with 4QAM modulation for different Doppler values d.

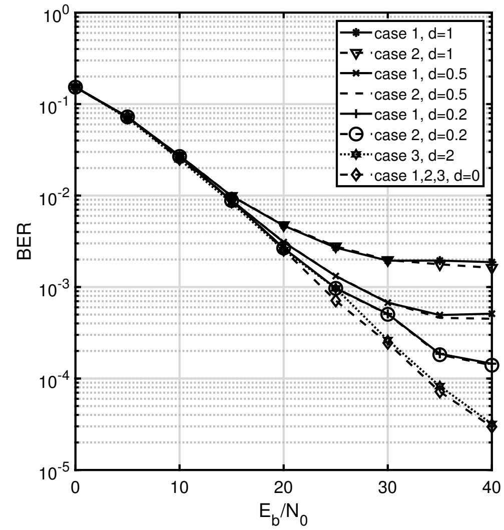

Fig. 12: BER vs E_b/N_0 for 2×2 MIMO OFDM system with 4QAM modulation for different Doppler values d.

consider three different cases based on the estimation technique that the receiver employs :

Case 1 : Per sub-carrier estimation from (90) where interference terms are completely ignored, such that $\mathrm{C}_N = \sigma_n^2 \cdot \mathrm{I}$.

Case 2 : Per sub-carrier estimation from (90) with interference treated as noise, such that $\mathrm{C}_N = (\sigma_n^2 \cdot \mathrm{I} + \sum_{q \neq p} \mathrm{H}^{(p,q)} \cdot \mathrm{H}^{(p,q)H})$.

Case 3 : Multi-linear MMSE estimation from (48) with $\mathcal{C}_{\mathcal{Y}}$ and $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ calculated from (73) and (74) respectively.

Figures 11 and 12 present the MSE and BER for the three cases with different values of the Doppler parameter d. It can be seen that as d increases, there is a significant performance degradation for case 1 and case 2 as compared to case 3. Case 2 slightly performs better than case 1 at higher SNRs as it accounts for interference albeit as noise. It can be observed in Figure 11 that for case 1 at $d = 1$, the mean square error increases at very high SNR. This is because at high SNR, the receiver of case 1 tends to a zero forcing receiver which tries to invert the channel while ignoring the interference terms totally. Hence at high value of d when the interfering terms are dominant in the received signal, the channel inversion further amplifies the interference part of the received signal leading to higher mean square error. This is easily remedied by making simultaneous use of information from all the domains with the tensor multi-linear MMSE estimator, as done in case 3. Also, case 3 shows robustness to a change in d, as the MSE or BER performance does not change significantly between $d = 0$ to $d = 2$. The robustness of the multi-linear MMSE estimation can be further seen in Figure 13 where the MSE results for case 3 are presented for very high values of Doppler parameter, d. It can be seen that only at extremely large values of Doppler, does the performance degrades. Such large values of the Doppler may not be very practical but are presented here only for a comparison and to illustrate the behavior of the proposed estimator in extreme cases.

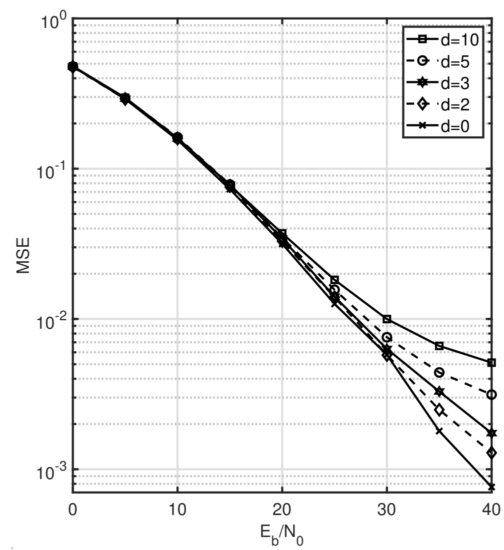

Fig. 13: MSE vs E_b/N_0 for 2×2 MIMO OFDM system with 4QAM modulation for higher Doppler values d.

V. Conclusions

This paper considered MMSE estimation techniques for tensor based signals. A unified framework for estimation of complex tensors, proper or improper, has been developed using the Einstein product. The tensor multi-linear (TL) and widely multi-linear (TWL) MMSE estimation techniques for multi-domain signals have been formulated while keeping the multi-way structure of the signals intact. We proved that for jointly complex Gaussian tensors, the best MMSE estimator is the widely multi-linear estimator. It was shown that the MSE achieved by a TWL estimator is always less than or equal to the MSE achieved by TL estimators. As compared to the TL estimator, the TWL estimator requires almost twice the computation since it finds two separate operators which act on the signal and its conjugate both for estimation. However,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2021.3084541, IEEE Open Journal of Signal Processing

the TWL estimator provides much better MSE performance compared to TL when the signal is improper by accounting for both covariance and pseudo-covariance. We compared the proposed estimator using the Einstein product with the Tucker approach. The Tucker product based estimator offers sub-optimal performance while providing lower computational complexity. The estimator using Einstein product has a higher computational complexity but provides much better MSE performance. We showed that the proposed estimator can also be used in applications where tensors are stored in TT formats. In such cases, the stored TT cores are used to find the estimate without constructing or reshaping the whole tensor. If the TT decomposition is computed with high accuracy, then the MSE performance of the proposed tensor estimation in TT format is indistinguishable from the case where tensors are stored in original format. We also presented a tensor based system model for multi-domain communication systems. As an example, we considered MIMO OFDM in a doubly selective channel using the tensor framework. The channel was represented using an order four tensor, accounting for inter-carrier and inter-antenna interference in a single framework. It was shown that tensor based MMSE estimation outperforms per sub-carrier estimation for MIMO OFDM by a significant margin when the inter-carrier interference is high.

Appendix A
Proof of Theorem 1

$$\begin{aligned}\mathbb{E}[||\mathcal{X}-h(\mathcal{Y})||^2] &= \mathbb{E}[||\underbrace{\mathcal{X}-g(\mathcal{Y})}_{\mathcal{E}}+\underbrace{g(\mathcal{Y})-h(\mathcal{Y})}_{\bar{h}(\mathcal{Y})}||^2]\\ &= \mathbb{E}[(\mathcal{E}+\bar{h}(\mathcal{Y}))*_N(\mathcal{E}+\bar{h}(\mathcal{Y}))^*]\\ &= \mathbb{E}[\mathcal{E}*_N\mathcal{E}^*+\mathcal{E}*_N\bar{h}(\mathcal{Y})^*+\bar{h}(\mathcal{Y})*_N\mathcal{E}^*+\bar{h}(\mathcal{Y})*_N\bar{h}(\mathcal{Y})^*]\\ &= \mathbb{E}[||\mathcal{E}||^2]+\underbrace{\mathbb{E}[\mathcal{E}*_N\bar{h}(\mathcal{Y})^*]+\mathbb{E}[\bar{h}(\mathcal{Y})*_N\mathcal{E}^*]}_{\text{cross-terms}}+\mathbb{E}[||\bar{h}(\mathcal{Y})||^2]\end{aligned}$$

Since $\bar{h}(\mathcal{Y})$ is another function of $\mathcal{Y}$, hence if (15) holds, then the cross terms above would be zero, which results into :

$$\mathbb{E}[||\mathcal{X}-h(\mathcal{Y})||^2]=\mathbb{E}[||\mathcal{E}||^2]+\underbrace{\mathbb{E}[||\bar{h}(\mathcal{Y})||^2]}_{\geq 0}\geq\mathbb{E}[||\mathcal{E}||^2]$$

Appendix B
Derivation of Error Covariance Tensor from (42)

$$\begin{aligned}\mathcal{Q}_{WL} = \mathbb{E}[(\mathcal{X}-\hat{\mathcal{X}}_{WL})\circ(\mathcal{X}-\hat{\mathcal{X}}_{WL})^*] = \mathbb{E}[\mathcal{X}\circ\mathcal{X}^*]-\\ \mathbb{E}[\mathcal{X}\circ\hat{\mathcal{X}}^*_{WL}]-\mathbb{E}[\hat{\mathcal{X}}_{WL}\circ\mathcal{X}^*]+\mathbb{E}[\hat{\mathcal{X}}_{WL}\circ\hat{\mathcal{X}}^*_{WL}]\end{aligned} \tag{92}$$

Using (6), $(\mathcal{A}_1 *_M \mathcal{Y})^* = (\mathcal{Y}^* *_M \mathcal{A}_1^H)$ and $(\mathcal{A}_2 *_M \mathcal{Y}^*)^* = (\mathcal{Y} *_M \mathcal{A}_2^H)$. So individual terms in (92) can be simplified as :

$$\begin{aligned}\mathbb{E}[\mathcal{X}\circ\hat{\mathcal{X}}^*_{WL}] &= \mathbb{E}[\mathcal{X}\circ(\mathcal{A}_1^* *_M \mathcal{Y}^*+\mathcal{A}_2^* *_M \mathcal{Y})]\\ &= \mathbb{E}[(\mathcal{X}\circ\mathcal{Y}^*)*_M\mathcal{A}_1^H+(\mathcal{X}\circ\mathcal{Y})*_M\mathcal{A}_2^H]\\ &= \mathcal{C}_{\mathcal{X}\mathcal{Y}}*_M\mathcal{A}_1^H+\tilde{\mathcal{C}}_{\mathcal{X}\mathcal{Y}}*_M\mathcal{A}_2^H\end{aligned} \tag{93}$$

$$\begin{aligned}\mathbb{E}[\hat{\mathcal{X}}_{WL}\circ\mathcal{X}^*] &= \mathbb{E}[(\mathcal{A}_1 *_M \mathcal{Y}+\mathcal{A}_2 *_M \mathcal{Y}^*)\circ\mathcal{X}^*]\\ &= \mathcal{A}_1 *_M \mathbb{E}[\mathcal{Y}\circ\mathcal{X}^*]+\mathcal{A}_2 *_M \mathbb{E}[\mathcal{Y}^*\circ\mathcal{X}^*]\\ &= \mathcal{A}_1 *_M \mathcal{C}^H_{\mathcal{X}\mathcal{Y}}+\mathcal{A}_2 *_M \tilde{\mathcal{C}}^H_{\mathcal{X}\mathcal{Y}}\end{aligned} \tag{94}$$

$$\begin{aligned}&\mathbb{E}[\hat{\mathcal{X}}_{WL}\circ\hat{\mathcal{X}}^*_{WL}]\\ &= \mathbb{E}[(\mathcal{A}_1 *_M \mathcal{Y}+\mathcal{A}_2 *_M \mathcal{Y}^*)\circ(\mathcal{A}_1^* *_M \mathcal{Y}^*+\mathcal{A}_2^* *_M \mathcal{Y})]\\ &= \mathbb{E}[(\mathcal{A}_1 *_M \mathcal{Y}+\mathcal{A}_2 *_M \mathcal{Y}^*)\circ(\mathcal{Y}^* *_M \mathcal{A}_1^H+\mathcal{Y} *_M \mathcal{A}_2^H)]\\ &= \mathcal{A}_1 *_M \mathcal{C}_{\mathcal{Y}} *_M \mathcal{A}_1^H+\mathcal{A}_2 *_M \tilde{\mathcal{C}}^*_{\mathcal{Y}} *_M \mathcal{A}_1^H+\\ &\quad \mathcal{A}_1 *_M \tilde{\mathcal{C}}_{\mathcal{Y}} *_M \mathcal{A}_2^H+\mathcal{A}_2 *_M \mathcal{C}^*_{\mathcal{Y}} *_M \mathcal{A}_2^H\\ &= (\mathcal{A}_1 *_M \mathcal{C}_{\mathcal{Y}}+\mathcal{A}_2 *_M \tilde{\mathcal{C}}^*_{\mathcal{Y}}) *_M \mathcal{A}_1^H+\\ &\quad (\mathcal{A}_1 *_M \tilde{\mathcal{C}}_{\mathcal{Y}}+\mathcal{A}_2 *_M \mathcal{C}^*_{\mathcal{Y}}) *_M \mathcal{A}_2^H\\ &= \mathcal{C}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{A}_1^H+\tilde{\mathcal{C}}_{\mathcal{X}\mathcal{Y}} *_M \mathcal{A}_2^H \quad \text{(from (37) and (38))}\end{aligned} \tag{95}$$

Substituting $\mathbb{E}[\mathcal{X}\circ\mathcal{X}^*]=\mathcal{C}_{\mathcal{X}}$ along with (93), (94) and (95) into (92), we get :

$$\mathcal{Q}_{WL}=\mathcal{C}_{\mathcal{X}}-\mathcal{A}_1 *_M \mathcal{C}^H_{\mathcal{X}\mathcal{Y}}-\mathcal{A}_2 *_M \tilde{\mathcal{C}}^H_{\mathcal{X}\mathcal{Y}} \tag{96}$$

Appendix C
Proof of Theorem 2

To prove Theorem 2, we show that given $\mathcal{Y}=\mathscr{Y}$, $\mathcal{X}$ is conditionally $\mathcal{CN}(\mathcal{A}_1 *_M \mathscr{Y}+\mathcal{A}_2 *_M \mathscr{Y}^*, \mathcal{Q}_{WL}, \tilde{\mathcal{Q}}_{WL})$ using the characteristic function.

The *characteristic function* of a complex random vector $\underline{\mathbf{x}}\in\mathbb{C}^N$ is defined as $\Phi_{\underline{\mathbf{x}}}(\underline{\omega})=\mathbb{E}[\exp(i\Re(\underline{\omega}^H\underline{\mathbf{x}}))]$ for $\underline{\omega}\in\mathbb{C}^N$ [76]. Using Einstein Product, the characteristic function of a complex random tensor $\mathcal{X}\in\mathbb{C}^{I_1\times\ldots\times I_N}$ is

$$\Phi_{\mathcal{X}}(\mathcal{W})=\mathbb{E}[\exp(i\Re(\mathcal{W}^* *_N \mathcal{X}))] \tag{97}$$

for tensor $\mathcal{W}\in\mathbb{C}^{I_1\times\ldots\times I_N}$. Notice that $\mathbf{z}=\mathcal{W}^* *_N \mathcal{X}$ is a complex scalar random variable. If $\mathcal{X}$ is a complex Gaussian tensor with mean $\mathcal{M}$, covariance $\mathcal{C}$ and pseudo-covariance $\tilde{\mathcal{C}}$, then $\mathbf{z}$ will be Gaussian distributed with mean $\mu_{\mathbf{z}}$ as:

$$\mu_{\mathbf{z}}=\mathbb{E}[\mathbf{z}]=\mathcal{W}^* *_N \mathbb{E}[\mathcal{X}]=\mathcal{W}^* *_N \mathcal{M} \tag{98}$$

with variance $\sigma^2_{\mathbf{z}}$ and pseudo-variance $\tilde{\sigma}^2_{\mathbf{z}}$ found using properties (5) and (9) as

$$\begin{aligned}\sigma^2_{\mathbf{z}} &= \mathbb{E}[(\mathcal{W}^* *_N (\mathcal{X}-\mathcal{M}))\circ(\mathcal{W}^* *_N (\mathcal{X}-\mathcal{M}))^*]\\ &= \mathcal{W}^* *_N \mathbb{E}[(\mathcal{X}-\mathcal{M})\circ((\mathcal{X}-\mathcal{M})^*] *_N \mathcal{W}\\ &= \mathcal{W}^* *_N \mathcal{C} *_N \mathcal{W}\end{aligned} \tag{99}$$

$$\begin{aligned}\tilde{\sigma}^2_{\mathbf{z}} &= \mathbb{E}[(\mathcal{W}^* *_N (\mathcal{X}-\mathcal{M}))\circ(\mathcal{W}^* *_N (\mathcal{X}-\mathcal{M}))]\\ &= \mathcal{W}^* *_N \tilde{\mathcal{C}} *_N \mathcal{W}^*\end{aligned} \tag{100}$$

The characteristic function $\Phi_{\mathbf{z}}(\omega)$ of a Gaussian scalar $\mathbf{z}\sim\mathcal{CN}(\mu_{\mathbf{z}},\sigma^2_{\mathbf{z}},\tilde{\sigma}^2_{\mathbf{z}})$ is given by [77]:

$$\Phi_{\mathbf{z}}(\omega)=\exp\left\{i\Re(\omega^*\mu_{\mathbf{z}})-\frac{1}{4}\Big(\omega^*\sigma^2_{\mathbf{z}}\omega+\Re(\omega^*\tilde{\sigma}^2_{\mathbf{z}}\omega^*)\Big)\right\} \tag{101}$$

Now on putting $\mathbf{z}=\mathcal{W}^* *_N \mathcal{X}$ in (97) we get :

$$\Phi_{\mathcal{X}}(\mathcal{W}) = \mathbb{E}[\exp(i\Re(\omega^*\cdot\mathbf{z}))]\Big|_{\omega=1} = \Phi_{\mathbf{z}}(\omega)\Big|_{\omega=1} \tag{102}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2021.3084541, IEEE Open Journal of Signal Processing

On substituting (98), (99), (100) and $\omega = 1$ in (101),

$$\Phi_{\mathfrak{X}}(\mathcal{W}) = \exp\Big\{ i\Re(\mathcal{W}^* *_N \mathcal{M}) - \frac{1}{4}\Big(\mathcal{W}^* *_N \mathcal{C} *_N \mathcal{W} + \Re(\mathcal{W}^* *_N \tilde{\mathcal{C}} *_N \mathcal{W}^*)\Big)\Big\} \quad (103)$$

The characteristic function of an improper complex Gaussian vector as given in [77], [78] can be seen as a specific case of (103). Further, the characteristic function of $\mathfrak{X}$ given $\mathfrak{Y} = \mathcal{Y}$ can be written as (from (97)) :

$$\begin{aligned}\Phi_{\mathfrak{X}|\mathfrak{Y}}(\mathcal{W}) &= \mathbb{E}\Big[\exp\Big(i\Re(\mathcal{W}^* *_N \mathfrak{X})\Big) \mid \mathfrak{Y} = \mathcal{Y}\Big] \\ &= \mathbb{E}\Big[\exp\Big(i\Re(\mathcal{W}^* *_N (\mathfrak{X} - \hat{\mathfrak{X}}_{WL} + \hat{\mathfrak{X}}_{WL}))\Big) \mid \mathfrak{Y} = \mathcal{Y}\Big] \\ &= \mathbb{E}\Big[\underbrace{\exp(i\Re(\mathcal{W}^* *_N (\mathfrak{X} - \hat{\mathfrak{X}}_{WL})))}_{a} \\ &\quad \underbrace{\exp\Big(i\Re(\mathcal{W}^* *_N \hat{\mathfrak{X}}_{WL})\Big)}_{b} \mid \mathfrak{Y} = \mathcal{Y}\Big] \end{aligned} \quad (104)$$

Term b can be taken out of the expectation as given $\mathfrak{Y} = \mathcal{Y}$, we also know $\mathcal{Y}^*$, so $\hat{\mathfrak{X}}_{WL}$ which is given as $\mathcal{A}_1 *_M \mathfrak{Y} + \mathcal{A}_2 *_M \mathfrak{Y}^*$ becomes deterministic. In term a, the vector $(\mathfrak{X} - \hat{\mathfrak{X}}_{WL})$ is the error tensor $\mathfrak{E}$, hence we get :

$$\begin{aligned}\Phi_{\mathfrak{X}|\mathfrak{Y}}(\mathcal{W}) = \exp\Big(i\Re(\mathcal{W}^* *_N (\mathcal{A}_1 *_M \mathcal{Y} + \mathcal{A}_2 *_M \mathcal{Y}^*))\Big) \\ \mathbb{E}\Big[\exp\Big(i\Re(\mathcal{W}^* *_N \mathfrak{E})\Big) \mid \mathfrak{Y} = \mathcal{Y}\Big]\end{aligned} \quad (105)$$

Since the error is orthogonal to $\mathfrak{Y}$, we drop the conditioning.

$$\begin{aligned}\Phi_{\mathfrak{X}|\mathfrak{Y}}(\mathcal{W}) = \exp\Big(i\Re(\mathcal{W}^* *_N (\mathcal{A}_1 *_M \mathcal{Y} + \mathcal{A}_2 *_M \mathcal{Y}^*))\Big) \\ \underbrace{\mathbb{E}\Big[\exp\Big(i\Re(\mathcal{W}^* *_N \mathfrak{E})\Big)\Big]}_{\Phi_{\mathfrak{E}}(\mathcal{W})}\end{aligned} \quad (106)$$

Since $\mathfrak{X}$ and $\mathfrak{Y}$ are assumed zero mean and jointly Gaussian, so the error tensor $\mathfrak{E}$ will also be Gaussian with zero mean. Hence its characteristic function is given as (from (103)):

$$\Phi_{\mathfrak{E}}(\mathcal{W}) = \exp\Big\{-\frac{1}{4}\Big(\mathcal{W}^* *_N \mathcal{C}_{\mathfrak{E}} *_N \mathcal{W} + \Re(\mathcal{W}^* *_N \tilde{\mathcal{C}}_{\mathfrak{E}} *_N \mathcal{W}^*)\Big)\Big\} \quad (107)$$

Substituting (107) into (106) with the error covariance $\mathcal{C}_{\mathfrak{E}} = \mathcal{Q}_{WL}$ and pseudo-covariance $\tilde{\mathcal{C}}_{\mathfrak{E}} = \tilde{\mathcal{Q}}_{WL}$, we get :

$$\begin{aligned}\Phi_{\mathfrak{X}|\mathfrak{Y}}(\mathcal{W}) = \exp\Big\{ & i\Re(\mathcal{W}^* *_N (\mathcal{A}_1 *_M \mathcal{Y} + \mathcal{A}_2 *_M \mathcal{Y}^*)) \\ & - \frac{1}{4}\Big(\mathcal{W}^* *_N \mathcal{Q}_{WL} *_N \mathcal{W} + \Re(\mathcal{W}^* *_N \tilde{\mathcal{Q}}_{WL} *_N \mathcal{W}^*)\Big)\Big\}\end{aligned}$$

which is the characteristic function of a complex Gaussian tensor with mean $(\mathcal{A}_1 *_M \mathcal{Y} + \mathcal{A}_2 *_M \mathcal{Y}^*)$, covariance $\mathcal{Q}_{WL}$ and pseudo-covariance $\tilde{\mathcal{Q}}_{WL}$ (based on (103)). So we have shown that given $\mathfrak{Y} = \mathcal{Y}$, $\mathfrak{X}$ is conditionally $\mathcal{CN}(\mathcal{A}_1 *_M \mathcal{Y} + \mathcal{A}_2 *_M \mathcal{Y}^*, \mathcal{Q}_{WL}, \tilde{\mathcal{Q}}_{WL})$. Hence the best MMSE estimate which is the conditional mean is same as the widely multi-linear MMSE estimate for jointly complex Gaussian tensors.

Appendix D
Faster implementation of Newton's Iteration

1) Reducing the complexity using an alternate NM form

The Newton Method recursion from (11) is used to iteratively find the inverse of a tensor. Most often the objective of finding the inverse is to find the solution to a system of multi-linear equations represented by $\mathcal{A} *_N \mathcal{Y} = \mathcal{X}$, where $\mathcal{A} \in \mathbb{C}^{I_1 \times \ldots \times I_N \times I_1 \times \ldots \times I_N}$, and $\mathcal{X}, \mathcal{Y} \in \mathbb{C}^{I_1 \times \ldots \times I_N}$. If we use Newton's iteration from (11) to find $\mathcal{A}^{-1}$, each iteration requires computing Einstein product ($*_N$) between tensors of order $2N$. Hence the complexity per iteration will be cubic in the size of tensor, i.e. $\mathcal{O}((I_1 \cdots I_N)^3)$. However, employing an alternate form of the Newton's iteration can reduce this complexity from cubic to square in the size of the tensor. An alternate way to express Newton's method has been presented in [79]. Using the Einstein product, we can write (11) in expanded form linking $\mathcal{B}^{(k)}$ to $\mathcal{B}^{(0)}$ as :

$$\mathcal{B}^{(k)} = \sum_{m=1}^{2^k - 1} c_{k,m} (\mathcal{B}^{(0)} *_N \mathcal{A})^m *_N \mathcal{B}^{(0)} \quad (108)$$

where $c_{k,m}$ is the coefficient of the mth summation term in (108) and $\mathcal{B}^{(k)}$ is the approximation of $\mathcal{A}^{-1}$ at kth iteration. For an order $2N$ tensor, the notation $(\mathcal{A})^m$ denotes :

$$(\mathcal{A})^m = \underbrace{\mathcal{A} *_N \mathcal{A} *_N \cdots *_N \mathcal{A}}_{m \text{ times}} \quad (109)$$

Equation (108) can be seen as another form of the Newton's method. By considering $c_{k,m}$ as coefficients of a polynomial $f_k(z) = c_{k,0} z^0 + c_{k,1} z^1 + \cdots + c_{k,2^k-1} z^{2^k - 1}$, we can write $f_{k+1}(z) = 2 f_k(z) - z[f_k(z)]^2$ with $f_0(z) = 1$ [79]. Thus the coefficients $c_{k,m}$ can be found recursively. In fact these coefficients do not depend on the tensor to be inverted, so can be calculated before hand and used in the solution. Since the objective is to find $\mathcal{X} = \mathcal{A}^{-1} *_N \mathcal{Y}$, rather than approximating $\mathcal{A}^{-1}$ and then taking its Einstein product with $\mathcal{Y}$, we can find the approximation of $\mathcal{A}^{-1} *_N \mathcal{Y}$ directly. Take Einstein product with $\mathcal{Y}$ on both sides in (108) to get :

$$\mathcal{B}^{(k)} *_N \mathcal{Y} = \sum_{m=1}^{2^k - 1} c_{k,m} \underbrace{(\mathcal{B}^{(0)} *_N \mathcal{A})^m *_N \mathcal{B}^{(0)} *_N \mathcal{Y}}_{\tilde{\mathcal{Y}}^{(m)}} \quad (110)$$

where the left hand side is the approximation of $\mathcal{X}$ at kth iteration, and thus we can write :

$$\mathcal{X}^{(k)} = \sum_{m=1}^{2^k - 1} c_{k,m} \tilde{\mathcal{Y}}^{(m)} \quad (111)$$

Since $\tilde{\mathcal{Y}}^{(m)} = (\mathcal{B}^{(0)} *_N \mathcal{A})^m *_N \mathcal{B}^{(0)} *_N \mathcal{Y}$, we have :

$$\tilde{\mathcal{Y}}^{(m+1)} = (\mathcal{B}^{(0)} *_N \mathcal{A}) *_N \tilde{\mathcal{Y}}^{(m)} \quad (112)$$

Hence $\tilde{\mathcal{Y}}^{(m)} \in \mathbb{C}^{I_1 \times \ldots \times I_N}$ can be found recursively. Using such an approach the Newton's method now requires taking Einstein product over N modes between a tensor of order $2N$ with a tensor of order N in each iteration, thus reducing the complexity to $\mathcal{O}((I_1 \cdots I_N)^2)$ from $\mathcal{O}((I_1 \cdots I_N)^3)$. The complexity still remains exponential in the number of domains. However, using parallel processing of NM, the exponential

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2021.3084541, IEEE Open Journal of Signal Processing

dependence on number of domains can be brought to linear dependence as shown next.

2) Reducing the complexity using parallel processing

Each iteration in (11) requires taking Einstein product ($*_N$) between two tensors, say $\mathcal{A}$ and $\mathcal{B}$ of size $I_1 \times \ldots \times I_N \times I_1 \times \ldots \times I_N$. The Einstein product can be implemented using multiple parallel processors. Consider $\mathcal{C} = \mathcal{A} *_N \mathcal{B}$, then by definition of Einstein product we have :

$$\mathcal{C}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N} = \sum_{j_1,\ldots,j_N} \mathcal{A}_{i_1,\ldots,i_N,j_1,\ldots,j_N} \mathcal{B}_{j_1,\ldots,j_N,i'_1,\ldots,i'_N} \tag{113}$$

For ease of notation, given a fixed $i_1,\ldots,i_N,i'_1,\ldots,i'_N$, we represent $\mathcal{A}_{i_1,\ldots,i_N,j_1,\ldots,j_N}\mathcal{B}_{j_1,\ldots,j_N,i'_1,\ldots,i'_N}$ using $f_{j_1,\ldots,j_N}$. Summing over $f_{j_1,\ldots,j_N}$ for all the indices $(j_1,\ldots,j_N)$ gives us an element in tensor $\mathcal{C}$ denoted by $\mathcal{C}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}$ as per (113). Each $f_{j_1,\ldots,j_N}$ can be concurrently computed for a given $i_1,\ldots,i_N,i'_1,\ldots,i'_N$, using $(I_1 \cdots I_N)$ processors. Further parallel computation of these factors for all $i_1,\ldots,i_N,i'_1,\ldots,i'_N$ elements would require $(I_1 \cdots I_N)^3$ processors. All these parallelly computed results can be combined using a binary tree approach as shown in Figure 14.

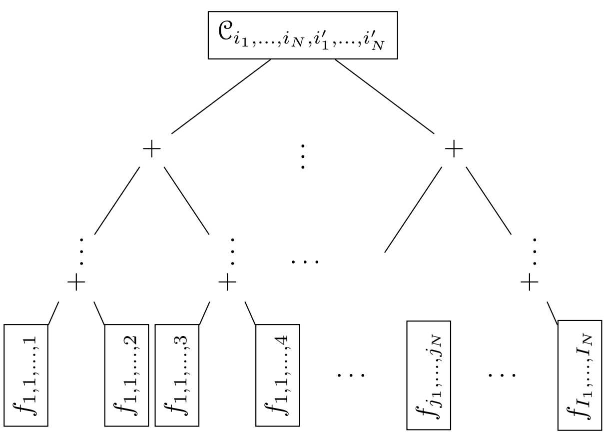

Fig. 14: Parallel execution of Einstein product

The height of the tree in Figure 14 is $\log(I_1 \cdots I_N)$. Hence assuming each level of the tree is computed parallelly, the time complexity of computing all the elements of $\mathcal{A} *_N \mathcal{B}$ using parallel processors is $\mathcal{O}(\log(I_1 \cdots I_N))$. Other than Einstein products, the NM iteration from (11) also requires one tensor subtraction which has a time complexity of $\mathcal{O}(1)$ using multiple processors. Hence the time complexity of each Newton iteration using parallel processors is $\mathcal{O}(\log(I_1 \cdots I_N))$. With $I_n = L$ for all n, the complexity is given as $\mathcal{O}(N \log L)$, which is linear in the number of domains.

References

[1] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Springer, 2nd Edition, 1994.

[2] T. G. Kolda and B. W. Bader, "Tensor Decompositions and Applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009. [Online]. Available: http://dx.doi.org/10.1137/07070111X

[3] A. Shashua and T. Hazan, "Non-negative Tensor Factorization with Applications to Statistics and Computer Vision," in *Proceedings of the 22nd International Conference on Machine Learning*, ser. ICML '05. New York, NY, USA: ACM, 2005, pp. 792–799. [Online]. Available: http://doi.acm.org/10.1145/1102351.1102451

[4] J. J. Guerrero, A. C. Murillo, and C. Sagues, "Localization and Matching Using the Planar Trifocal Tensor With Bearing-Only Data," *Trans. Rob.*, vol. 24, no. 2, pp. 494–501, Apr. 2008. [Online]. Available: http://dx.doi.org/10.1109/TRO.2008.918043

[5] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Transactions on Signal Processing*, vol. 48, no. 8, pp. 2377–2388, 2000.

[6] M. Bousse, O. Debals, and L. De Lathauwer, "A tensor-based method for large-scale blind source separation using segmentation," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 346–358, 2017.

[7] A. Cichocki, D. P. Mandic, A. H. Phan, C. F. Caiafa, G. Zhou, Q. Zhao, and L. D. Lathauwer, "Tensor Decompositions for Signal Processing Applications From Two-way to Multiway Component Analysis," *CoRR*, vol. abs/1403.4462, 2014. [Online]. Available: http://arxiv.org/abs/1403.4462

[8] P. Comon, "Tensors : A brief introduction," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 44–53, May 2014.

[9] L. Kuang, F. Hao, L. T. Yang, M. Lin, C. Luo, and G. Min, "A tensor-based approach for big data representation and dimensionality reduction," *IEEE Transactions on emerging topics in computing*, vol. 2, no. 3, pp. 280–291, 2014.

[10] N. Vervliet, O. Debals, L. Sorber, and L. De Lathauwer, "Breaking the curse of dimensionality using decompositions of incomplete tensors: Tensor-based scientific computing in big data analysis," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 71–79, 2014.

[11] M. Mardani, G. Mateos, and G. B. Giannakis, "Subspace learning and imputation for streaming big data matrices and tensors," *IEEE Transactions on Signal Processing*, vol. 63, no. 10, pp. 2663–2677, 2015.

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

[13] D. Muti and S. Bourennane, "Multidimensional filtering based on a tensor approach," *Signal Processing*, vol. 85, no. 12, pp. 2338–2353, 2005.

[14] L.-B. Cui, C. Chen, W. Li, and M. K. Ng, "An Eigenvalue problem for even order tensors with its applications," *Linear and Multilinear Algebra*, vol. 64, no. 4, pp. 602–621, 2016. [Online]. Available: https://doi.org/10.1080/03081087.2015.1071311

[15] A. Venugopal and H. Leib, "A tensor based framework for multi-domain communication systems," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 606–633, 2020.

[16] A. L. De Almeida, G. Favier, and L. R. Ximenes, "Space-Time-Frequency (STF) MIMO communication systems with blind receiver based on a generalized PARATUCK2 model," *IEEE Transactions on Signal Processing*, vol. 61, no. 8, pp. 1895–1909, 2013.

[17] G. Favier and A. L. F. de Almeida, "Tensor Space-Time-Frequency Coding With Semi-Blind Receivers for MIMO Wireless Communication Systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 22, pp. 5987–6002, Nov 2014.

[18] M. N. da Costa, G. Favier, and J. M. T. Romano, "Tensor modelling of MIMO communication systems with performance analysis and Kronecker receivers," *Signal Processing*, vol. 145, pp. 304–316, 2018.

[19] C. Chen, A. Surana, A. Bloch, and I. Rajapakse, "Multilinear Time Invariant System Theory," in *2019 Proceedings of the Conference on Control and its Applications*, 2019, pp. 118–125. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611975758.18

[20] C. Chen, A. Surana, A. M. Bloch, and I. Rajapakse, "Multilinear Control Systems Theory," *SIAM Journal on Control and Optimization*, vol. 59, no. 1, pp. 749–776, 2021.

[21] F. Roemer and M. Haardt, "Tensor-based channel estimation (TENCE) for two-way relaying with multiple antennas and spatial reuse," in *Acoustics, Speech and Signal Processing, ICASSP. IEEE International Conference on*. IEEE, 2009, pp. 3641–3644.

[22] X. Han, A. L. de Almeida, and Z. Yang, "Channel estimation for MIMO multi-relay systems using a tensor approach," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, p. 163, 2014.

[23] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 810–823, 2000.

[24] C. A. Fernandes, G. Favier, and J. C. Mota, "PARAFAC-based channel estimation and data recovery in nonlinear MIMO spread spectrum communication systems," *Signal Processing*, vol. 91, no. 2, pp. 311–322, 2011.

[25] A. de Baynast, L. De Lathauwer, and B. Aazhang, "Blind PARAFAC receivers for multiple access-multiple antenna systems," in *2003 IEEE*

58th Vehicular Technology Conference. VTC 2003-Fall (IEEE Cat. No. 03CH37484), vol. 2. IEEE, 2003, pp. 1128–1132.

[26] D. C. Araújo, A. L. De Almeida, J. P. Da Costa, and R. T. de Sousa, "Tensor-Based Channel Estimation for Massive MIMO-OFDM Systems," *IEEE Access*, vol. 7, pp. 42 133–42 147, 2019.

[27] Y. Lin, S. Jin, M. Matthaiou, and X. You, "Tensor-Based Channel Estimation for Millimeter Wave MIMO-OFDM With Dual-Wideband Effects," *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4218–4232, 2020.

[28] D. Muti and S. Bourennane, "Multidimensional estimation based on a tensor decomposition," in *IEEE Workshop on Statistical Signal Processing, 2003*. IEEE, 2003, pp. 98–101.

[29] M. Brazell, N. Li, C. Navasca, and C. Tamon, "Solving Multilinear Systems via Tensor Inversion," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 2, pp. 542–570, 2013. [Online]. Available: https://doi.org/10.1137/100804577

[30] J. Marot, C. Fossati, and S. Bourennane, "About advances in tensor data denoising methods," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 1, p. 235357, 2008.

[31] R. Tong, G. Bao, and Z. Ye, "A Higher Order Subspace Algorithm for Multichannel Speech Enhancement," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 2004–2008, 2015.

[32] T. Lin and S. Bourennane, "Survey of hyperspectral image denoising methods based on tensor decompositions," *EURASIP journal on Advances in Signal Processing*, vol. 2013, no. 1, pp. 1–11, 2013.

[33] T. Yokota, Q. Zhao, and A. Cichocki, "Smooth PARAFAC Decomposition for Tensor Completion," *IEEE Transactions on Signal Processing*, vol. 64, no. 20, pp. 5423–5436, 2016.

[34] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and R. S. Blum, "Low-Rank Tensor Decomposition-Aided Channel Estimation for Millimeter Wave MIMO-OFDM Systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1524–1538, 2017.

[35] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-Valued Signal Processing: The Proper Way to Deal With Impropriety," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5101–5125, Nov 2011.

[36] D. Pandey and H. Leib, "Tensor multi-linear mmse estimation using the einstein product," in *Advances in Information and Communication (FICC 2021)*, K. Arai, Ed. Cham: Springer International Publishing, 2021, pp. 47–64.

[37] M. lin Liang, B. Zheng, and R. juan Zhao, "Tensor inversion and its application to the tensor equations with Einstein product," *Linear and Multilinear Algebra*, vol. 67, no. 4, pp. 843–870, 2019. [Online]. Available: https://doi.org/10.1080/03081087.2018.1500993

[38] I. Kisil, G. G. Calvi, B. S. Dees, and D. P. Mandic, "Tensor decompositions and practical applications: A hands-on tutorial," in *Recent Trends in Learning From Data*. Springer, 2020, pp. 69–97.

[39] A. L. F. De Almeida, "Tensor Modeling and Signal Processing for Wireless Communication Systems," Theses, Université de Nice Sophia Antipolis, Nov. 2007. [Online]. Available: https://tel.archives-ouvertes.fr/tel-00460157

[40] Q.-W. Wang and X. Xu, "Iterative algorithms for solving some tensor equations," *Linear and Multilinear Algebra*, vol. 67, no. 7, pp. 1325–1349, 2019. [Online]. Available: https://doi.org/10.1080/03081087.2018.1452889

[41] A. Ben-Israel, "A note on an iterative method for generalized inversion of matrices," *Mathematics of Computation*, vol. 20, no. 95, pp. 439–440, 1966.

[42] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge University Press, 2010.

[43] A. van den Bos, "The multivariate complex normal distribution-a generalization," *IEEE Transactions on Information Theory*, vol. 41, no. 2, pp. 537–539, 1995.

[44] F. D. Neeser and J. L. Massey, "Proper Complex Random Processes with applications to Information Theory," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1293–1302, Jul 1993.

[45] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Transactions on Signal Processing*, vol. 43, pp. 2030–2033, 1995.

[46] L. Sun, B. Zheng, C. Bu, and Y. Wei, "Moore–Penrose inverse of tensors via Einstein product," *Linear and Multilinear Algebra*, vol. 64, no. 4, pp. 686–698, 2016. [Online]. Available: https://doi.org/10.1080/03081087.2015.1083933

[47] D. Muti, S. Bourennane, and J. Marot, "Lower-rank tensor approximation and multiway filtering," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1172–1204, 2008. [Online]. Available: https://doi.org/10.1137/060653263

[48] V. Pan and J. Reif, "Fast and efficient parallel solution of dense linear systems," *Computers and Mathematics with Applications*, vol. 17, no. 11, pp. 1481–1491, 1989. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0898122189900813

[49] B. Huang and W. Li, "Numerical subspace algorithms for solving the tensor equations involving Einstein product," *Numerical Linear Algebra with Applications*, vol. 28, no. 2, p. e2351, 2021.

[50] R. Behera and D. Mishra, "Further results on generalized inverses of tensors via the Einstein product," *Linear and Multilinear Algebra*, vol. 65, no. 8, pp. 1662–1682, 2017.

[51] K. Panigrahy and D. Mishra, "Extension of Moore–Penrose inverse of tensor via Einstein product," *Linear and Multilinear Algebra*, vol. 0, no. 0, pp. 1–24, 2020. [Online]. Available: https://doi.org/10.1080/03081087.2020.1748848

[52] B. Huang, "Numerical study on Moore-Penrose inverse of tensors via Einstein product," *Numerical Algorithms*, pp. 1–31, 2021.

[53] J. K. Sahoo and R. Behera, "Reverse-order law for core inverse of tensors," *Computational and Applied Mathematics*, vol. 39, no. 2, pp. 1–22, 2020.

[54] N. Costa and S. Haykin, "A Novel Wideband MIMO Channel Model and Experimental Validation," *IEEE Transactions on Antennas and Propagation*, vol. 56, no. 2, pp. 550–562, Feb 2008.

[55] D. Pandey and H. Leib, "A Tensor based Precoder and Receiver for MIMO GFDM systems," in *IEEE International Conference on Communications, ICC'21 (accepted)*. IEEE, 2021.

[56] A. Cichocki, "Era of big data processing: A new approach via tensor networks and tensor decompositions," *arXiv preprint arXiv:1403.2048*, 2014.

[57] I. V. Oseledets, "Tensor-train decomposition," *SIAM Journal on Scientific Computing*, vol. 33, no. 5, pp. 2295–2317, 2011.

[58] X. Wang, L. T. Yang, Y. Wang, L. Ren, and M. J. Deen, "ADTT: A Highly Efficient Distributed Tensor-Train Decomposition Method for IIoT Big Data," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 1573–1582, 2021.

[59] H. Liu, L. T. Yang, Y. Guo, X. Xie, and J. Ma, "An Incremental Tensor-Train Decomposition for Cyber-Physical-Social Big Data," *IEEE Transactions on Big Data*, pp. 1–1, 2018.

[60] J. Zhang, X. Ma, J. Qi, and S. Jin, "Designing Tensor-Train Deep Neural Networks For Time-Varying MIMO Channel Estimation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 759–773, 2021.

[61] Y. Zniyed, R. Boyer, A. L. de Almeida, and G. Favier, "Tensor train representation of MIMO channels using the JIRAFE method," *Signal Processing*, vol. 171, p. 107479, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168420300220

[62] Y. Zniyed, R. Boyer, A. L. F. de Almeida, and G. Favier, "Tensor-Train Modeling for MIMO OFDM Tensor Coding-and-Forwarding Relay Systems," in *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5.

[63] H. Liu, L. T. Yang, J. Ding, Y. Guo, and S. S. Yau, "Tensor-train-based high-order dominant eigen decomposition for multimodal prediction services," *IEEE Transactions on Engineering Management*, vol. 68, no. 1, pp. 197–211, 2021.

[64] "5G NR : Physical channels and modulation (3GPP TS 38.211 version 15.8.0 Release 15) ," European Telecommunications Standards Institute, Tech. Rep., 2020. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/138200_138299/138211/15.08.00_60/ts_138211v150800p.pdf

[65] A. Stamoulis, S. N. Diggavi, and N. Al-Dhahir, "Intercarrier interference in MIMO OFDM," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2451–2464, Oct 2002.

[66] A. Burg, S. Haene, D. Perels, P. Luethi, N. Felber, and W. Fichtner, "Algorithm and VLSI architecture for linear MMSE detection in MIMO-OFDM systems," in *IEEE International Symposium on Circuits and Systems*. IEEE, 2006.

[67] D. N. Liu and M. P. Fitz, "Low complexity affine MMSE detector for iterative detection-decoding MIMO OFDM systems," *IEEE Transactions on Communications*, vol. 56, no. 1, pp. 150–158, 2008.

[68] M. Wu, B. Yin, G. Wang, C. Dick, J. R. Cavallaro, and C. Studer, "Large-scale MIMO detection for 3GPP LTE: Algorithms and FPGA implementations," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 916–929, 2014.

[69] K. Kim and H. Park, "Modified successive interference cancellation for MIMO OFDM on doubly selective channels," in *VTC Spring 2009-IEEE 69th Vehicular Technology Conference*. IEEE, 2009, pp. 1–5.

[70] L. Rugini, P. Banelli, K. Fang, and G. Leus, "Enhanced turbo MMSE equalization for MIMO-OFDM over rapidly time-varying frequency-

selective channels," in *2009 IEEE 10th Workshop on Signal Processing Advances in Wireless Communications*. IEEE, 2009, pp. 36–40.

[71] D. Li, B. Guo, and K. Yamashita, "A Novel Channel Estimation and Equalization for MIMO-OFDM Systems," *IEEJ Transactions on Electronics, Information and Systems*, vol. 126, no. 3, pp. 339–346, 2006.

[72] W. Yi and H. Leib, "OFDM symbol detection integrated with channel multipath gains estimation for doubly-selective fading channels," *Physical communication*, vol. 22, pp. 19–31, 2017.

[73] K. A. D. Teo and S. Ohno, "Pilot-Aided Channel Estimation and Viterbi Equalization for OFDM over Doubly-Selective Channel," in *IEEE Globecom 2006*, Nov 2006, pp. 1–5.

[74] A. Alimohammad, S. F. Fard, B. F. Cockburn, and C. Schlegel, "An improved SOS-based fading channel emulator," in *2007 IEEE 66th Vehicular Technology Conference*. IEEE, 2007, pp. 931–935.

[75] P. Mukunthan and P. Dananjayan, "PAPR Reduction based on a Modified PTS with Interleaving and Pulse Shaping method for STBC MIMO-OFDM System," in *2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12)*. IEEE, 2012, pp. 1–6.

[76] H. H. Andersen, M. Hojbjerre, D. Sorensen, and P. S. Eriksen, *Linear and graphical models: for the multivariate complex normal distribution*. Springer Science & Business Media, 1995, vol. 101.

[77] B. Picinbono, "Second-order complex random vectors and normal distributions," *IEEE Transactions on Signal Processing*, vol. 44, no. 10, pp. 2637–2640, Oct 1996.

[78] G. R. Ducharme, P. L. De Micheaux, and B. Marchina, "The complex multinormal distribution, quadratic forms in complex random vectors and an omnibus goodness-of-fit test for the complex normal distribution," *Annals of the Institute of Statistical Mathematics*, vol. 68, no. 1, pp. 77–104, 2016.

[79] C.-Y. Hsu and W.-R. Wu, "Low-Complexity ICI Mitigation Methods for High-Mobility SISO/MIMO-OFDM Systems," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 6, pp. 2755–2768, 2009.

Divyanshu Pandey received the B.Tech degree in Communication and Computer Engineering from the LNM Institute of Information Technology, Jaipur, India, in 2011 and the M.S. degree in Electrical Engineering from University of Minnesota, Twin Cities, USA, in 2014. He worked as a Systems Engineer in WLAN PHY R&D team at Marvell Semiconductors Inc., Santa Clara, CA, USA from February 2015 to August 2017. He is currently pursuing the Ph.D. degree in Electrical Engineering from McGill University, Montreal, QC, Canada. His research interests include Information Theory, Wireless Communications, and Tensor Algebra with applications to communications and signal processing.

Harry Leib received the B.Sc. (cum laude) and M.Sc. degrees in Electrical Engineering from the Technion - Israel Institute of Technology, Israel in 1977 and 1984 respectively. In 1987 he received the Ph.D. degree in Electrical Engineering from the University of Toronto, Canada.

During 1977-1984 he was with the Israel Ministry of Defense, working in Communication Systems. After completing his Ph.D. studies, he was with the University of Toronto as a Post-doctoral Research Associate and as an Assistant Professor. Since September 1989 he has been with the Department of Electrical and Computer Engineering at McGill University, where he is now a Full Professor. His current research activities are in the areas of Digital Communications, Wireless Communication Systems, Global Navigation Satellite Systems, Detection, Estimation, and Information Theory.

Dr. Leib was Editor for the IEEE Transactions on Communications 2000-2013, and Associate Editor for the IEEE Transactions on Vehicular Technology 2001-2007. He was a guest co-editor for special issues of the IEEE Journal on Selected Areas in Communication on "Differential and Noncoherent Wireless Communication" 2003-2005, and on "Spectrum and Energy Efficient Design of Wireless Communication Networks" 2012-2013. Since 2017 he has been the founding Editor-in-Chief of AIMS Electronics and Electrical Engineering Journal.