Channel Estimation for Beyond Diagonal RIS via Tensor Decomposition

André L. F. de Almeida, *Senior Member, IEEE,* Bruno Sokal, *Member, IEEE,*
Hongyu Li, *Graduate Student Member, IEEE,* and Bruno Clerckx, *Fellow, IEEE*

Abstract—This paper addresses the channel estimation problem for beyond diagonal reconfigurable intelligent surface (BD-RIS) from a tensor decomposition perspective. We first show that the received pilot signals can be arranged as a three-way tensor, allowing us to recast the cascaded channel estimation problem as a block Tucker decomposition problem that yields decoupled estimates for the involved channel matrices while offering a substantial performance gain over the conventional (matrix-based) least squares (LS) estimation method. More specifically, we develop two solutions to solve the problem. The first one is a closed-form solution that extracts the channel estimates via a block Tucker Kronecker factorization (BTKF), which boils down to solving a set of parallel rank-one matrix approximation problems. Exploiting such a low-rank property yields a noise rejection gain compared to the standard LS estimation scheme while allowing the two involved channels to be estimated separately. The second solution is based on a block Tucker alternating least squares (BTALS) algorithm that directly estimates the involved channel matrices using an iterative estimation procedure. We discuss the uniqueness and identifiability issues and their implications for training design. We also propose a tensor-based design of the BD-RIS training tensor for each algorithm that ensures unique decoupled channel estimates under trivial scaling ambiguities. Our numerical results shed light on the tradeoffs offered by BTKF and BTALS methods. Specifically, while the first enjoys fast and parallel extraction of the channel estimates in closed form, the second has a more flexible training design, allowing for a significantly reduced training overhead compared to the state-of-the-art LS method.

Index Terms—Beyond diagonal reconfigurable intelligent surfaces, channel estimation, tensor decomposition, alternating least squares, Kronecker factorization.

I. INTRODUCTION

As a new advance of conventional reconfigurable intelligent surface (RIS) techniques with diagonal phase shift matrices [1], [2], beyond diagonal (BD) RIS has been recently proposed and theoretically proved to achieve enhanced channel gain and enlarged coverage [3]–[5]. The benefits of BD-RIS are enabled by interconnecting elements via additional tunable components to mathematically generate scattering matrices with nonzero off-diagonal entries, increasing flexibility to manipulate waves. The fundamental modeling and architecture (group/fully-connected) design of BD-RIS based on the circuit topology has been first studied in [6]. Following [6],

other architectures (forest/tree-connected) with reduced circuit complexity yet satisfactory performance have been proposed based on the graph theory [7]. In [8], a frequency-dependent model is proposed to optimize a BD-RIS architecture for multi-band MIMO networks. Meanwhile, to enlarge the coverage, BD-RIS with hybrid transmissive and reflective mode and multi-sector mode have been proposed based on the antenna array arrangement [3].

It is worth noting that the enhanced performance of BD-RIS architectures and modes depends highly on the accuracy of the channel state information (CSI). However, it is difficult to effectively and efficiently acquire the CSI for BD-RIS-aided wireless systems for the following reasons. *First*, since the BD-RIS is nearly passive without the ability to sense signals, a straightforward strategy is to estimate the *combined* channel constructed by the transmitter-RIS and RIS-user channels, as well studied in conventional RIS literature [9]–[11], [12], [13]. This strategy relies on the pre-design of BD-RIS for pilot training, while each BD-RIS architecture leads to unique mathematical constraints of the scattering matrix, which indicates that the design for conventional RISs does not work for BD-RIS architectures. *Second*, each BD-RIS architecture leads to unique constructions of the combined channel with increasing dimensions, which requires additional training overhead to obtain the CSI. To solve the above two challenges, one recent work [5] has proposed a closed-form solution based on the least squares (LS) estimation to pre-design the BD-RIS with group/fully-connected architectures. Nevertheless, there are two limitations in [5]. *First*, the combined CSI is obtained at the cost of a large training overhead, which grows heavily with the circuit complexity of BD-RIS architectures. *Second*, the built-in block Kronecker structure of the combined channel is ignored, which could be further exploited to facilitate channel estimation. Thus, estimating the BD-RIS-aided channels with high accuracy and low training overhead remains an important yet challenging open problem.

To address the above limitations, this paper studies channel estimation for BD-RIS via a tensor decomposition perspective. Tensor methods have been successfully exploited in different knowledge areas, including signal processing and machine learning, as well as in wireless communications and multi-sensor processing (see [14]–[18] and references therein). Recently, tensor modeling has also found applications in RIS-assisted communications [19]–[23]. Here, we derive efficient channel estimation methods for BD-RIS that leverage the tensor decomposition structure of the received pilots to

provide accurate CSI acquisition of each individual channel while operating at low training overheads. Specifically, we first show that the combined channel for BD-RIS can be formulated as a (block) Tucker tensor decomposition. Then, decoupled estimates for the involved channel matrices can be obtained by exploiting the different unfoldings of the received pilot tensor. We develop two solutions to solve this problem. The first one is a closed-form solution that extracts the individual channel estimates via a block Tucker Kronecker factorization (BTKF), which boils down to solving a set of parallel rank-one matrix approximation problems. The second one is based on a block Tucker alternating least squares (BTALS) algorithm that directly estimates the involved channels using an iterative estimation procedure. The proposed algorithms offer substantial performance gains over conventional LS estimation while operating at much lower training overheads by capitalizing on the tensor signal structure.

The contributions of this paper are summarized as follows:

First, we link the channel estimation problem for BD-RIS to a tensor decomposition problem. Specifically, we show that the received pilot signals can be organized as a three-dimensional (3D) array or a third-order tensor that follows a (block) Tucker decomposition model. In addition, we discuss the implications of the specific BD-RIS architecture to the resulting tensor decomposition structure.

Second, we propose two tensor decomposition-based channel estimation schemes for BD-RIS that capitalize on tensor modeling. The BTKF method is a closed-form scheme that yields decoupled estimates of the involved channels after an LS estimation step by solving a block-Kronecker factorization problem that boils down to solving rank-one approximation problems based on the 3-mode unfolding of the received pilot tensor. The second method, BTALS, directly estimates the channel matrices separately using the 1-mode and 2-mode unfoldings of the received pilot tensor.

Third, we discuss the trade-offs involving the channel estimation methods and their implications for the training design. On the one hand, we show that the BTKF method effectively exploits the inherent Kronecker structure of the combined channel to provide a more accurate reconstruction than the reference LS method, thanks to the noise rejection property of the channel separation step. On the other hand, by efficiently exploiting the tensor decomposition structure of the received pilots, BTALS yields decoupled channel estimates with a much lower training overhead, which can be orders of magnitude smaller than that of LS and BTKF methods. Such savings in training resources offered by BTALS are even more pronounced for BD-RIS configurations with higher levels of couplings among the scattering elements.

Fourth, we study the identifiability conditions and uniqueness associated with the proposed algorithms and their implications for training design. We propose a new training design for the BD-RIS scattering matrix and derive the structure of the BD-RIS training tensor used in each proposed channel estimation method. The proposed designs fulfill the physical constraints of the BD-RIS architecture and the identifiability conditions of the associated block Tucker model.

Finally, simulation results show the superiority of the proposed BTKF and BTALS algorithms compared to the baseline LS estimator while highlighting the trade-offs involving these methods in terms of normalized mean square error (NMSE) performance, required training overhead, and computational complexity. In particular, we show that BTKF and BTALS offer performance gains over the reference LS method thanks to decoupling the estimated channel matrices, yielding a more accurate reconstruction of the combined channel. For a group-connected BD-RIS architecture with Q groups, the BTKF method achieves channel separation by solving a set of Q rank-one matrix approximations in parallel. On the other hand, the BTALS method directly estimates the involved channel matrices by intertwining the estimation of the transmitter-RIS and RIS-receiver channels in an iterative way using an alternating least squares procedure. Our results also show that the NMSE performance of the estimated channels with group-connected BD-RIS architectures is the same as that with conventional RIS in some scenarios.

This work is organized as follows. Section II provides the basic material and definitions related to tensor decomposition, along with the main notations and properties. Section III describes the system model and discusses the baseline LS method and BD-RIS design. Section IV gives a detailed presentation of the tensor modeling of the received pilot signals. Section V formulates the proposed channel estimation methods. This section also discusses computational complexity and uniqueness issues. Section VI describes the proposed BD-RIS tensor design. Numerical results are discussed in Section VII, and the paper is concluded in Section VIII.

II. TENSOR PREREQUISITES

In this section, we provide the useful notations and main operators used in this papers as well an overview of the Tucker decomposition, in which will be used to develop the proposed channel estimation algorithm.

A. Notation and properties

Scalars are represented as non-bold lower-case letters a, column vectors as lower-case boldface letters $\boldsymbol{a}$, matrices as upper-case boldface letters $\boldsymbol{A}$, and tensors as calligraphic upper-case letters $\boldsymbol{\mathcal{A}}$. The superscripts $\{\cdot\}^{\mathrm{T}}$, $\{\cdot\}^{*}$, $\{\cdot\}^{\mathrm{H}}$ and $\{\cdot\}^{+}$ stand for transpose, conjugate, conjugate transpose, and pseudo-inverse operations, respectively. An identity matrix of dimension K is denoted as $\boldsymbol{I}_{K}$. The operator $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm of a matrix or tensor, and $\mathbb{E}\{\cdot\}$ is the expectation operator. Given a matrix $\boldsymbol{A} \in \mathbb{C}^{I \times R}$, the operator $\mathrm{D}_{i}(\boldsymbol{A})$ defines a diagonal matrix of size $R \times R$ constructed from the i-th row of $\boldsymbol{A}$, for $i \in \{1, \ldots, I\}$. From a set of Q matrices $\boldsymbol{X}^{(q)} \in \mathbb{C}^{M \times N}$, $q = \{1, \ldots, Q\}$, we can construct a block diagonal matrix as $\boldsymbol{X} = \mathrm{blkdiag}\big(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(Q)}\big) \in \mathbb{C}^{MQ \times NQ}$. Moreover, $\mathrm{vec}\,(\boldsymbol{A})$ converts $\boldsymbol{A} \in \mathbb{C}^{I \times R}$ to a column vector $\boldsymbol{a} \in \mathbb{C}^{IR \times 1}$ by stacking its columns on top of each other, while the operator $\mathrm{unvec}\big(\boldsymbol{a}\big)_{I \times R}$ returns to the matrix $\boldsymbol{A} \in \mathbb{C}^{I \times R}$. The symbols $\circ$, $\otimes$, $\diamond$, and $|\otimes|$ denote the outer product, the Kronecker product, the Khatri-Rao product (also known as the column-wise Kronecker product), and the block Kronecker product, respectively. The Khatri-Rao

product of matrices $X \in \mathbb{C}^{I \times R}$ and $Y \in \mathbb{C}^{J \times R}$, is defined as $Z = X \diamond Y = [x_1 \otimes y_1, \ldots, x_R \otimes y_R] \in \mathbb{C}^{JI \times R}$, where x_r and y_r are the r-th column of X and Y, respectively, $r = 1, \ldots, R$. Likewise, let us define $H = [H^{(1)}, \ldots, H^{(Q)}] \in \mathbb{C}^{M \times LQ}$ and $G = [G^{(1)}, \ldots, G^{(Q)}] \in \mathbb{C}^{N \times LQ}$, matrices formed each by Q block matrices, i.e., $H^{(q)} \in \mathbb{C}^{M \times L}$ and $G^{(q)} \in \mathbb{C}^{N \times L}$, $q = 1, \ldots, Q$. The block Kronecker product[1] between H and G, denoted as $W = H |\otimes| G$, is given by

$$W = [H^{(1)} \otimes G^{(1)}, \ldots, H^{(Q)} \otimes G^{(Q)}] \in \mathbb{C}^{NM \times L^2 Q}. \quad (1)$$

We also use the following property of the Kroneker product

$$\text{vec}(ABC) = (C^{\mathrm{T}} \otimes A)\,\text{vec}(B), \quad (2)$$

where the involved matrices have compatible dimensions.

B. Slices and unfoldings

Consider a set of matrices $Y_k \in \mathbb{C}^{I \times J}$, $\forall k = 1, \ldots, K$. Concatenating all K matrices, we form the third-order tensor $\mathcal{Y} = Y_1 \sqcup_3 Y_2 \sqcup_3 \ldots \sqcup_3 Y_K \in \mathbb{C}^{I \times J \times K}$, where $\sqcup_3$ indicates a concatenation along the third dimension. We can interpret Y_k as the k-th frontal slice of $\mathcal{Y}$, defined as the matrix $\mathcal{Y}_{..k} = Y_k \in \mathbb{C}^{I \times J}$. This matrix is built by varying the first and second dimensions for a fixed third-dimension index k. The tensor $\mathcal{Y}$ can be *matricized* by letting one dimension vary along the rows and the remaining two dimensions along the columns. From $\mathcal{Y}$, we can form three different matrices, referred to as the *n-mode unfolding*, $n = 1, 2, 3$, which can be respectively obtained as a function of the frontal slices as

$$[\mathcal{Y}]_{(1)} = [\mathcal{Y}_{..1}, \ldots, \mathcal{Y}_{..K}] \in \mathbb{C}^{I \times JK}, \quad (3)$$

$$[\mathcal{Y}]_{(2)} = [\mathcal{Y}_{..1}^{\mathrm{T}}, \ldots, \mathcal{Y}_{..K}^{\mathrm{T}}] \in \mathbb{C}^{J \times IK}, \quad (4)$$

$$[\mathcal{Y}]_{(3)} = [\text{vec}(\mathcal{Y}_{..1}), \ldots, \text{vec}(\mathcal{Y}_{..K})]^{\mathrm{T}} \in \mathbb{C}^{K \times IJ}. \quad (5)$$

For convenience, we can also refer to the unfolding operation as $[\mathcal{Y}]_{(n)} = \text{unfold}(\mathcal{Y}, n)$, $n = 1, 2, 3$. The n-mode product, denoted as "$\times_n$", defines the multiplication between a tensor $\mathcal{Y}$ and a matrix A, leading to a tensor $\mathcal{Z}$ with compatible dimensions, i.e., $\mathcal{Z} = \mathcal{Y} \times_n A$. It can be computed by pre-multiplying the n-mode unfolding of $\mathcal{Y}$ by the matrix A, i.e., $[\mathcal{Z}]_{(n)} = A[\mathcal{Y}]_{(n)}$. For example, the 1-mode product between $\mathcal{Y} \in \mathbb{C}^{I \times J \times K}$ and $A \in \mathbb{C}^{L \times I}$ yields $\mathcal{Z} = \mathcal{Y} \times_1 A \in \mathbb{C}^{L \times J \times K}$. It can be computed by $[\mathcal{Z}]_{(1)} = A[\mathcal{Y}]_{(1)} \in \mathbb{C}^{L \times JK}$.

C. Tucker decomposition

The Tucker decomposition [25] defines the concept of multilinear transformation. For a third-order tensor $\mathcal{Y} \in \mathbb{C}^{I \times J \times K}$, it expresses the tensor as multiple sums of rank-one tensor components, which can be defined as

$$\mathcal{Y} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} g_{r_1, r_2, r_3} a_{r_1} \circ b_{r_2} \circ c_{r_3}, \quad (6)$$

where $a_{r_1} \in \mathbb{C}^{I \times 1}$, $b_{r_2} \in \mathbb{C}^{J \times 1}$, and $c_{r_3} \in \mathbb{C}^{K \times 1}$ are the column vectors of the factor matrices $A \in \mathbb{C}^{I \times R_1}$,

[1]The block Kronecker product defined in (1) is also referred to in the literature as the Khatri-Rao product between partitioned matrices [24].

$B \in \mathbb{C}^{J \times R_2}$, and $C \in \mathbb{C}^{K \times R_3}$, respectively, and $\mathcal{G} \in \mathbb{C}^{R_1 \times R_2 \times K}$ is referred to as the *core tensor*, with typical element $g_{r_1, r_2, r_3} \doteq [\mathcal{G}]_{r_1, r_2, r_3}$. Adopting the n-mode product notation, the Tucker decomposition can be written as

$$\mathcal{Y} = \mathcal{G} \times_1 A \times_2 B \times_3 C \quad (7)$$

A special case is the Tucker-2 decomposition, where one of its factor matrices equals the identity matrix, e.g., $C = I_K \in \mathbb{R}^{K \times K}$ (with $K = R_3$). In this case, (7) simplifies to

$$\mathcal{Y} = \mathcal{G} \times_1 A \times_2 B \quad (8)$$

The 3-mode (frontal) slices $\mathcal{Y}_{..k} \in \mathbb{C}^{I \times J}$ can be expressed as

$$\mathcal{Y}_{..k} = A \mathcal{G}_{..k} B^{\mathrm{T}} \in \mathbb{C}^{I \times J}, \ k = 1, \ldots, K. \quad (9)$$

By properly staking these frontal slices according to equations (3)-(5), the three matrix unfolding of the Tucker-2 decomposition can be factorized as

$$[\mathcal{Y}]_{(1)} = A[\mathcal{G}]_{(1)} (I_K \otimes B)^{\mathrm{T}} \in \mathbb{C}^{I \times JK}, \quad (10)$$

$$[\mathcal{Y}]_{(2)} = B[\mathcal{G}]_{(2)} (I_K \otimes A)^{\mathrm{T}} \in \mathbb{C}^{J \times IK}, \quad (11)$$

$$[\mathcal{Y}]_{(3)} = [\mathcal{G}]_{(3)} (B \otimes A)^{\mathrm{T}} \in \mathbb{C}^{K \times IJ}. \quad (12)$$

The Tucker decomposition is not unique due to rotational freedom involving the factor matrices and the core tensor. Indeed, the multiplication of each factor matrix by a nonsingular matrix is compensated by transforming the corresponding modes of the core tensor by the inverse of these matrices without changing the output tensor [25]. However, when the core tensor $\mathcal{G}$ is known, a unique estimation of the factor matrices under trivial scaling ambiguities is possible under certain conditions [26], which is the case of this work.

It is worth mentioning that the well-known parallel factor (PARAFAC) decomposition, also known as the canonical polyadic decomposition (CPD) [27], [14], is a special case of the Tucker decomposition, in which $R_1 = R_2 = R_3 = R$ while the core tensor reduces to an identity tensor. In this case, we have $\mathcal{Y} = \mathcal{I}_{3, R} \times_1 A \times_2 B \times_3 C \in \mathbb{C}^{I \times J \times K}$, with $A \in \mathbb{C}^{I \times R}$, $B \in \mathbb{C}^{J \times R}$, and $C \in \mathbb{C}^{K \times R}$ being the associated factor matrices, where R denotes the tensor rank.

III. SYSTEM MODEL

Let us consider a multiple-input multiple-output (MIMO) system assisted by a beyond diagonal reconfigurable intelligent surface (BD-RIS), as illustrated in Fig. 1, where the transmitter and the receiver are equipped with M_T and M_R antennas, respectively, and a BD-RIS with N elements. For simplification, we assume that the direct link between the transmitter and the receiver is blocked.

A. Signal and channel models

We adopt a two-timescale protocol, where the transmission consists of K blocks of consecutive T time slots each. We assume that the length-T pilot sequences are repeatedly transmitted through each block, while the BD-RIS response repeats itself within one block and varies among blocks. This training protocol follows the idea proposed in [19] for diagonal
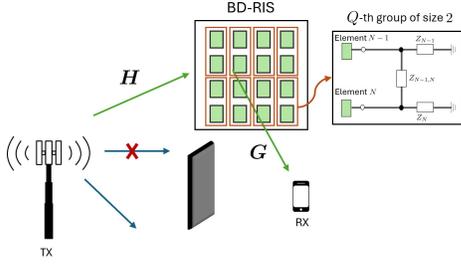
Fig. 1: A communication system aided by a BD-RIS.

RIS, in which the scattering matrix of the BD-RIS is fixed during the time window corresponding to a block of T time slots while varying between blocks. The received pilot signal at the t-th time slot and k-th block is $\bar{\boldsymbol{y}}_{t,k} = \boldsymbol{G}\boldsymbol{S}_k\boldsymbol{H}^{\mathrm{T}}\boldsymbol{x}_t + \bar{\boldsymbol{b}}_{t,k} \in \mathbb{C}^{M_R \times 1}$, where $\boldsymbol{H} \in \mathbb{C}^{M_T \times N}$ and $\boldsymbol{G} \in \mathbb{C}^{M_R \times N}$ are the TX-RIS and RIS-RX channels, respectively, and $\boldsymbol{S}_k \in \mathbb{C}^{N \times N}$ is the BD-RIS scattering matrix associated with the k-th training block, with $\boldsymbol{S}_k^{\mathrm{H}}\boldsymbol{S}_k = \boldsymbol{I}_N$. Collecting the T time slots of the k-th block yields $\bar{\boldsymbol{Y}}_k = \boldsymbol{G}\boldsymbol{S}_k\boldsymbol{H}^{\mathrm{T}}\boldsymbol{X} + \bar{\boldsymbol{B}}_k \in \mathbb{C}^{M_R \times T}$. The matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T] \in \mathbb{C}^{M_T \times T}$ collects the transmitted pilot symbols during the T time slots, and $\bar{\boldsymbol{B}}_k = [\bar{\boldsymbol{b}}_{1,k}, \ldots, \bar{\boldsymbol{b}}_{T,k}] \in \mathbb{C}^{M_R \times T}$ is the additive noise at the receiver modeled as a complex Gaussian random variable with zero mean and unitary variance, i.e., $\sim \mathcal{CN}(\boldsymbol{0}, M_R\boldsymbol{I}_{M_R})$.

Assuming that the transmitter sends orthogonal pilot sequences, which requires $T \geq M_T$, and after matched-filtering using the known pilots, we get

$$\boldsymbol{Y}_k = \boldsymbol{G}\boldsymbol{S}_k\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{B}_k \in \mathbb{C}^{M_R \times T}, \qquad (13)$$

where $\boldsymbol{Y}_k \doteq \bar{\boldsymbol{Y}}_k\boldsymbol{X}^{\mathrm{H}}$ and $\boldsymbol{B}_k = \bar{\boldsymbol{B}}_k\boldsymbol{X}^{\mathrm{H}}$ are the filtered pilot signals and noise. Considering a group-connected BD-RIS architecture [6], the N reflecting elements are divided into Q groups, each with $\bar{N}$ elements connected to each other, i.e., $N = \bar{N} \cdot Q$. In this case, the scattering matrix is expressed as $\boldsymbol{S}_k = \mathrm{blkdiag}(\boldsymbol{S}_k^{(1)}, \ldots \boldsymbol{S}_k^{(Q)}) \in \mathbb{C}^{N \times N}$, where the q-th matrix $\boldsymbol{S}_k^{(q)} \in \mathbb{C}^{\bar{N} \times \bar{N}}$ satisfies $\boldsymbol{S}_k^{(q)\mathrm{H}}\boldsymbol{S}_k^{(q)} = \boldsymbol{I}_{\bar{N}}$. Hence, (13) translates into a sum of Q blocks

$$\boldsymbol{Y}_k = \sum_{q=1}^{Q} \boldsymbol{G}^{(q)}\boldsymbol{S}_k^{(q)}\boldsymbol{H}^{(q)\mathrm{T}} + \boldsymbol{B}_k \in \mathbb{C}^{M_R \times T}, \qquad (14)$$

where $\boldsymbol{H}^{(q)} \in \mathbb{C}^{M_T \times \bar{N}}$ and $\boldsymbol{G}^{(q)} \in \mathbb{C}^{M_R \times \bar{N}}$ correspond to the q-th block of $\boldsymbol{H} \in \mathbb{C}^{M_T \times \bar{N}Q}$ and $\boldsymbol{G} \in \mathbb{C}^{M_R \times \bar{N}Q}$, respectively, defined as follows

$$\boldsymbol{H}^{(q)} = \boldsymbol{H}_{.,[(q-1)\bar{N}+1,\ldots,q\bar{N}]} \in \mathbb{C}^{M_T \times \bar{N}}, \ q = 1, \ldots, Q, \quad (15)$$
$$\boldsymbol{G}^{(q)} = \boldsymbol{G}_{.,[(q-1)\bar{N}+1,\ldots,q\bar{N}]} \in \mathbb{C}^{M_R \times \bar{N}}, \ q = 1, \ldots, Q. \quad (16)$$

Hence, the channel matrices can be seen as a concatenation of smaller submatrices such that $\boldsymbol{H} = [\boldsymbol{H}^{(1)}, \ldots, \boldsymbol{H}^{(Q)}] \in \mathbb{C}^{M_T \times \bar{N}Q}$ and $\boldsymbol{G} = [\boldsymbol{G}^{(1)}, \ldots, \boldsymbol{G}^{(Q)}] \in \mathbb{C}^{M_R \times \bar{N}Q}$.

B. Least squares channel estimation

We start by recalling the conventional LS channel estimation as a reference for the proposed solutions. Defining $\boldsymbol{y}_k =$

$\mathrm{vec}(\boldsymbol{Y}_k) \in \mathbb{C}^{M_R M_T \times 1}$, and using (2), the noiseless vectorized received signal at the k-th block can be expressed as

$$\boldsymbol{y}_k = \mathrm{vec}\Big(\sum_{q=1}^{Q} \boldsymbol{G}^{(q)}\boldsymbol{S}_k^{(q)}\boldsymbol{H}^{(q)\mathrm{T}} \Big) = \sum_{q=1}^{Q} \big(\boldsymbol{H}^{(q)} \otimes \boldsymbol{G}^{(q)} \big) \mathrm{vec}\big(\boldsymbol{S}_k^{(q)} \big)$$
$$= (\boldsymbol{H} \,|\!\otimes\!|\, \boldsymbol{G})\mathrm{vec}(\bar{\boldsymbol{S}}_k),$$

where $\bar{\boldsymbol{S}}_k = [\mathrm{vec}(\boldsymbol{S}_k^{(1)}), \ldots, \mathrm{vec}(\boldsymbol{S}_k^{(Q)})] \in \mathbb{C}^{\bar{N}^2 \times Q}$, and

$$\boldsymbol{H} \,|\!\otimes\!|\, \boldsymbol{G} = \big[\boldsymbol{H}^{(1)} \otimes \boldsymbol{G}^{(1)}, \ldots, \boldsymbol{H}^{(Q)} \otimes \boldsymbol{G}^{(Q)} \big] \in \mathbb{C}^{M_R M_T \times N},$$

is the combined block-Kronecker-structured MIMO channel matrix that concatenates the combined channels associated with the Q BD-RIS groups. Defining $\boldsymbol{T} \doteq \boldsymbol{H} \,|\!\otimes\!|\, \boldsymbol{G}$ and $\bar{\boldsymbol{S}} = [\mathrm{vec}(\bar{\boldsymbol{S}}_1), \ldots, \mathrm{vec}(\bar{\boldsymbol{S}}_K)]^{\mathrm{T}} \in \mathbb{C}^{K \times \bar{N}^2 Q}$ and collecting the received signal over K blocks we have

$$\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_K] = \boldsymbol{T}\bar{\boldsymbol{S}}^{\mathrm{T}} + \boldsymbol{B} \in \mathbb{C}^{M_R M_T \times K}, \qquad (17)$$

where $\boldsymbol{B}$ is the corresponding noise term. An estimate of the combined channel $\boldsymbol{T}$ can be obtained by right-filtering using the known BD-RIS training matrix, i.e., $\hat{\boldsymbol{T}} = \boldsymbol{Y}(\bar{\boldsymbol{S}}^{\mathrm{T}})^{\dagger}$ as a solution to the following least squares (LS) problem

$$\hat{\boldsymbol{T}} = \underset{\boldsymbol{T}}{\mathrm{argmin}} \big\| \boldsymbol{Y} - \boldsymbol{T}\bar{\boldsymbol{S}}^{\mathrm{T}} \big\|_F^2, \qquad (18)$$

where $\hat{\boldsymbol{T}} \approx \hat{\boldsymbol{H}} \,|\!\otimes\!|\, \hat{\boldsymbol{G}}$ is an estimate of the combined channel. Note that assuming an orthogonal design for BD-RIS training matrix $\bar{\boldsymbol{S}} \in \mathbb{C}^{K \times \bar{N}^2 Q}$, the solution is found by simplified matched filtering [5]. In this case, the estimate of the combined channel can also be found as $\hat{\boldsymbol{T}} = \boldsymbol{Y}\bar{\boldsymbol{S}}^*$. The LS solution requires $K \geq \bar{N}^2 Q$ to ensure a unique estimation of the combined channel. This constraint may be too restrictive, especially for a moderate number of scattering elements, due to the quadratic dependency on the number of connected BD-RIS elements in each group.

C. BD-RIS matrix design and motivation

Although the design proposed in [5] is optimal in the mean square error (MSE) sense, LS channel estimation based on (18) ignores the built-in block Kronecker structure of the effective MIMO channel since only an estimate of the "combined" channel is obtained. Hence, it requires a large training overhead, which has a quadratic growth with the group size $\bar{N}$. As discussed in the next sections, the block Kronecker product structure linking the involved channel matrices can be efficiently exploited, allowing us to obtain enhanced channel estimates compared to the baseline LS solution. Additionally, it turns out that the received signal in (14) can be recast using a tensor modeling approach. This is possible by reformulating the training design and rebuilding the BD-RIS training structure as a third-order tensor. Then, by resorting to tensor decomposition algorithms and capitalizing on their intrinsic uniqueness properties, decoupled estimates of the individual channels $\boldsymbol{G}$ and $\boldsymbol{H}$ can be obtained with a significantly reduced training overhead and improved accuracy. Additionally, as will be shown later (Section VI), the proposed BD-RIS training design is flexible to allow operation under more challenging system setups with $K \ll \bar{N}^2 Q$, which have not yet been considered in the literature.

IV. TENSOR SIGNAL MODELING

In this section, we recast the signal model for the received pilots using a tensor decomposition approach. In correspondence with the background material presented in Section II, we formulate the tensor signal models by starting from the general (fully-connected) case for presentation convenience. Then, we discuss the tensor signal formulation for the group-connected case, which is of practical interest due to its lower implementation complexity. The main expressions derived in this section will be exploited later in the derivation of the proposed channel estimation methods.

A. General fully-connected case

Starting from the signal model in equation (13) and omitting the noise term for notation convenience, the received pilot signal at the k-th block can be expressed as

$$Y_k = GS_k H^{\mathrm{T}} \in \mathbb{C}^{M_R \times M_T}. \tag{19}$$

By analogy with equation (9), we can interpret the k-th received pilot signal matrix as the k-th frontal slice of the *received pilot tensor* $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{M_R \times M_T \times K}$ that follows a Tucker-2 decomposition [25], [26]. Using (8) with the correspondences $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{\mathcal{G}}) \leftrightarrow (\boldsymbol{G}, \boldsymbol{H}, \boldsymbol{I}_K, \boldsymbol{\mathcal{S}})$, we can express the received pilot signal tensor using the n-mode product notation as

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{S}} \times_1 \boldsymbol{G} \times_2 \boldsymbol{H}, \tag{20}$$

where $\boldsymbol{\mathcal{Y}} = \boldsymbol{Y}_1 \sqcup_3 \boldsymbol{Y}_2 \sqcup_3 \dots \sqcup_3 \boldsymbol{Y}_K \in \mathbb{C}^{M_R \times M_T \times K}$, and $\boldsymbol{\mathcal{S}} = \boldsymbol{S}_1 \sqcup_3 \boldsymbol{S}_2 \sqcup_3 \dots \sqcup_3 \boldsymbol{S}_K \in \mathbb{C}^{N \times N \times K}$. Note that $\boldsymbol{\mathcal{S}}$ results from concatenating the BD-RIS scattering matrices along the third dimension. We refer to $\boldsymbol{\mathcal{S}}$ as the *BD-RIS training tensor*.

For the fully-connected case and adopting the Tucker representation in (20), in correspondence with (10), (11), and (12), we can deduce the following matrix unfoldings for the received pilot signal tensor:

$$[\boldsymbol{\mathcal{Y}}]_{(1)} = \boldsymbol{G}[\boldsymbol{\mathcal{S}}]_{(1)} (\boldsymbol{I}_K \otimes \boldsymbol{H})^{\mathrm{T}} \in \mathbb{C}^{M_R \times M_T K}, \tag{21}$$

$$[\boldsymbol{\mathcal{Y}}]_{(2)} = \boldsymbol{H}[\boldsymbol{\mathcal{S}}]_{(2)} (\boldsymbol{I}_K \otimes \boldsymbol{G})^{\mathrm{T}} \in \mathbb{C}^{M_T \times M_R K}, \tag{22}$$

$$[\boldsymbol{\mathcal{Y}}]_{(3)} = [\boldsymbol{\mathcal{S}}]_{(3)} (\boldsymbol{H} \otimes \boldsymbol{G})^{\mathrm{T}} \in \mathbb{C}^{K \times M_R M_T}, \tag{23}$$

where $[\boldsymbol{\mathcal{S}}]_{(n)}$ is the n-mode unfolding of the BD-RIS training tensor, $n = 1, 2, 3$. Following (3), (4), and (5), these matrix unfoldings are respectively given by

$$[\boldsymbol{\mathcal{S}}]_{(1)} = [\boldsymbol{\mathcal{S}}_{..1}, \dots, \boldsymbol{\mathcal{S}}_{..K}] \in \mathbb{C}^{N \times NK}, \tag{24}$$

$$[\boldsymbol{\mathcal{S}}]_{(2)} = [\boldsymbol{\mathcal{S}}_{..1}^{\mathrm{T}}, \dots, \boldsymbol{\mathcal{S}}_{..K}^{\mathrm{T}}] \in \mathbb{C}^{N \times NK}, \tag{25}$$

$$[\boldsymbol{\mathcal{S}}]_{(3)} = [\mathrm{vec}(\boldsymbol{\mathcal{S}}_{..1}), \dots, \mathrm{vec}(\boldsymbol{\mathcal{S}}_{..K})]^{\mathrm{T}} \in \mathbb{C}^{K \times N^2}, \tag{26}$$

where $\boldsymbol{\mathcal{S}}_{..k} \in \mathbb{C}^{N \times N}$, the k-th frontal slice of the tensor $\boldsymbol{\mathcal{S}} \in \mathbb{C}^{N \times N \times K}$, corresponds to the BD-RIS scattering matrix associated with the k-th training block, $k = 1, \dots, K$.

B. Group-connected case

Let us now consider the group-connected architecture. Starting from the signal model in (14), and omitting the noise term, we have

$$Y_k = \sum_{q=1}^{Q} \boldsymbol{G}^{(q)} \boldsymbol{S}_k^{(q)} \boldsymbol{H}^{(q)\mathrm{T}}. \tag{27}$$

In this case, the received pilot tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{M_R \times M_T \times K}$ corresponds to a block Tucker-2 decomposition, i.e., we can rewrite (20) as a sum of Q tensor blocks

$$\boldsymbol{\mathcal{Y}} = \sum_{q=1}^{Q} \boldsymbol{\mathcal{S}}^{(q)} \times_1 \boldsymbol{G}^{(q)} \times_2 \boldsymbol{H}^{(q)}, \tag{28}$$

where $\boldsymbol{\mathcal{S}}^{(q)} \in \mathbb{C}^{\bar{N} \times \bar{N} \times K}$ is the BD-RIS training tensor associated with the q-th group.

Group-connected architectures are commonly adopted due to implementation complexity. In this case, the BD-RIS training tensor is "sparse" due to its block-diagonal structure. Figure 2 illustrates the decomposition of received pilot signals in tensor form, corresponding to a Tucker-2 decomposition composed of Q blocks. This decomposition can also be viewed as a special block-term decomposition (BTD), more specifically, also referred to as a "type-2 BTD", which represents a tensor into a sum of rank-$(\bar{N}, \bar{N}, \cdot)$ tensor blocks [24], [28]. From this figure, we can see that the assumption of a group-connected architecture implies a BD-RIS training tensor having a "block-diagonal" structure, the k-th frontal slice of which corresponds to the BD-RIS training matrix associated with the k-th block. Note that a fully-connected BD-RIS architecture is equivalent to having a fully dense BD-RIS training tensor in Figure 2, which does not have zero off-diagonal blocks. Although we focus on the group-connected case based on (28), any (sub-connected) architecture can be captured by the general tensor representation in (20), the difference being in the structure of the BD-RIS training tensor $\boldsymbol{\mathcal{S}}$.

Similarly to (24)-(26), we construct the unfoldings of the q-th group BD-RIS tensor as

$$[\boldsymbol{\mathcal{S}}^{(q)}]_{(1)} = [\boldsymbol{\mathcal{S}}_{..1}^{(q)}, \dots, \boldsymbol{\mathcal{S}}_{..K}^{(q)}] \in \mathbb{C}^{\bar{N} \times \bar{N} K}, \tag{29}$$

$$[\boldsymbol{\mathcal{S}}^{(q)}]_{(2)} = [\boldsymbol{\mathcal{S}}_{..1}^{(q)\mathrm{T}}, \dots, \boldsymbol{\mathcal{S}}_{..K}^{(q)\mathrm{T}}] \in \mathbb{C}^{\bar{N} \times \bar{N} K}, \tag{30}$$

$$[\boldsymbol{\mathcal{S}}^{(q)}]_{(3)} = [\mathrm{vec}(\boldsymbol{\mathcal{S}}_{..1}^{(q)}), \dots, \mathrm{vec}(\boldsymbol{\mathcal{S}}_{..K}^{(q)})]^{\mathrm{T}} \in \mathbb{C}^{K \times \bar{N}^2}. \tag{31}$$

Hence, for the group-connected case, equivalent expressions can be obtained for the unfoldings of the received pilot tensor. The corresponding expressions for the 1-mode and 2-mode matrix unfoldings of the received pilot tensor can be obtained by rewriting (21) and (22) as a sum of Q blocks

$$[\boldsymbol{\mathcal{Y}}]_{(1)} = \boldsymbol{G} \cdot \underbrace{\mathrm{blkdiag}\left([\boldsymbol{\mathcal{S}}^{(1)}]_{(1)}, \dots, [\boldsymbol{\mathcal{S}}^{(Q)}]_{(1)}\right)}_{S_1} \begin{bmatrix} \boldsymbol{I}_K \otimes \boldsymbol{H}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{I}_K \otimes \boldsymbol{H}_Q^{\mathrm{T}} \end{bmatrix}$$

$$[\boldsymbol{\mathcal{Y}}]_{(2)} = \boldsymbol{H} \cdot \underbrace{\mathrm{blkdiag}\left([\boldsymbol{\mathcal{S}}^{(1)}]_{(2)}, \dots, [\boldsymbol{\mathcal{S}}^{(Q)}]_{(2)}\right)}_{S_2} \begin{bmatrix} \boldsymbol{I}_K \otimes \boldsymbol{G}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{I}_K \otimes \boldsymbol{G}_Q^{\mathrm{T}} \end{bmatrix}$$
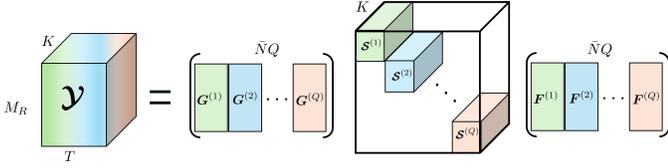
Fig. 2: Illustration of the decomposition of the noiseless 3-D received signal tensor $\boldsymbol{\mathcal{Y}}$ of dimensions $M_R \times T \times K$ for a group-connected BD-RIS communication system. The resulting receiver pilot tensor can be viewed as a sum of Q tensor blocks. The full BD-RIS training tensor is given by a block-diagonal concatenation of Q component tensors $\boldsymbol{\mathcal{S}}^{(q)}$, $q = 1, \ldots, Q$, each of dimensions $\bar{N} \times \bar{N} \times K$.

or, compactly,

$$[\boldsymbol{\mathcal{Y}}]_{(1)} = \boldsymbol{G} \boldsymbol{S}_1 \left(\overline{\boldsymbol{I}} \,|\otimes| \, \boldsymbol{H} \right)^{\mathrm{T}}, \qquad (32)$$

$$[\boldsymbol{\mathcal{Y}}]_{(2)} = \boldsymbol{H} \boldsymbol{S}_2 \left(\overline{\boldsymbol{I}} \,|\otimes| \, \boldsymbol{G} \right)^{\mathrm{T}}, \qquad (33)$$

where

$$\boldsymbol{S}_1 \doteq \mathrm{blkdiag}\left([\boldsymbol{\mathcal{S}}^{(1)}]_{(1)}, \ldots, [\boldsymbol{\mathcal{S}}^{(Q)}]_{(1)} \right) \in \mathbb{C}^{\bar{N}Q \times \bar{N}KQ} \quad (34)$$

$$\boldsymbol{S}_2 \doteq \mathrm{blkdiag}\left([\boldsymbol{\mathcal{S}}^{(1)}]_{(2)}, \ldots, [\boldsymbol{\mathcal{S}}^{(Q)}]_{(2)} \right) \in \mathbb{C}^{\bar{N}Q \times \bar{N}KQ} \quad (35)$$

with $\overline{\boldsymbol{I}} \doteq [\boldsymbol{I}_K, \ldots, \boldsymbol{I}_K] \in \mathbb{C}^{K \times KQ}$. Likewise, the 3-mode matrix unfolding of the received pilot tensor can be obtained by rewriting (23) as a sum of Q blocks

$$[\boldsymbol{\mathcal{Y}}]_{(3)} = \underbrace{\left[[\boldsymbol{\mathcal{S}}^{(1)}]_{(3)}, \ldots, [\boldsymbol{\mathcal{S}}^{(Q)}]_{(3)} \right]}_{\boldsymbol{S}_3} \begin{bmatrix} \boldsymbol{H}_1^{\mathrm{T}} \otimes \boldsymbol{G}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{H}_Q^{\mathrm{T}} \otimes \boldsymbol{G}_Q^{\mathrm{T}} \end{bmatrix}$$

$$= \boldsymbol{S}_3 (\boldsymbol{H} \,|\otimes| \, \boldsymbol{G})^{\mathrm{T}}, \qquad (36)$$

where

$$\boldsymbol{S}_3 \doteq \left[[\boldsymbol{\mathcal{S}}^{(1)}]_{(3)}, \ldots, [\boldsymbol{\mathcal{S}}^{(Q)}]_{(3)} \right] \in \mathbb{C}^{K \times \bar{N}^2 Q}, \qquad (37)$$

is referred to as the *compact 3-mode unfolding* that concatenates the 3-mode unfoldings of the Q tensors along its columns. This matrix is key to deriving our first channel estimation method, which relies on channel separation after the LS estimation step using a Kronecker-product factorization method. Likewise, the 1-mode and 2-mode unfoldings of the received pilot tensor, given by the two expressions in (32)-(33), are the basis for the formulation of the second algorithm, which yields direct decoupled estimations of the two involved channels using an iterative algorithm, as will be shown later.

Remark 1: The pairs of expressions (21)-(22) and (32)-(33) representing the 1-mode and 2-mode unfoldings are equivalent and interchangeable due to the following property

$$(\overline{\boldsymbol{I}} \,|\otimes| \, \boldsymbol{H}) = (\boldsymbol{I}_K \otimes \boldsymbol{H})\boldsymbol{P} \text{ and } (\overline{\boldsymbol{I}} \,|\otimes| \, \boldsymbol{G}) = (\boldsymbol{I}_K \otimes \boldsymbol{G})\boldsymbol{P},$$

$$\boldsymbol{S}_n = [\boldsymbol{\mathcal{S}}]_{(n)}\boldsymbol{P}, \ n = 1, 2, \qquad (38)$$

where $\boldsymbol{P} \doteq [\boldsymbol{I}_Q \otimes \boldsymbol{e}_1^{(K)}, \ldots, \boldsymbol{I}_Q \otimes \boldsymbol{e}_K^{(K)}] \otimes \boldsymbol{I}_{\bar{N}}$ is a permutation matrix of dimensions $\bar{N}QK \times \bar{N}KQ$ and $\boldsymbol{e}_k^{(K)}$ is a unit vector corresponding to the k-th column of $\boldsymbol{I}_K$. This equivalence implies that the 1-mode and 2-mode expressions in (32)-(33) are equivalent to those in (21)-(22) since the first ones are just permuted versions of the latter. Hence, both are valid for formulating the block Tucker alternating least squares

(BTALS) channel estimation method. On the other hand, one should note that the 3-mode unfolding expression in (37) provides a more compact representation than the one in (23) when considering a group-connected architecture. This is due to the fact that working with the 3-mode unfolding of the BD-RIS tensor $\boldsymbol{\mathcal{S}} \in \mathbb{C}^{\bar{N}Q \times \bar{N}Q \times K}$ introduced unnecessary zeros due to its block structure (see Figure 2). $\boldsymbol{S}_3$ is instead a concatenation of the 3-mode unfoldings of Q tensor blocks. Consequently, it is not a permuted version of $[\boldsymbol{\mathcal{S}}]_{(3)}$ (as opposed to the 1-mode and 2-mode unfoldings). Indeed, the first is a $K \times \bar{N}^2 Q$ matrix, while the second is a longer and sparse $K \times \bar{N}^2 Q^2$ matrix whose number of columns is increased by a factor of Q. The difference in the structures of $\{\boldsymbol{S}_1, \boldsymbol{S}_2\}$ and $\boldsymbol{S}_3$ is a consequence of the "asymmetric" structure of the block Tucker model in (28) with respect to its third mode, which is an identity matrix and hence does not depend on the number Q of blocks. Hence, in the next section, we will adopt the compact 3-mode expression in (37) to formulate the block Tucker Kronecker factorization (BTKF) algorithm for channel estimation.

Remark 2: It is easy to see that the tensor model for conventional (diagonal) RIS corresponds to a special case of (20), where $\boldsymbol{\mathcal{S}} \in \mathbb{C}^{N \times N \times K}$ is such that its K frontal slices $\boldsymbol{\mathcal{S}}_{..1}, \ldots, \boldsymbol{\mathcal{S}}_{..K}$ are diagonal matrices. In tensor notation, this is equivalent to the reduced representation $\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{I}}_{3,N} \times_3 \boldsymbol{S}$, where $\boldsymbol{S} \in \mathbb{C}^{K \times N}$ is the diagonal RIS training matrix holding the set of N phase shifts of each training block along its K rows. In this case, the tensor signal model (20) reduces to

$$\boldsymbol{\mathcal{Y}} = (\boldsymbol{\mathcal{I}}_{3,N} \times_3 \boldsymbol{S}) \times_1 \boldsymbol{G} \times_2 \boldsymbol{H} = \boldsymbol{\mathcal{I}}_{3,N} \times_1 \boldsymbol{G} \times_2 \boldsymbol{H} \times_3 \boldsymbol{S},$$

which corresponds to a PARAFAC tensor model for the received pilots. It is clear that the proposed BD-RIS Tucker model generalizes the diagonal RIS PARAFAC model. For further details on tensor modeling and algorithms for the diagonal RIS case, we refer the reader to [19].

V. TENSOR-BASED CHANNEL ESTIMATION METHODS

In this section, we formulate the proposed channel estimation methods by capitalizing on the tensor signal structures discussed in the previous section. Two algorithms are proposed to solve the channel estimation problem *via* decoupling the estimates of the involved channel matrices $\boldsymbol{G}$ and $\boldsymbol{H}$. The first one, BTKF, is a closed-form solution that extracts the channel estimates via a block Kronecker factorization problem, which boils down to solving a set of rank-one matrix approximation problems. The second one is termed BTALS, which directly estimates the channel matrices using an iterative procedure. We also discuss identifiability and uniqueness issues and their overall implications for system design. The design of the BD-RIS training tensor is a separate topic that will be discussed in the next section.

A. Block Tucker Kronecker factorization algorithm (BTKF) algorithm for decoupled channel estimation

Taking into consideration the noise term, we recall the received signal model in tensor form as

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{S}} \times_1 \boldsymbol{G} \times_2 \boldsymbol{H} + \boldsymbol{\mathcal{B}}, \qquad (39)$$

where $\mathcal{B} \in \mathbb{C}^{M_R \times M_T \times K}$ is the additive noise tensor.

Closed-form estimates of $\boldsymbol{H}$ and $\boldsymbol{G}$ can be obtained by exploiting the block Kronecker structure of the compact 3-mode unfolding of the received pilot signal tensor in (36), which is given by $[\mathcal{Y}]_{(3)} = \boldsymbol{S}_3 (\boldsymbol{H} |\otimes| \boldsymbol{G})^{\mathrm{T}} + [\mathcal{B}]_{(3)}$. Assuming $K \geq \bar{N}^2 Q$, a right-filtering of the received pilot signal gives

$$\boldsymbol{Z} = \left(\boldsymbol{S}_3^{\dagger} [\mathcal{Y}]_{(3)} \right)^{\mathrm{T}} \approx \boldsymbol{H} |\otimes| \boldsymbol{G} \in \mathbb{C}^{M_R M_T \times \bar{N}^2 Q}. \quad (40)$$

To obtain decoupled estimates of the involved channel matrices $\boldsymbol{H}$ and $\boldsymbol{G}$ from the filtered signal in (40), we formulate the following optimization problem

$$\{\hat{\boldsymbol{H}}, \hat{\boldsymbol{G}}\} = \operatorname*{argmin}_{\boldsymbol{H}, \boldsymbol{G}} \| \boldsymbol{Z} - \boldsymbol{H} |\otimes| \boldsymbol{G} \|_{\mathrm{F}}^2. \quad (41)$$

By making use of the partition of the BD-RIS phase shifts into Q groups, one can easily note that

$$[\boldsymbol{Z}^{(1)}, \ldots, \boldsymbol{Z}^{(Q)}] \approx [\boldsymbol{H}^{(1)} \otimes \boldsymbol{G}^{(1)}, \ldots, \boldsymbol{H}^{(Q)} \otimes \boldsymbol{G}^{(Q)}], \quad (42)$$

where $\boldsymbol{Z}^{(q)} = \boldsymbol{Z}_{.,[(q-1)\bar{N}^2+1,\ldots,q\bar{N}^2]}$ is the q-th block matrix of the filtered signal, $q = 1, \ldots, Q$. From such a block structure, we can recast this problem as Q independent sub-problems executed in parallel, with the q-th problem being defined as

$$\{\hat{\boldsymbol{H}}^{(q)}, \hat{\boldsymbol{G}}^{(q)}\} = \operatorname*{argmin}_{\boldsymbol{H}^{(q)}, \boldsymbol{G}^{(q)}} \left\| \boldsymbol{Z}^{(q)} - \boldsymbol{H}^{(q)} \otimes \boldsymbol{G}^{(q)} \right\|_{\mathrm{F}}^2, \quad (43)$$

$q = 1, \ldots, Q$. This problem can be solved using the classical nearest Kronecker approximation method originally proposed in [29], a closed-form approach that yields the best factorization of the Kronecker product of two matrices.

Specifically, by properly permuting the elements of $\boldsymbol{Z}^{(q)}$, problem (43) can be recast as a simple rank-one matrix approximation, such that it can be rewritten as

$$\{\hat{\boldsymbol{h}}^{(q)}, \hat{\boldsymbol{g}}^{(q)}\} = \operatorname*{argmin}_{\boldsymbol{h}^{(q)}, \boldsymbol{g}^{(q)}} \left\| \overline{\boldsymbol{Z}}^{(q)} - \boldsymbol{g}^{(q)} \boldsymbol{h}^{(q)\mathrm{T}} \right\|_{\mathrm{F}}^2, \quad (44)$$

where $\overline{\boldsymbol{Z}}^{(q)} \in \mathbb{C}^{M_R \bar{N} \times M_T \bar{N}}$ is a reshaped version of $\boldsymbol{Z}^{(q)} \in \mathbb{C}^{M_R M_T \times \bar{N}^2}$, $\boldsymbol{g}^{(q)} = \operatorname{vec}(\boldsymbol{G}^{(q)}) \in \mathbb{C}^{M_R \bar{N} \times 1}$, and $\boldsymbol{h}^{(q)} = \operatorname{vec}(\boldsymbol{H}^{(q)}) \in \mathbb{C}^{M_T \bar{N} \times 1}$. Defining the truncated singular value decomposition (SVD) of $\overline{\boldsymbol{Z}}^{(q)} = \boldsymbol{U}^{(q)} \boldsymbol{\Sigma}^{(q)} \boldsymbol{V}^{\mathrm{H}(q)}$, the estimates of $\boldsymbol{h}^{(q)}$ and $\boldsymbol{g}^{(q)}$ corresponding respectively to the dominant left and right singular vectors $\boldsymbol{u}_1^{(q)} = \boldsymbol{U}_{.1}^{(q)} \in \mathbb{C}^{M_R \bar{N} \times 1}$ and $\boldsymbol{v}_1^{*(q)} = \boldsymbol{V}_{.1}^{*(q)} \in \mathbb{C}^{M_T \bar{N} \times 1}$

$$\hat{\boldsymbol{g}}^{(q)} = \boldsymbol{u}_1^{(q)}, \quad \hat{\boldsymbol{G}}^{(q)} = \operatorname{unvec}(\hat{\boldsymbol{g}}^{(q)})_{M_R \times \bar{N}}, \quad (45)$$

$$\hat{\boldsymbol{h}}^{(q)} = \boldsymbol{v}_1^{*(q)}, \quad \hat{\boldsymbol{H}}^{(q)} = \operatorname{unvec}(\hat{\boldsymbol{h}}^{(q)})_{M_T \times \bar{N}}. \quad (46)$$

The global estimates of the channel matrices are then formed by collecting the Q estimated blocks

$$\hat{\boldsymbol{G}} = [\hat{\boldsymbol{G}}^{(1)}, \ldots, \hat{\boldsymbol{G}}^{(Q)}] \in \mathbb{C}^{M_R \times \bar{N} Q}, \quad (47)$$

$$\hat{\boldsymbol{H}} = [\hat{\boldsymbol{H}}^{(1)}, \ldots, \hat{\boldsymbol{H}}^{(Q)}] \in \mathbb{C}^{M_T \times \bar{N} Q}. \quad (48)$$

A block diagram of the main steps of the BTFK algorithm is provided in Figure 3.
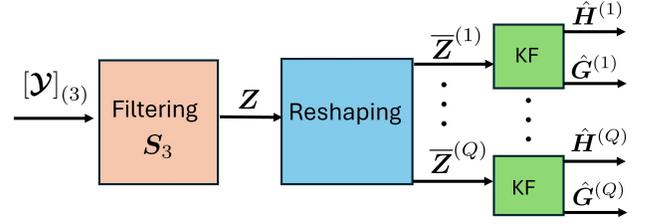
Fig. 3: Block-diagram of the BTKF algorithm.
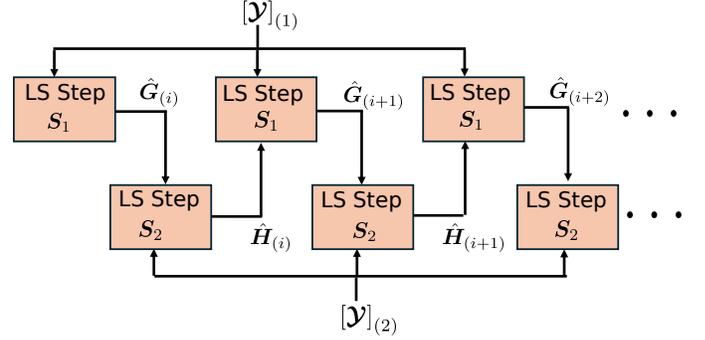
Fig. 4: Block-diagram of the BTALS algorithm.

B. Block Tucker alternating least squares (BTALS) algorithm for decoupled channel estimation

Starting from the noisy received pilot model in the tensor form shown in equation (39), let us consider the following problem

$$\min_{\boldsymbol{G}, \boldsymbol{H}} \left\| \mathcal{Y} - \mathcal{S} \times_1 \boldsymbol{G} \times_2 \boldsymbol{H} \right\|_{\mathrm{F}}^2. \quad (49)$$

It is clear that this is a nonlinear problem in the unknowns since the solution involves products of the coefficients of the channel matrices $\boldsymbol{H}$ and $\boldsymbol{G}$. However, it is known from tensor decomposition theory [30] that this problem can be efficiently solved through an alternating least squares (ALS) estimation algorithm [30], [31], a popular iterative method for fitting a tensor model thanks to its simplicity and monotonic convergence property. In our case, the algorithm yields decoupled estimates of the channel matrices by converting the bilinear problem in (49) into the following two simplest linear LS sub-problems. Specifically, to estimate the channels $\boldsymbol{H}$ and $\boldsymbol{G}$, we exploit the 1-mode and 2-mode unfoldings of the received pilots tensor $\mathcal{Y}$, defined in (32)-(33), respectively. The bilinear alternating least squares (BALS) method solves the following two LS problems in an iterative way

$$\hat{\boldsymbol{G}} = \operatorname*{argmin}_{\boldsymbol{G}} \left\| [\mathcal{Y}]_{(1)} - \boldsymbol{G} \boldsymbol{S}_1 (\overline{\boldsymbol{I}} |\otimes| \boldsymbol{H})^{\mathrm{T}} \right\|_{\mathrm{F}}^2,$$

$$\hat{\boldsymbol{H}} = \operatorname*{argmin}_{\boldsymbol{H}} \left\| [\mathcal{Y}]_{(2)} - \boldsymbol{H} \boldsymbol{S}_2 (\overline{\boldsymbol{I}} |\otimes| \boldsymbol{G})^{\mathrm{T}} \right\|_{\mathrm{F}}^2,$$

the solutions of which are respectively given by

$$\hat{\boldsymbol{G}} = [\mathcal{Y}]_{(1)} \left[\boldsymbol{S}_1 (\overline{\boldsymbol{I}} |\otimes| \boldsymbol{H})^{\mathrm{T}} \right]^{+}, \quad (50)$$

$$\hat{\boldsymbol{H}} = [\mathcal{Y}]_{(2)} \left[\boldsymbol{S}_2 (\overline{\boldsymbol{I}} |\otimes| \boldsymbol{G})^{\mathrm{T}} \right]^{+}. \quad (51)$$

The algorithm intertwines the LS estimates of $\boldsymbol{G}$ and $\boldsymbol{H}$. At each iteration, one matrix is updated based on a previously

obtained estimate of the other matrix. This procedure is repeated until the convergence. Figure 4 illustrates the BTALS iterations process. An error measure computed at the end of the i-th iteration is given by $\epsilon_{(i)} \doteq \| [\boldsymbol{\mathcal{Y}}]_{(3)} - \hat{\boldsymbol{Y}}_{(i)} \|_{\mathrm{F}}^2 / \| [\boldsymbol{\mathcal{Y}}]_{(3)} \|_{\mathrm{F}}^2$, where $\hat{\boldsymbol{Y}}_{(i)} \doteq [\boldsymbol{\mathcal{S}}]_{(3)} \left(\hat{\boldsymbol{H}}_{(i)} \otimes \hat{\boldsymbol{G}}_{(i)} \right)^{\mathrm{T}}$. The convergence is declared when this error does not significantly change between two successive iterations, which implies that $|\epsilon_{(i)} - \epsilon_{(i-1)}| \leq \eta$, where η is a convergence threshold. A summary of the steps of the proposed BALS algorithm is provided in Algorithm 2.

Such a bilinear ALS procedure may improve or maintain but cannot worsen the current fit, usually leading to a global minimum solution [15]. For simplicity, random initialization can be adopted. However, one may resort to enhanced initialization and acceleration procedures [15], [30]. Other initialization strategies that could be used to exploit the structure of the channel matrices are possible, although not advocated here. In our case, where the core tensor (represented by the BD-RIS structure) is known at the receiver, such enhancements are not required, and random initialization is enough to obtain satisfactory results.

The main steps of the proposed BTKF and BTALS algorithms are summarized in Algorithm 1 and Algorithm 2.

C. Discussion

From the above discussion, it is clear that the BTALS algorithm has less restrictive requirements on the training length K than the BTKF algorithm. This comes from the fact that the latter only exploits the 3-mode unfolding of the received pilot tensor (as is the case with the LS method [5]), and channel decoupling is carried out after the LS filtering step, which is the bottleneck restricting the training length. On the other hand, the BTALS algorithm exploits the 1-mode and 2-mode unfoldings of the received pilot tensor to yield direct decoupled estimates of the two involved channels. The associated LS steps involve more relaxed requirements for the training length K compared to BTKF.

It is worth highlighting the trade-offs between BTKF and BTALS regarding computational complexity and processing delay. To calculate the complexity, we assume a complexity $\mathcal{O}(mn)$ for computing a rank-one approximation of a matrix $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ [32], while the Moore-Penrose pseudo-inverse of $\boldsymbol{A}$ has a cost $\mathcal{O}(m^2 n)$. Hence, the BTKF algorithm requires $\mathcal{O}(Q\bar{N}^2 M_R M_T)$ for computing the Q rank-one approximations to find the individual channel estimates for each BD-RIS group, as shown in (44). In addition, since BTKF also includes a prior LS estimation step, its overall complexity corresponds to $\mathcal{O}(M_R T K \bar{N}^2 Q)$. In contrast, the BTALS algorithm involves two LS estimation steps at each iteration, where each step requires the computation of two matrix inverses, as shown in steps 4 and 5 of Algorithm 2. Considering the dimensions of these matrix inverses and summing up their individual complexities, we arrive at a total cost of $\mathcal{O}(I_{\max} \bar{N}^2 Q K (M_R + M_T))$, where $I_{\max}$ denotes the maximum number of iterations assumed.

It is worth mentioning that the BTFK method accomplishes channel estimation in closed form utilizing Q independent and parallel processing routines, which implies a shorter

---

Algorithm 1 Block Tucker Kronecker factorization algorithm (BTKF) for decoupled channel estimation

1: **Inputs**: Received signal tensor $\boldsymbol{\mathcal{Y}}$ and BD-RIS training tensors $\{\boldsymbol{\mathcal{S}}^{(1)}, \ldots, \boldsymbol{\mathcal{S}}^{(Q)}\}$.
2: Compute an estimate of $\boldsymbol{Z}$ by filtering

$$\left(\boldsymbol{S}_3^\dagger [\boldsymbol{\mathcal{Y}}]_{(3)} \right)^{\mathrm{T}} \approx \boldsymbol{H} |\otimes| \boldsymbol{G} \in \mathbb{C}^{M_R M_T \times \bar{N}^2 Q}.$$

3: **for** $q = 1 : Q$ **do**
4:     Partition $\boldsymbol{Z}$ into Q matrix blocks $\{\boldsymbol{Z}^{(1)}, \ldots, \boldsymbol{Z}^{(Q)}\}$:

$$\boldsymbol{Z}^{(q)} = \boldsymbol{Z}_{., [(q-1)\bar{N}^2, \ldots, q\bar{N}^2]} \in \mathbb{C}^{M_R M_T \times \bar{N}^2}.$$

5:     Reshape $\boldsymbol{Z}^{(q)}$ to obtain $\overline{\boldsymbol{Z}}^{(q)} \in \mathbb{C}^{M_R \bar{N} \times M_T \bar{N}}$.
6:     Define the SVD of $\overline{\boldsymbol{Z}}^{(q)} = \boldsymbol{U}^{(q)} \boldsymbol{\Sigma}^{(q)} \boldsymbol{V}^{(q)\mathrm{H}}$.
7:     Obtain an estimate of $\boldsymbol{G}^{(q)}$ and $\boldsymbol{H}^{(q)}$ as

$$\hat{\boldsymbol{G}}^{(q)} = \mathrm{unvec}\left(\boldsymbol{u}_1^{(q)} \right)_{M_R \times \bar{N}}, \ \hat{\boldsymbol{H}}^{(q)} = \mathrm{unvec}\left(\boldsymbol{v}_1^{*(q)} \right)_{M_T \times \bar{N}}.$$

8: **end for**
9: Return $\hat{\boldsymbol{G}} = [\hat{\boldsymbol{G}}^{(1)}, \ldots, \hat{\boldsymbol{G}}^{(Q)}]$, $\hat{\boldsymbol{H}} = [\hat{\boldsymbol{H}}^{(1)}, \ldots, \hat{\boldsymbol{H}}^{(Q)}]$.

Algorithm 2 Block Tucker alternating least squares (BTALS) algorithm for decoupled channel estimation

1: **Inputs**: Received signal tensor $\boldsymbol{\mathcal{Y}}$, BD-RIS training tensor $\boldsymbol{\mathcal{S}}$, number I of iterations, and convergence threshold η.
2: Set $i = 0$. Randomly initialize $\hat{\boldsymbol{H}}_{(i=0)}$.
3: **for** $i = 1 : I$ **do**
4:     Compute an LS estimate of $\boldsymbol{G}_{(i)}$ as

$$\hat{\boldsymbol{G}}_{(i)} = [\boldsymbol{\mathcal{Y}}]_{(1)} \left[\boldsymbol{S}_1 \left(\overline{\boldsymbol{I}} |\otimes| \boldsymbol{H}_{(i-1)} \right)^{\mathrm{T}} \right]^+.$$

5:     Compute an LS estimate of $\boldsymbol{H}_{(i)}$ as

$$\hat{\boldsymbol{H}}_{(i)} = [\boldsymbol{\mathcal{Y}}]_{(2)} \left[\boldsymbol{S}_2 \left(\overline{\boldsymbol{I}} |\otimes| \boldsymbol{G}_{(i)} \right)^{\mathrm{T}} \right]^+.$$

6:     Compute $\hat{\boldsymbol{Y}}_{(i)} = [\boldsymbol{\mathcal{S}}]_{(3)} \left(\hat{\boldsymbol{H}}_{(i)} \otimes \hat{\boldsymbol{G}}_{(i)} \right)^{\mathrm{T}}$
    and calculate the error $\epsilon_{(i)} = \left\| [\boldsymbol{\mathcal{Y}}]_{(3)} - \hat{\boldsymbol{Y}}_{(i)} \right\|_{\mathrm{F}}^2$.
7:     Check convergence and stop if $|\epsilon_{(i)} - \epsilon_{(i-1)}| \leq \eta$.
8: **end for**
9: Return $\hat{\boldsymbol{G}}_{(i)}$ and $\hat{\boldsymbol{H}}_{(i)}$.

---

processing delay than BTALS. The latter consists of a sequential process of alternating estimation steps, where the channel matrices associated with the Q BD-RIS groups are jointly estimated. Therefore, there is a longer processing delay than BTKF. From a practical perspective, optimized procedures for computing rank-one approximations and the possibility of parallel processing are attractive features of BTKF despite its higher training overhead than BTALS. Thus, there is clearly a trade-off between both algorithms involving complexity, training overhead, and processing delay.

Remark 3: The steps 4 and 5 of Algorithm 2, derived from the unfolding expressions (32)-(33), can be replaced by the updating steps $\hat{\boldsymbol{G}}_{(i)} = [\boldsymbol{\mathcal{Y}}]_{(1)} \left[[\boldsymbol{\mathcal{S}}]_{(1)} (\boldsymbol{I}_K \otimes \boldsymbol{H}_{(i-1)})^{\mathrm{T}} \right]^+$ and $\hat{\boldsymbol{H}}_{(i)} = [\boldsymbol{\mathcal{Y}}]_{(2)} \left[[\boldsymbol{\mathcal{S}}]_{(2)} (\boldsymbol{I}_K \otimes \boldsymbol{G}_{(i)})^{\mathrm{T}} \right]^+$ that result from the pair of expressions (21)-(22). As mentioned in *Remark 1*, both representations produce the same result. Numerical experiments have shown us that the second choice yields a

faster runtime in the MATLAB environment. This is because the traditional Kronecker product can be computed more efficiently than the block Kronecker product.

D. Uniqueness and identifiability issues

We discuss the identifiability issues and requirements in terms of parameter settings related to an essentially unique estimate of the channel matrices using the proposed BTKF and BTALS algorithms. In particular, capitalizing on the uniqueness conditions for the BTD [24] allows us to derive useful relations involving the system parameters, such as the number of transmit/receive antennas and training length for the proposed algorithms.

First, let us recall that the received pilot tensor in the noise-free case follows a Tucker-2 decomposition, as shown in (20). In the group-connected case, it can be expressed as a sum of Q blocks (see (28)). In the general case, Tucker models are not essentially unique due to rotational freedom since nonsingular transformations that compensate each other can be applied to the core tensor and the corresponding factor matrix in each mode without changing the tensor fit [25], [30]. Specifically, defining $\boldsymbol{T_G} \in \mathbb{C}^{N \times N}$ and $\boldsymbol{T_H} \in \mathbb{C}^{N \times N}$, any alternative solution $\hat{\boldsymbol{G}} = \boldsymbol{G} \boldsymbol{T_G}$ and $\hat{\boldsymbol{H}} = \boldsymbol{H} \boldsymbol{T_H}$ and $\hat{\boldsymbol{S}} = \boldsymbol{S} \times_1 \boldsymbol{T_G}^{-1} \times_2 \boldsymbol{T_H}^{-1}$ yields the same result since $\boldsymbol{\mathcal{Y}} = \hat{\boldsymbol{S}} \times_1 \hat{\boldsymbol{G}} \times_2 \hat{\boldsymbol{H}} = (\boldsymbol{S} \times_1 \boldsymbol{T_G}^{-1} \times_2 \boldsymbol{T_H}^{-1}) \times_1 (\boldsymbol{G} \boldsymbol{T_G}) \times_2 (\boldsymbol{H} \boldsymbol{T_H}) = \boldsymbol{S} \times_1 (\boldsymbol{G}(\boldsymbol{T_G} \boldsymbol{T_G}^{-1})) \times_2 (\boldsymbol{H}(\boldsymbol{T_H} \boldsymbol{T_H}^{-1})) = \boldsymbol{S} \times_1 \boldsymbol{G} \times_2 \boldsymbol{H}$. In the group-connected case, these nonsingular transformation matrices are confined within each block, and one can also arbitrarily permute the Q blocks without changing the result [24]. However, the core tensor is fixed in our case since we know the BD-RIS training tensor $\boldsymbol{S}$ at the receiver. In this case, these transformation matrices reduce to trivial scalar indeterminacies affecting the factor matrices that compensate each other, i.e., $\boldsymbol{T_G} = \alpha \boldsymbol{I}_N$ and $\boldsymbol{T_H} = \beta \boldsymbol{I}_N$., with $\alpha \beta = 1$.

Then, in the noise-free case, the estimated channel matrices $\hat{\boldsymbol{G}}$ and $\hat{\boldsymbol{H}}$ are related to the true ones by the identities

$$\hat{\boldsymbol{G}} = \alpha \boldsymbol{G}, \quad \hat{\boldsymbol{H}} = \beta \boldsymbol{H}, \quad \alpha \beta = 1.$$

Considering the 3-mode unfolding of the received pilot tensor given in (23) and introducing the transformation matrices, we have $[\boldsymbol{\mathcal{Y}}]_{(3)} = [\boldsymbol{\mathcal{S}}]_{(3)}(\hat{\boldsymbol{H}} \otimes \hat{\boldsymbol{G}})^{\mathrm{T}} = [\boldsymbol{\mathcal{S}}]_{(3)}((\boldsymbol{H} \boldsymbol{T_H}) \otimes (\boldsymbol{G} \boldsymbol{T_G}))^{\mathrm{T}} = [\boldsymbol{\mathcal{S}}]_{(3)}(\boldsymbol{T_H} \otimes \boldsymbol{T_G})^{\mathrm{T}}(\boldsymbol{H} \otimes \boldsymbol{G})^{\mathrm{T}}$. Note that the only choice for $\boldsymbol{T_G}$ and $\boldsymbol{T_H}$ satisfying the Kronecker-product equation $\boldsymbol{T_H} \otimes \boldsymbol{T_G} = \boldsymbol{I}_{N^2}$ are scaled identity matrices that compensate each other.

The reasoning for the group-connected case is similar but applies to each group. Hence, the channel matrices associated with the q-th group are related to the true ones by the identities

$$\hat{\boldsymbol{G}}^{(q)} = \alpha^{(q)} \boldsymbol{G}^{(q)}, \quad \hat{\boldsymbol{H}}^{(q)} = \beta^{(q)} \boldsymbol{H}^{(q)}, \quad \alpha^{(q)} \beta^{(q)} = 1$$

for $q = 1, \ldots, Q$. It is worth noting that these scaling ambiguities affecting the individual channels are irrelevant to the optimization of the BD-RIS scattering coefficients and transmit/receive beamforming vectors since they cancel each other when building the combined channel used to optimize the system parameters using state-of-the-art methods [5].

As discussed in Sections V-A and V-B (and summarized in Algorithms 1 and 2), the decoupled estimation of the channel matrices relies on the different unfoldings of the received pilot tensor. For the BTKF algorithm, the essential uniqueness of the channel estimates requires $\bar{\boldsymbol{S}}_3$ to be full row-rank so that the filtering step in (40) yields a unique solution. This implies that $K \geq \bar{N}^2 Q$ is required. The BTALS algorithm estimates the channel matrices by solving the LS problems in (50) and (51) (see also steps 4 and 5 of Algorithm 2). Defining $\boldsymbol{P}_1 \doteq \boldsymbol{S}_1 \left(\bar{\boldsymbol{I}} |\otimes| \boldsymbol{H} \right)^{\mathrm{T}} \in \mathbb{C}^{N \times K M_T}$ and $\boldsymbol{P}_2 \doteq \boldsymbol{S}_2 \left(\bar{\boldsymbol{I}} |\otimes| \boldsymbol{G} \right)^{\mathrm{T}} \in \mathbb{C}^{N \times K M_R}$ and from the LS problems in (50) and (51), we conclude that identifiability of $\hat{\boldsymbol{G}} = [\boldsymbol{\mathcal{Y}}]_{(1)} \boldsymbol{P}_1^{\dagger}$ and $\hat{\boldsymbol{H}} = [\boldsymbol{\mathcal{Y}}]_{(2)} \boldsymbol{P}_2^{\dagger}$ requires that $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ be right-invertible, which implies $K M_T \geq N$ and $K M_R \geq N$, respectively. In addition, it is required that $\boldsymbol{S}_1 \in \mathbb{C}^{\bar{N} Q \times \bar{N} K Q}$ and $\boldsymbol{S}_2 \in \mathbb{C}^{\bar{N} Q \times \bar{N} K Q}$ have full row-rank. Note, however, that these conditions are necessary but not sufficient for uniqueness. Nevertheless, these conditions are useful for the system designer to eliminate "bad" configurations that do not yield unique estimates of $\boldsymbol{G}$ and $\boldsymbol{H}$ separately. Otherwise stated, system setups that violate one of these conditions do not lead to unique channel estimates and can be discarded.

Sufficient condition: We can obtain a condition for guaranteed identifiability of the channel matrices by adopting a similar reasoning as that of [24] to study the uniqueness of the block Tucker2 decomposition (therein referred to as type-2 decomposition). The discussion is simpler in our context since the core tensor (represented by the BD-RIS training tensor) is known. Under the assumption that $\boldsymbol{G}$ and $\boldsymbol{H}$ are full column-rank, their essential uniqueness is guaranteed if $K \geq 3$ and the BD-RIS training tensor $\boldsymbol{\mathcal{S}}$ does not have proportional frontal slices, which means that $\boldsymbol{\mathcal{S}}_{..k} \neq \alpha \boldsymbol{\mathcal{S}}_{..k'}$, $\forall k \neq k'$, $k, k' \in \{1, \ldots, K\}$, where α is a scalar. This result implies that the design of the BD-RIS tensor is crucial to the uniqueness of the channel estimates for both BTKF and BTALS algorithms.

VI. BD-RIS TRAINING TENSOR DESIGN

As shown in the previous section, the BTKF and BTALS algorithms exploit the BD-RIS training tensor differently to estimate the involved channel matrices. While the BTKF relies on its-mode unfolding (37) for a two-step channel separation using Kronecker factorization, BTALS exploits its 1-mode and 2-mode unfoldings for a direct and iterative channel estimation method. It is clear that a proper design of the BD-RIS training tensor is crucial to ensuring unique estimates of the channel matrices (up to trivial scaling ambiguities) for both algorithms. We are interested in a more flexible (lower overhead) design, which is valid for $K << \bar{N}^2 Q$, in contrast to the design of [5] that requires $K = \bar{N}^2 Q$ for the LS estimation. It also allows us to go from conventional RIS to BD-RIS fully-connected architectures while covering a broader range of values of K. In the following, we discuss the design of the BD-RIS training tensor in two steps. We first formulate the baseline BD-RIS tensor design and then discuss the BD-RIS training structures used by the BTKF and BTALS algorithms, respectively.

Let us define the tensor $\boldsymbol{\mathcal{Z}} \in \mathbb{C}^{\bar{N} \times \bar{N} \times K_1}$ such that its k_1-th frontal slice corresponds to

$$\boldsymbol{\mathcal{Z}}_{..k_1} \doteq \mathrm{unvec}_{\bar{N} \times \bar{N}}(\boldsymbol{z}_{k_1}) \in \mathbb{C}^{\bar{N} \times \bar{N}}, \ k_1 = 1, \ldots, K_1. \quad (52)$$

Note that these frontal slices are column/row-permuted versions of the DFT matrix obtained from the applied circular shifts, as detailed earlier in this section. The basic structure of the BD-RIS tensor $\boldsymbol{\mathcal{S}}^{(q)} \in \mathbb{C}^{\bar{N} \times \bar{N} \times K}$ is given, as a function of its k-th frontal slice, as follows

$$\boldsymbol{\mathcal{S}}_{..k}^{(q)} = [\boldsymbol{\Theta}]_{k_2,q} \boldsymbol{\mathcal{Z}}_{..k_1} \in \mathbb{C}^{\bar{N} \times \bar{N}}, \ k \doteq (k_2 - 1)K_1 + k_1 \quad (53)$$

where $\boldsymbol{\Theta} \doteq [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_Q] \in \mathbb{C}^{K_2 \times Q}$ is a (possibly truncated) $K_2 \times Q$ Hadamard ($\boldsymbol{\Omega}_{\mathrm{HAD}}$) or a DFT ($\boldsymbol{\Omega}_{\mathrm{DFT}}$) matrix, where $K_2 = K/K_1$ must be an integer. Note that each slice $\boldsymbol{\mathcal{S}}_{..k}^{(q)} \in \mathbb{C}^{N \times N}$ is orthogonal, i.e., $\boldsymbol{\mathcal{S}}_{..k}^{(q)} \boldsymbol{\mathcal{S}}_{..k}^{(q)\mathrm{H}} = \boldsymbol{\mathcal{S}}_{..k}^{(q)\mathrm{H}} \boldsymbol{\mathcal{S}}_{..k}^{(q)} = \boldsymbol{I}_{\bar{N}}$, $k = 1, \ldots, K$ since $\boldsymbol{\mathcal{Z}}_{..k_1}$ defined in (52) is orthogonal $\forall k_1 \in \{1, \ldots, K_1\}$. Collecting the K slices in (53) according to (29)-(31) gives $[\boldsymbol{\mathcal{S}}^{(q)}]_{(1)} = \boldsymbol{\theta}_q^{\mathrm{T}} \otimes [\boldsymbol{\mathcal{Z}}]_{(1)} \in \mathbb{C}^{\bar{N} \times \bar{N}K}$, $[\boldsymbol{\mathcal{S}}^{(q)}]_{(2)} = \boldsymbol{\theta}_q^{\mathrm{T}} \otimes [\boldsymbol{\mathcal{Z}}]_{(2)} \in \mathbb{C}^{\bar{N} \times \bar{N}K}$, and $[\boldsymbol{\mathcal{S}}^{(q)}]_{(3)} = \boldsymbol{\theta}_q \otimes [\boldsymbol{\mathcal{Z}}]_{(3)} \in \mathbb{C}^{K \times \bar{N}^2}$.

Substituting into (34)-(35), and (37), we obtain

$$\boldsymbol{S}_1 = \mathrm{blkdiag}(\boldsymbol{\theta}_1^{\mathrm{T}} \otimes [\boldsymbol{\mathcal{Z}}]_{(1)}, \ldots, \boldsymbol{\theta}_Q^{\mathrm{T}} \otimes [\boldsymbol{\mathcal{Z}}]_{(1)}), \quad (54)$$

$$\boldsymbol{S}_2 = \mathrm{blkdiag}(\boldsymbol{\theta}_1^{\mathrm{T}} \otimes [\boldsymbol{\mathcal{Z}}]_{(2)}, \ldots, \boldsymbol{\theta}_Q^{\mathrm{T}} \otimes [\boldsymbol{\mathcal{Z}}]_{(2)}), \quad (55)$$

$$\boldsymbol{S}_3 = [\boldsymbol{\theta}_1 \otimes [\boldsymbol{\mathcal{Z}}]_{(3)}, \ldots, \boldsymbol{\theta}_Q \otimes [\boldsymbol{\mathcal{Z}}]_{(3)}] = \boldsymbol{\Theta} \otimes [\boldsymbol{\mathcal{Z}}]_{(3)}, \quad (56)$$

where $[\boldsymbol{\mathcal{Z}}]_{(1)} \in \mathbb{C}^{\bar{N} \times \bar{N}K_1}$, $[\boldsymbol{\mathcal{Z}}]_{(2)} \in \mathbb{C}^{\bar{N} \times \bar{N}K_1}$, and $[\boldsymbol{\mathcal{Z}}]_{(3)} \in \mathbb{C}^{K_1 \times \bar{N}^2}$ are the the unfoldings of the circulant tensor $\boldsymbol{\mathcal{Z}}$ constructed from its frontal slices similarly to (24)-(26). Since the frontal slices of the tensor $\boldsymbol{\mathcal{Z}}$ are symmetric due to (52), we have $[\boldsymbol{\mathcal{Z}}]_{(1)} = [\boldsymbol{\mathcal{Z}}]_{(2)}$, which implies $[\boldsymbol{\mathcal{S}}^{(q)}]_{(1)} = [\boldsymbol{\mathcal{S}}^{(q)}]_{(2)}$, $q = 1, \ldots, Q$, and, hence, $\boldsymbol{S}_1 = \boldsymbol{S}_2$ in (54)-(55).

A. BD-RIS design for the BTKF algorithm

The first step of the BTKF algorithm (see Figure 3) corresponds to filtering the received pilot tensor by the compact 3-mode unfolding $\boldsymbol{S}_3 = \boldsymbol{\Theta} \otimes [\boldsymbol{\mathcal{Z}}]_{(3)}$ of the BD-RIS tensor. This step is the same as that of the traditional LS estimation method and is subject to the same requirements.

If $K \geq \bar{N}^2 Q$, or, equivalently, if $K_1 = \bar{N}^2$ and $K_2 \geq Q$, then $\boldsymbol{S}_3$ is column-orthogonal, which implies $\boldsymbol{S}_3^{\mathrm{H}} \boldsymbol{S}_3 = (K/\bar{N}) \boldsymbol{I}_{\bar{N}^2 Q}$. This property can be checked as follows. We can first note that $K \geq \bar{N}^2 Q$ implies $K_1 = \bar{N}^2$ and $K_2 = K/\bar{N}^2$. In this case, the 3-mode unfolding of the circulant tensor $\boldsymbol{\mathcal{Z}} \in \mathbb{C}^{\bar{N} \times \bar{N} \times \bar{N}^2}$ is a square $\bar{N}^2 \times \bar{N}^2$ circulant matrix with orthogonal rows constructed from circular shifts of the DFT basis in vectorized form, which implies $[\boldsymbol{\mathcal{Z}}]_{(3)}^{\mathrm{H}} [\boldsymbol{\mathcal{Z}}]_{(3)} = \bar{N} \boldsymbol{I}_{\bar{N}^2}$. From (56), we have $[\boldsymbol{\mathcal{S}}^{(q)}]_{(3)}^{\mathrm{H}} [\boldsymbol{\mathcal{S}}^{(q)}]_{(3)} = (\boldsymbol{\Theta}^{\mathrm{H}} \boldsymbol{\Theta}) \otimes ([\boldsymbol{\mathcal{Z}}]_{(3)}^{\mathrm{H}} [\boldsymbol{\mathcal{Z}}]_{(3)}) = K/\bar{N}^2 \otimes \bar{N} \boldsymbol{I}_{\bar{N}^2} = (K/\bar{N}) \boldsymbol{I}_{\bar{N}^2}$. Hence, it follows that $\boldsymbol{S}_3^{\mathrm{H}} \boldsymbol{S}_3 = \boldsymbol{I}_{\bar{N}^2 Q}$.

Due to this column-orthogonality property, a simple matched filtering can replace the pseudo-inverse in step 3 of Algorithm 1 while offering optimized performance [5].

B. BD-RIS design for the BTALS algorithm

The BTALS algorithm can provide unique estimates of the individual channel matrices by capitalizing on the essential uniqueness properties of block Tucker decompositions [24], [33]. As discussed in section V-D, in our context, the uniqueness of our BD-RIS block Tucker model (28) assumes that the BD-RIS training tensor does not have proportional frontal slices, which means no slices in the set $\{\boldsymbol{\mathcal{S}}_{..1}^{(q)}, \ldots, \boldsymbol{\mathcal{S}}_{..K}^{(q)}\}$, $\forall q \in \{1, \ldots, Q\}$ are scalar multiples of each other. To cope with this assumption while preserving structural properties, we design the BD-RIS training tensor as

$$\boldsymbol{\mathcal{S}}_{..k}^{(q)} = \mathrm{D}_k(\boldsymbol{W}^{(q)}) \bar{\boldsymbol{\mathcal{S}}}_{..k}^{(q)} \mathrm{D}_k(\boldsymbol{W}^{(q)*}), \ k = 1, \ldots, K, \quad (57)$$

where $\bar{\boldsymbol{\mathcal{S}}}_{..k}^{(q)}$ follows the design in (53), and $\boldsymbol{W}^{(q)} \doteq [\boldsymbol{w}_1^{(q)}, \ldots, \boldsymbol{w}_K^{(q)}]^{\mathrm{T}} \in \mathbb{C}^{K \times \bar{N}}$ is a matrix of exponentials, whose k-th row is given by $\boldsymbol{w}_k^{(q)} \doteq [1 \ e^{j\psi_{k,1}^{(q)}}, \ldots, e^{j\psi_{k,\bar{N}-1}^{(q)}}]^{\mathrm{T}}$, with phases $\{\psi_{k,n}^{(q)}\}$ uniformly distributed between $[0, 2\pi)$, $k = 1, \ldots, K$, $q = 1, \ldots, Q$.

The random phase shifts generated by the transformations in (57) ensure that the K frontal slices of the BD-RIS training tensor are uncorrelated. This eliminates proportional slices that may result from the design in (53) depending on the values assumed for $\bar{N}$ and Q, ensuring the uniqueness of the channel estimates. Note also that the random left/right rotations applied to the basis tensor preserve the orthogonality of the frontal slices, i.e., $\boldsymbol{\mathcal{S}}_{..k}^{(q)} \boldsymbol{\mathcal{S}}_{..k}^{(q)\mathrm{H}} = \boldsymbol{\mathcal{S}}_{..k}^{(q)\mathrm{H}} \boldsymbol{\mathcal{S}}_{..k}^{(q)} = \boldsymbol{I}_{\bar{N}}$, $k = 1, \ldots, K$.

VII. NUMERICAL RESULTS

We evaluate the performance of the proposed BTKF and BTALS receivers under different system setups. We also discuss the tradeoffs involving the proposed channel estimation methods and their superior performance compared to the LS scheme [5]. We consider the NMSE of the combined channel $\boldsymbol{C} = \boldsymbol{H} |\otimes| \boldsymbol{G} \in \mathbb{C}^{M_R M_T \times \bar{N}^2 Q}$. The NMSE is given by

$$\mathrm{NMSE}(\hat{\boldsymbol{C}}) = \frac{\left\| \boldsymbol{C} - \hat{\boldsymbol{C}} \right\|_{\mathrm{F}}^2}{\|\boldsymbol{C}\|_{\mathrm{F}}^2}.$$

As in [5], we assume i.i.d. channels, and we compare the performance of the different methods for a variety of parameter setups that are shown in each figure.

In Figures 5 to 7, we evaluate the performance of the proposed receiver BTKF with the baseline LS of [5]. It is worth mentioning that the BTALS is omitted here due to the fact that it achieves the same performance as the BTKF in the considered scenarios. Since the BTALS is the only receiver that can operate at a lower training overhead, we will discuss its performance in a separate section (Figure 8 to 10). For all results, orthogonal pilot sequences are assumed, and their length is fixed to its minimum value $T_{\min} = M_T$.

In Figure 5, we compare the NMSE of the combined channel by fixing $K = 256$, which is the minimum value ($K_{\min} = \bar{N}^2 Q$) for the configuration with largest group size $\bar{N} = 4$. First, we observe that BTKF offers higher estimation accuracies than the baseline LS method, regardless of the group size. Indeed, the LS method estimates the combined channel as a classical MIMO channel estimation problem, which is *blind* to its inherent Kronecker factorization structure. Thus, the performance of the reference method is sensitive to the number of groups and degrades as the group size

Fig. 5: Comparison between BTKF and LS for different BD-RIS configurations.

Fig. 6: Comparison between BTKF and LS considering the minimum training overhead ($K_{\min}$) for each configuration.

Fig. 7: Comparison between BTKF and LS for different numbers of transmit and receive antennas.

Fig. 8: Performance of BTALS as a function of the number K of training blocks. The SNR is fixed to 20 dB.

increases as the group size, which explains the performance gap between the two methods as $\bar{N}$ increases. Note that, in the single-connected case ($\bar{N} = 1$), the numbers of estimated channel coefficients are the same for LS and BTKF. However, even in this case, the BTKF still offers a performance gain over LS. From this set of results, one can conclude that as the group size increases, the noise rejection gain offered by the rank-one approximation step of BTKF also increases. In contrast, the LS method has nearly the same performance regardless of the group size since it does not take advantage of the Kronecker structure of the combined channel.

In Figure 7, we fix the group size $\bar{N} = 4$ and the total number of groups $Q = 16$, implying $N = 64$ RIS elements. In this experiment, we compare the performance of BTKF and LS for different numbers of antennas M_T and M_R at the transmitter and receiver, respectively. Note that the pilot sequence length is adjusted in each configuration to its minimum value ($T = M_T$). We observe that the performance of the LS estimator is the same regardless of the number of transmit/receive antennas. Indeed, although the pilot length increases to ensure orthogonality, the number of channel coefficients of the combined channel also increases with M_T and M_R. This means that the LS estimator cannot benefit from the added spatial degrees of freedom to improve channel estimation performance. In contrast, the BTKF method efficiently benefits from more transmit and receive spatial degrees of freedom by capitalizing on the Kronecker structure of the combined channel. Such a gain comes from the noise rejection gains provided by the rank-one approximation problems in (44), whose dimensions increase with M_T and M_R, yielding more accurate estimates of $\boldsymbol{G}$ and $\boldsymbol{H}$ and, consequently, of the combined channel $\boldsymbol{C} = \hat{\boldsymbol{H}} \boxtimes \hat{\boldsymbol{G}}$.

In the next experiments (Figures 8 to 10), we turn our attention to the BTALS algorithm and study its performance for several system configurations. The focus is on very low training overhead setups, where $K << \bar{N}^2 Q$. Hence, these figures do not show the results of the LS and BTKF methods since they cannot operate in the considered challenging setups over the full range of values of K due to their more restrictive requirements on this training parameter.

Figure 8 depicts the NMSE performance of BTALS as a function of the number K of training blocks, by varying

($\bar{N}$) increases, while the BTKF method is insensitive to it. Additionally, we can see that the performance gains of BTKF over LS increase with $\bar{N}$. Such a gain comes from the channel separation property of BTKF that exploits the Kronecker structure of the combined channel.

In Figure 6, we assume the minimum training overhead for each considered BD-RIS configuration, i.e., $K_{\min} = \bar{N}^2 Q$ for each setup, that ensure a unique solution. It is interesting to note that the baseline LS scheme [5] has the same performance in all configurations. Hence, the total number of channel coefficients to be estimated is the same in all the training configurations considered in this figure. In contrast, the BTKF method estimates fewer channel coefficients than LS. The gap in terms of the number of channel coefficients to be estimated

Fig. 9: BTALS performance vs. the number Q of groups.
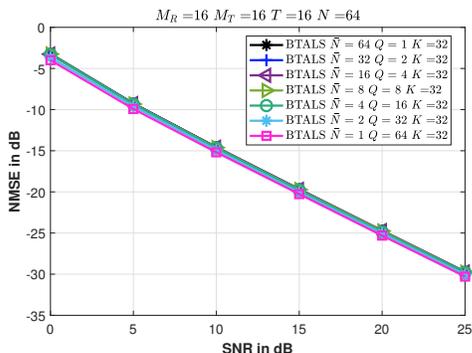
Fig. 10: BTALS performance for different configurations.

Fig. 11: Average number of iterations for BTALS convergence as a function of the number K of training blocks.

Fig. 12: Computational complexity of LS [5], BTKF, and BTALS for different group sizes $\bar{N}$.

from $K = 24$ to $K = 128$, for a fixed SNR equal to 20 dB. It can be observed that all configurations achieve very close performance. In particular, the performance gap between group-connected architectures and the fully connected architecture is negligible. The group-connected configurations have a slight performance gap ($\approx$ 0.5 dB) over the single-connected one ($\bar{N} = 1$, $Q = 64$). Note that for most of the system setups ($\bar{N} = \{4, 8, 16, 32, 64\}$), the range considered for the BD-RIS training length K is far below the minimum value required by the LS and BTKF methods, which is $K_{\min} = \bar{N}^2 Q$ in each case. As an example, for the configuration $(\bar{N}, Q) = (4, 16)$, LS and BTKF would require $K = 256$ blocks while for $(\bar{N}, Q) = (64, 1)$, they would require $K = 4096$ blocks. These results corroborate the remarkable savings of training resources provided by the BTALS algorithm, which can operate over a broader set of choices for K with significantly lower training overheads.

In Figure 9, we show the performance of BTALS as a function of the number Q of groups, going from the fully-connected case ($Q = 1$) to the single-connected case ($Q = 64$). We assume $N = \bar{N}Q = 64$ RIS elements, $K = 64$ blocks, and an SNR of 20 dB while considering configurations with different numbers of transmit and receive antennas. As expected, the performance increases as more transmit/receive antennas are used, showing the effectiveness of BTALS in converting spatial diversity gains at both link ends into more accurate channel estimates despite the increase in the number of estimated channel coefficients. These results corroborate the gains of tensor-based processing for BD-RIS channel estimation. We can also see that when the number of

antennas and the transmit/receive increases, the gap among the single-, group- and fully-connected architectures reduces.

Figure 10 shows the NMSE performance of BTALS as a function of the SNR, assuming $N = 64$ and different BD-RIS configurations (combinations of $\bar{N}$ and Q), including the single-connected architecture to the fully-connected one as extreme cases. In all configurations, the training length $K = 32$ (corresponding to $N/2$) is much smaller than the product $\bar{N}^2 Q$ that is the minimum overhead required by the LS and BTKF methods (both cannot operate in this case). We can see similar results in all the considered configurations. These results show that BTALS is an attractive solution in terms of training overhead compared to the competing schemes.

In Figure 11, we study the convergence performance of BTALS. First, we can see that lower training lengths imply more iterations to achieve convergence. Indeed, increased time diversity is available when more training blocks are used, which helps BTALS to speed up its iterative process. For very low training lengths ($K \leq 32$), the required number of iterations is more sensitive to the group size $\bar{N}$. However, for most of the considered values of K, the convergence behavior is similar regardless of the chosen group-connected configuration. It can also be noted that the single-connected case using optimal DFT design for the RIS training matrix needs fewer iterations than the group-connected cases, although the gap is not significant.

In Figure 12, we compare LS, BTKF, and BTALS in terms of computational complexity. We plot the complexity

order of each algorithm as a function of the group size $\bar{N}$ for a BD-RIS with $N = 64$ elements. Recall that $\bar{N} = 1$ corresponds to the single-connected case, while $\bar{N} = 64$ to the fully-connected case. As expected, the computational complexity increases with the group size in all cases. Moreover, since the convergence of BTALS is sensitive to the SNR (see Fig. 11), its overall complexity will increase at low SNRs. This is not the case with BTKF, whose complexity does not depend on the SNR due to its closed-form nature. Such a complexity gap is the price BTALS pays to operate at a very low overhead compared to the BTFK and the LS methods. In addition, BTKF and LS have comparable complexities in this scenario since the complexity is dominated by the matched filtering step (the same for both LS and BTKF).

VIII. CONCLUSION AND PERSPECTIVES

The channel estimation problem for BD-RIS can be addressed from a tensor decomposition perspective. Recasting the received pilot signals as a block Tucker tensor model yields individual estimates of the involved channels by exploiting the multilinear structure of the received pilot signals. Decoupling the individual channels for BD-RIS has some key benefits. First, it provides improved CSI estimation accuracy over traditional LS estimation by capitalizing on the inherent (block) Kronecker structure of the combined channel. The gains are more pronounced as more transmit/receive antennas are available. Second, the channel estimates are obtained with significantly lower training overheads than the LS method.

The proposed BTKF and BTALS algorithms have interesting tradeoffs. When training overhead is not critical, BTKF is preferable due to its lower complexity, closed-form solution, and small processing delay. On the other hand, when minimizing the training overhead is crucial (especially for strongly connected BD-RIS), BTALS is the most attractive choice due to its flexibility in operating with considerably fewer training blocks than the LS estimator with no performance degradation. Moreover, the proposed BD-RIS training yields decoupled estimates of the individual channels up to a scalar factor, which is irrelevant to system optimization.

REFERENCES

[1] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, Aug. 2019.

[2] M. Jian, G. C. Alexandropoulos, E. Basar, C. Huang, R. Liu, Y. Liu, and C. Yuen, "Reconfigurable intelligent surfaces for wireless communications: Overview of hardware designs, channel models, and estimation techniques," *Intelligent Conv. Networks*, vol. 3, 2022.

[3] H. Li, S. Shen, M. Nerini, and B. Clerckx, "Reconfigurable intelligent surfaces 2.0: Beyond diagonal phase shift matrices," *IEEE Communications Magazine*, vol. 62, no. 3, pp. 102–108, 2024.

[4] H. Li, S. Shen, and B. Clerckx, "Beyond diagonal reconfigurable intelligent surfaces: From transmitting and reflecting modes to single-, group-, and fully-connected architectures," *IEEE Transactions on Wireless Communications*, vol. 22, no. 4, pp. 2311–2324, 2023.

[5] H. Li, S. Shen, Y. Zhang, and B. Clerckx, "Channel estimation and beamforming for beyond diagonal reconfigurable intelligent surfaces," *IEEE Transactions on Signal Processing*, vol. 72, pp. 3318–3332, 2024.

[6] S. Shen, B. Clerckx, and R. Murch, "Modeling and architecture design of reconfigurable intelligent surfaces using scattering parameter network analysis," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1229–1243, 2021.

[7] M. Nerini, S. Shen, H. Li, and B. Clerckx, "Beyond diagonal reconfigurable intelligent surfaces utilizing graph theory: Modeling, architecture design, and optimization," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2024.

[8] A. S. de Sena, M. Rasti, N. H. Mahmood, and M. Latva-aho, "Beyond diagonal RIS for multi-band multi-cell MIMO networks: A practical frequency-dependent model and performance analysis," *arXiv preprint arXiv:2401.06475*, 2024.

[9] A. L. Swindlehurst, G. Zhou, R. Liu, C. Pan, and M. Li, "Channel estimation with reconfigurable intelligent surfaces–A general framework," *Proc. of the IEEE*, vol. 110, no. 9, pp. 1312–1338, 2022.

[10] B. Zheng, C. You, W. Mei, and R. Zhang, "A survey on channel estimation and practical passive beamforming design for intelligent reflecting surface aided wireless communications," *IEEE Comm. Surv. & Tutorials*, vol. 24, no. 2, pp. 1035–1071, 2022.

[11] C. Pan, G. Zhou, K. Zhi, S. Hong, T. Wu, Y. Pan, H. Ren, M. D. Renzo, A. Lee Swindlehurst, R. Zhang, and A. Y. Zhang, "An overview of signal processing techniques for RIS/IRS-aided wireless systems," *IEEE J. Sel. Topics Signal Process*, vol. 16, no. 5, pp. 883–917, 2022.

[12] C. Hu, L. Dai, S. Han, and X. Wang, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7736–7747, 11 2021.

[13] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, and M. Debbah, "Channel estimation for RIS-empowered multi-user MISO wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 4144–4157, 6 2021.

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

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

[16] A. L. F. de Almeida, G. Favier, J. P. da Costa, and J. C. M. Mota, "Overview of tensor decompositions with applications to communications," in *Signals and Images: Adv. Results in Speech, Estimation, Compression, Recognition, Filtering, and Processing*. CRC-Press, 1 2016, no. Chapter 12, pp. 325–356.

[17] S. Miron, Y. Zniyed, R. Boyer, A. L. F. de Almeida, G. Favier, D. Brie, and P. Comon, "Tensor methods for multisensor signal processing," *IET Signal Processing*, vol. 14, no. 10, pp. 693–709, 2020.

[18] H. Chen, F. Ahmad, S. Vorobyov, and F. Porikli, "Tensor decompositions in wireless communications and MIMO radar," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 438–453, 2021.

[19] G. T. de Araújo, A. L. F. de Almeida, and R. Boyer, "Channel estimation for intelligent reflecting surface assisted MIMO systems: A tensor modeling approach," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 3, pp. 789–802, Feb. 2021.

[20] G. T. de Araújo, A. L. F. de Almeida, R. Boyer, and G. Fodor, "Semi-blind joint channel and symbol estimation for IRS-assisted MIMO systems," *IEEE Trans. on Signal Process.*, vol. 71, pp. 1184–1199, 2023.

[21] K. Ardah, S. Gherekhloo, A. L. de Almeida, and M. Haardt, "TRICE: A channel estimation framework for RIS-aided millimeter-wave MIMO systems," *IEEE Signal Processing Letters*, vol. 28, pp. 513–517, 2021.

[22] B. Sokal, P. R. B. Gomes, A. L. F. de Almeida, B. Makki, and G. Fodor, "Reducing the control overhead of intelligent reconfigurable surfaces via a tensor-based low-rank factorization approach," *IEEE Transactions on Wireless Communications*, vol. 22, no. 10, pp. 6578–6593, 2023.

[23] K. B. A. Benício, A. L. F. de Almeida, B. Sokal, Fazal-E-Asim, B. Makki, and G. Fodor, "Tensor-based channel estimation and data-aided tracking in IRS-assisted MIMO systems," *IEEE Wireless Communications Letters*, vol. 13, no. 2, pp. 333–337, 2024.

[24] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms—part II: Definitions and uniqueness," *SIAM J. Mat. Anal. Appl.*, vol. 30, no. 3, pp. 1033–1066, 2008.

[25] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, Aug. 2009.

[26] G. Favier and A. L. F. de Almeida, "Overview of constrained PARAFAC models," *EURASIP Journal of Advances in Signal Processing*, vol. 2014, no. 142, pp. 1–25, Sep. 2014.

[27] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.

[28] L. De Lathauwer and D. Nion, "Decompositions of a higher-order tensor in block terms—part III: Alternating least squares algorithms," *SIAM J. Mat. Anal. Appl*, vol. 30, no. 3, pp. 1067–1083, 2008.

[29] C. Van Loan and N. Pitsianis, "Approximation with Kronecker Products," Cornell University, Tech. Rep., 1992.

[30] P. Comon, X. Luciani, and A. L. F. de Almeida, "Tensor decompositions, alternating least squares and other tales," *Journal of Chemometrics*, vol. 23, no. 7-8, pp. 393–405, 2009.

[31] A. Smilde, R. Bro, and P. Geladi, *Multi-way analysis: applications in the chemical sciences*. John Wiley & Sons, 2005.

[32] N. Kishore Kumar and J. Schneider, "Literature survey on low rank approximation of matrices," *Linear and Multilinear Algebra*, vol. 65, no. 11, pp. 2212–2244, 2017.

[33] A. Stegeman and A. L. F. de Almeida, "Uniqueness conditions for constrained three-way factor decompositions with linearly dependent loadings," *SIAM J. Mat. Anal. Appl.*, vol. 31, no. 3, pp. 1469–1490, 2010.