

Channel Estimation for Intelligent Reflecting Surface Assisted MIMO Systems: A Tensor Modeling Approach

Gilderlan T. de Araújo, André L. F. de Almeida, Rémy Boyer

Abstract—Intelligent reflecting surface (IRS) is an emerging technology for future wireless communications including 5G and especially 6G. It consists of a large 2D array of (semi-)passive scattering elements that control the electromagnetic properties of radio-frequency waves so that the reflected signals add coherently at the intended receiver or destructively to reduce co-channel interference. The promised gains of IRS-assisted communications depend on the accuracy of the channel state information. In this paper, we address the receiver design for an IRS-assisted multiple-input multiple-output (MIMO) communication system *via* a tensor modeling approach aiming at the channel estimation problem using supervised (pilot-assisted) methods. Considering a structured time-domain pattern of pilots and IRS phase shifts, we present two channel estimation methods that rely on a parallel factor (PARAFAC) tensor modeling of the received signals. The first one has a closed-form solution based on a Khatri-Rao factorization of the cascaded MIMO channel, by solving rank-1 matrix approximation problems, while the second one is an iterative alternating estimation scheme. The common feature of both methods is the decoupling of the estimates of the involved MIMO channel matrices (base station-IRS and IRS-user terminal), which provides performance enhancements in comparison to competing methods that are based on unstructured LS estimates of the cascaded channel. Design recommendations for both methods that guide the choice of the system parameters are discussed. Numerical results show the effectiveness of the proposed receivers, highlight the involved trade-offs, and corroborate their superior performance compared to competing LS-based solutions.

Index Terms—Intelligent reflecting surface, channel estimation, MIMO, tensor modeling, PARAFAC, Khatri-Rao factorization.

I. INTRODUCTION

In a typical wireless propagation environment, the transmitted signals suffer attenuation and scattering caused by absorption and reflection, diffraction, and refraction phenomena. In general, multipath propagation is known as one of the main limiting factors to the performance of a

wireless communication system [1]. Indeed, the randomness of the physical radio environment turns the wireless propagation uncontrollable.

Intelligent reflecting surface (IRS) (also referred to as reconfigurable intelligent surface or software-controlled metasurface) [2]–[7] is an emergent and promising technology for future (beyond 5G) wireless communications. It consists of a 2D array with a large number of passive or semi-passive elements that can control the electromagnetic properties of the radio-frequency waves so that the reflected signals add coherently at the intended receiver or destructively to reduce the co-channel interference. Each element can act independently and can be reconfigured in a software-defined manner by means of an external controller. The IRS does not require dedicated radio-frequency chains and is usually wirelessly powered by an external RF-based source, as opposed to amplify-and-forward or decode-and-forward relays, which require dedicated power sources [6]. In the literature, IRS is being considered in a number of application scenarios, such as to provide coverage to users located in a dead zone and to suppress co-channel interference when the user is in the edge of the cell [8], [9], and to improve the physical layer security [10], [11]. Besides, the IRS can be employed for simultaneous wireless information and power transfer in an IoT network [8]. Regarding wireless communication systems, recently, [12] established a connection between IRS technology and a millimeter wave (mmWave) hybrid MIMO systems. In this case, the authors consider a hybrid MIMO-OFDM assisted by IRS working in the mmWave band.

Recent works have discussed the potentials and challenges of IRS-assisted wireless communications (see, e.g., [2], [3] and references therein). Among the several open issues, we highlight the acquisition of channel state information. One challenge is related to the assumption that the IRS usually consists of passive elements, which means that the estimation of the cascaded channel should be performed at the receiver based on pilots sent by the transmitter *via* the IRS. At this point, the pattern of phase shifts used by the IRS during the training phase plays an important role. In addition, the large number of IRS elements imposes an extra challenge to the address the channel estimation problem. Two approaches have been proposed in the literature. The first one assumes a semi-passive structure, where the IRS has a few active elements connected to receive radio-frequency chains. In this case, the availability of some baseband processing at the IRS

Gilderlan T. de Araújo with Federal Institute of Ceará, Canindé, CE, e-mail: gilderlan.tavares@ifce.edu.br.

André Lima Férrer de Almeida is with Wireless Telecommunication Research Group (GTel), Department of Teleinformatics, Federal University of Ceará, Fortaleza, CE, e-mail: andre@gtel.ufc.br.

Rémy Boyer with University of Lille-1, CRISTAL Laboratory, France e-mail: remy.boyer@univ-lille.fr.

This work was supported by the Ericsson Research, Sweden, and Ericsson Innovation Center, Brazil, under UFC.48 Technical Cooperation Contract Ericsson/UFC. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES)-Finance Code 001, and CAPES/PRINT Proc. 88887.311965/2018-00. André L. F. de Almeida like to acknowledge CNPq for its financial support under the grant 306616/2016-5.

facilitates the CSI acquisition. An example of this approach is discussed in [13], where the involved channels are estimated by means of compressive sensing. The second approach, which is the one adopted in this paper, assumes a fully passive structure, where the IRS operates by reflecting the impinging waves according to some phase-shift pattern. This is a more challenging scenario, where the estimation of the cascaded (transmitter-IRS-receiver) should be done at the receiver based on pilots sent by the transmitter and reflected by the IRS.

A few works have addressed the channel estimation problem and provided different solutions to the passive IRS case. In [14], a minimum variance unbiased estimator is proposed, and an optimal design of the IRS phase shift matrix is found. The authors of [15] propose a two-stage algorithm by exploiting sparse representations of low-rank multipath channels. In [16], links between massive MIMO and IRS are discussed in the context of Terahertz communications, and a cooperative channel estimation *via* beam training is presented. In [17], IRS is proposed as a solution to mitigate the blockage problem in mmWave communications and a channel estimation approach is presented. The work [18] proposes an uplink channel estimation protocol for an IRS aided multi-user MIMO system applying compressing sensing (CS) methods. In [19], an IRS-aided MIMO system is considered, and channel estimation is carried out in a two-stage approach, and the IRS-assisted link is estimated by means of an approximate message-passing method. Considering an IRS-assisted internet of things scenario, [20] formulates a joint active detection and channel estimation based on sparse matrix factorization, matrix completion, and multiple measurement vector problems.

The authors of [21] propose a channel estimation framework where the BS-IRS, IRS-UT, and BS-UT channels are estimated in a two-timescale approach, while in [22] a practical transmission protocol is proposed to accomplish channel estimation and passive beamforming. In [23], channel estimation is carried out by resorting to an on-off strategy that sequentially activates the IRS elements one-by-one. The work [24] proposes a parallel factor model to solve the channel estimation problem in a multi-user MISO setting. In general, most of the existing works on IRS-assisted communications consider the multiple input single output (MISO) case, where the receiver station is equipped with a single antenna.

In the last decade, tensor modeling has been successfully applied in a variety of signal processing problems [25]–[30], in particular in the context of wireless communications, involving the design of semi-blind receivers for MIMO systems [31], [32], channel estimation methods for cooperative communications [33], [34], direction of arrival estimation and beamforming in array processing [35]–[37], and, more recently, compressed channel estimation in massive MIMO systems [38], [39]. As discussed in most of these works, tensor-based signal processing benefits from the powerful uniqueness properties of tensor decompositions while exploiting the multi-dimensional nature of the transmitted/received signals and communication channels. In this work, we establish an existing connection between IRS-assisted MIMO communications and tensor modeling. By assuming a structured time-domain pattern of pilots and

IRS phase shifts, we show that the received signal follows a parallel factor (PARAFAC) tensor model. By exploiting the PARAFAC signal structure in two different ways, we propose two simple and effective algorithms to estimate the cascaded MIMO channel via decoupling the transmitter-IRS and IRS-receiver MIMO channels, respectively. The first algorithm is a closed-form solution based on the Khatri-Rao factorization (KRF) of the combined BS-IRS and IRS-UT channels, while the second one consists of an iterative bilinear alternating least squares (BALS) algorithm. While the first algorithm is a closed-form algebraic and less complex solution, the second one can operate under less restrictive conditions on the system parameters.

The common feature of the two algorithms is that the estimation of the cascaded channel is achieved via decoupling the estimation of the two involved channel matrices, which provides a performance enhancement compared to the direct estimation of the cascaded channel via conventional least squares. By focusing on pilot-assisted channel estimation schemes, this work extends the results of our previous conference paper [40] by presenting a more comprehensive formulation of tensor-based IRS-assisted channel estimation methods, while bringing a discussion on the uniqueness conditions for the channel estimation problem considering the proposed receivers, from which useful design recommendations on the training parameters are derived. We discuss how to deal with a nonideal setup where the IRS phase shifts are not perfectly known at the receiver, and provide a solution to handle this problem. In addition, we also present generalizations of the proposed approach to multi-user scenarios. Numerical results corroborate the effectiveness of the proposed channel estimation methods and highlight the involved tradeoffs.

The contributions of this work are summarized as follows.

- Resorting to tensor modeling, we connect the channel estimation problem for IRS-assisted MIMO systems to that of fitting PARAFAC model to a third-order tensor;
- We derive two simple pilot-assisted channel estimation algorithms (namely, KRF and BALS) that exploit the algebraic structure of the PARAFAC model of the received signals in two different ways;
- We provide system design recommendations for the proposed KRF and BALS receivers that ensure the uniqueness of the channel estimation problem;
- We discuss how to handle perturbations/fluctuations on the IRS phase shifts by means of a joint channel and IRS matrix estimation at the receiver;
- Generalizations of the proposed tensor signal model to multi-user scenarios is provided, which include the multi-UT and the multi-BS cases;
- A detailed derivation of the analytical expressions of the CRB is provided.

Notation and properties: Matrices are represented with boldface capital letters ($\mathbf{A}, \mathbf{B}, \dots$), and vectors are denoted by boldface lowercase letters ($\mathbf{a}, \mathbf{b}, \dots$). Tensors are symbolized by calligraphic letters ($\mathcal{A}, \mathcal{B}, \dots$). Transpose and pseudo-inverse of a matrix \mathbf{A} are denoted as \mathbf{A}^T and \mathbf{A}^\dagger ,

respectively. The operator $\text{diag}(\mathbf{a})$ forms a diagonal matrix out of its vector argument, while $*$, \odot , \diamond , \oslash and \otimes denote the conjugate, outer product, Khatri Rao, Hadamard and Kronecker products, respectively. \mathbf{I}_N denotes the $N \times N$ identity matrix. The operator $\text{vec}(\cdot)$ vectorizes an $I \times J$ matrix argument, while $\text{unvec}_{I \times J}(\cdot)$ does the opposite operation. Moreover, $\text{vecd}(\cdot)$ forms a vector out of the diagonal of its matrix argument. The n -mode product between a tensor $\mathcal{Y} \in \mathbb{C}^{I \times J \times \dots \times K}$ and a matrix $\mathbf{A} \in \mathbb{C}^{I \times R}$ is denoted as $\mathcal{X} \times_n \mathbf{A}$, for $1 \leq n \leq N$. An identity N -way tensor of dimension $R \times R \times \dots \times R$ is denoted as $\mathcal{I}_{N,R}$. The operator $D_i(\mathbf{A})$ forms a diagonal matrix from the i -th row of its matrix argument \mathbf{A} . Moreover, \mathbf{A}_i denotes the i th row of the matrix \mathbf{A} . The operator $\lceil x \rceil$ rounds its fractional argument up to the nearest integer. In this paper, we make use of the following identities:

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A}\mathbf{C}) \otimes (\mathbf{B}\mathbf{D}). \quad (1)$$

$$(\mathbf{A} \otimes \mathbf{B})^H(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A}^H \mathbf{C}) \otimes (\mathbf{B}^H \mathbf{D}). \quad (2)$$

$$\text{vec}(\mathbf{A}\mathbf{B}\mathbf{C}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B}). \quad (3)$$

$$\text{diag}(\mathbf{a})\mathbf{b} = \text{diag}(\mathbf{b})\mathbf{a}. \quad (4)$$

If \mathbf{B} is a diagonal matrix, we have:

$$\text{vec}(\mathbf{A}\mathbf{B}\mathbf{C}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vecd}(\mathbf{B}). \quad (5)$$

II. SYSTEM MODEL

We consider a MIMO communication system assisted by an IRS. Both the transmitter and the receiver are equipped with multiple antennas. Although the terminology adopted in this paper assumes a downlink communication, where the transmitter is the base station (BS) and the receiver is the user terminal (UT), our signal model also applies to the uplink case by just inverting the roles of the transmitter and the receiver. The BS and UT are equipped with arrays of M and L antennas, respectively. The IRS is composed of N elements, or unit cells, capable of individually adjusting their reflection coefficients (i.e., phase shifts). The system model is illustrated in Figure 1. In a time-slotted transmission, we assume that the IRS adjusts its phase-shifts as a function of the time $t = 1, \dots, T$. We also assume a block-fading channel, which means that the BS-IRS and IRS-UT channels are constant during T time slots. The received signal is given as [15]

$$\mathbf{y}[t] = \mathbf{G}(\mathbf{s}[t] \odot \mathbf{H}\mathbf{x}[t]) + \mathbf{b}[t], \quad 1 \leq t \leq T, \quad (6)$$

or, alternatively,

$$\mathbf{y}[t] = \mathbf{G}\text{diag}(\mathbf{s}[t])\mathbf{H}\mathbf{x}[t] + \mathbf{b}[t], \quad (7)$$

where $\mathbf{x}[t] \in \mathbb{C}^{M \times 1}$ is the vector containing the transmitted pilot signals at time t , $\mathbf{s}[t] = [s_{1,t}e^{j\phi_1}, \dots, s_{N,t}e^{j\phi_N}]^T \in \mathbb{C}^{N \times 1}$ is the vector that models the phase shifts and activation pattern of the IRS, where $\phi_n \in (0, 2\pi]$, and $s_{n,t} \in \{0, 1\}$ controls the on-off state of the corresponding element at time t . The matrices $\mathbf{H} \in \mathbb{C}^{N \times M}$ and $\mathbf{G} \in \mathbb{C}^{L \times N}$ denote the BS-IRS and IRS-UT MIMO channels, respectively, while $\mathbf{b}[t] \in \mathbb{C}^{L \times 1}$ is the additive white Gaussian noise (AWGN) vector.

The channel training time T_s is divided into K blocks, where each block has T time slots so that $T_s = KT$. Let us

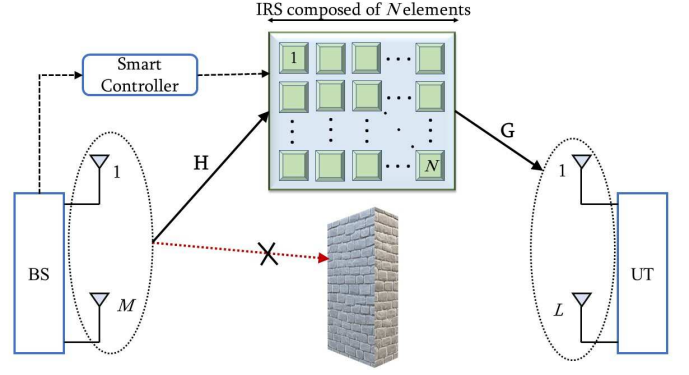


Fig. 1. IRS-assisted MIMO system

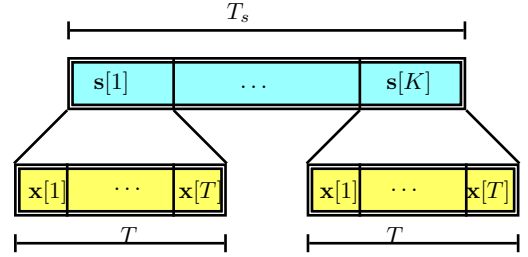


Fig. 2. Structured pilot pattern in the time domain

define $\mathbf{y}[k, t] \doteq \mathbf{y}[(k-1)T + t]$ as the received signal at the t -th time slot of the k -th block, $t = 1, \dots, T$, $k = 1, \dots, K$. Likewise, denote $\mathbf{x}[k, t]$ and $\mathbf{s}[k, t]$ as the pilot signal and phase shift vectors associated with the t -th time slot of the k -th block. We propose the following structured time-domain protocol: i) the IRS phase shift vector is constant during the T time slots of the k -th block and varies from block to block; ii) the pilot signals $\{\mathbf{x}[1], \dots, \mathbf{x}[T]\}$ are repeated over the K blocks. Mathematically, this means that

$$\mathbf{s}[k, t] = \mathbf{s}[k], \quad \text{for } 1 \leq t \leq T, \quad (8)$$

$$\mathbf{x}[k, t] = \mathbf{x}[t], \quad \text{for } 1 \leq k \leq K. \quad (9)$$

An illustration of this time-domain protocol is shown in Figure 2. Under these assumptions, the received signal model (7) can be written as

$$\mathbf{y}[k, t] = \mathbf{G}\text{diag}(\mathbf{s}[k])\mathbf{H}\mathbf{x}[t] + \mathbf{b}[k, t]. \quad (10)$$

Collecting the received signals during T time slots for the k -th block in $\mathbf{Y}[k] = [\mathbf{y}[k, 1] \dots \mathbf{y}[k, T]] \in \mathbb{C}^{L \times T}$ leads to

$$\mathbf{Y}[k] = \mathbf{G}\text{diag}(\mathbf{s}[k])\mathbf{H}\mathbf{X}^T + \mathbf{B}[k], \quad (11)$$

where $\mathbf{X} \doteq [\mathbf{x}[1], \dots, \mathbf{x}[T]]^T \in \mathbb{C}^{T \times M}$, and $\mathbf{B} \doteq [\mathbf{b}[1], \dots, \mathbf{b}[T]]^T \in \mathbb{C}^{L \times T}$.

A. Least squares channel estimation

A baseline method consists of estimating a combined version of the communication channels \mathbf{G} and \mathbf{H} using least squares (LS) approach. To derive the LS estimator, we apply property (5) to (11) to obtain

$$\begin{aligned} \mathbf{y}[k] &= (\mathbf{X}\mathbf{H}^T \otimes \mathbf{G}) \mathbf{s}_k \\ &= (\mathbf{X} \otimes \mathbf{I}_L) (\mathbf{H}^T \otimes \mathbf{G}) \mathbf{s}[k] + \mathbf{b}[k], \end{aligned} \quad (12)$$

where $\mathbf{y}[k] \doteq \text{vec}(\mathbf{Y}[k]) \in \mathbb{C}^{LT}$, $\mathbf{b}[k] \doteq \text{vec}(\mathbf{B}[k]) \in \mathbb{C}^{LT}$, and we have used property (1). Defining $\tilde{\mathbf{Y}} \doteq [\mathbf{y}[1] \dots \mathbf{y}[K]] \in \mathbb{C}^{LT \times K}$, and $\tilde{\mathbf{X}} \doteq (\mathbf{X} \otimes \mathbf{I}_L) \in \mathbb{C}^{TL \times ML}$, we have

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}} (\mathbf{H}^T \diamond \mathbf{G}) \mathbf{S}^T + \mathbf{B}, \quad (13)$$

where $\mathbf{S} \doteq [\mathbf{s}[1], \dots, \mathbf{s}[K]]^T \in \mathbb{C}^{K \times N}$, $\mathbf{B} \in \mathbb{C}^{LT \times K}$ is the noise matrix constructed in the same way as \mathbf{Y} . Finally, defining $\mathbf{y} \doteq \text{vec}(\tilde{\mathbf{Y}}) \in \mathbb{C}^{LT \times K}$, and applying property (3) to (13), we get

$$\mathbf{y} = (\mathbf{S} \otimes \tilde{\mathbf{X}}) \text{vec}(\mathbf{H}^T \diamond \mathbf{G}) + \mathbf{b}, \quad (14)$$

or, compactly,

$$\mathbf{y} = \mathbf{U}\boldsymbol{\theta} + \mathbf{b}, \quad (15)$$

where $\mathbf{U} \doteq \mathbf{S} \otimes \tilde{\mathbf{X}} \in \mathbb{C}^{KTL \times NML}$ and $\boldsymbol{\theta} \doteq \text{vec}(\mathbf{H}^T \diamond \mathbf{G}) \in \mathbb{C}^{MLN}$ is the *composite channel parameter* combining the BS-IRS and IRS-UT channels. The LS estimate of this composite channel minimizes the problem

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{U}\boldsymbol{\theta}\|^2, \quad (16)$$

the solution of which is known to be $\boldsymbol{\theta} = \mathbf{U}^\dagger \mathbf{y}$. The computation of this solution can indeed be simplified to $\boldsymbol{\theta} = (\mathbf{S}^\dagger \otimes \tilde{\mathbf{X}}^\dagger) \mathbf{y}$, due to the Kronecker structure of \mathbf{U} .

It should be noted that the conventional LS problem ignores the Khatri-Rao structure of the composite channel that is present in the linearized parameter vector $\boldsymbol{\theta}$. Indeed, the signal model (11), or equivalently, (13) has a tensor structure, and can be recast as a PARAFAC tensor model. As we show in Section III, adopting this tensor modeling allows us to enhance the channel estimation accuracy (compared to conventional LS methods). This is achieved by decoupling the estimates of \mathbf{H} and \mathbf{G} , rather than estimating $\boldsymbol{\theta} = \text{vec}(\mathbf{H}^T \diamond \mathbf{G})$ as a whole. Moreover, useful system design recommendations can be derived from the proposed modeling approach.

B. Tensor signal modeling

In order to simplify the exposition of the signal model, we remove the noise term from the following developments. The noise term will be taken into account later. We can rewrite the signal part of equation (11) as

$$\bar{\mathbf{Y}}[k] = \mathbf{G} \mathbf{D}_k(\mathbf{S}) \mathbf{Z}^T, \quad \mathbf{Z} \doteq \mathbf{X} \mathbf{H}^T \in \mathbb{C}^{T \times N}, \quad (17)$$

where $\mathbf{D}_k(\mathbf{S}) \doteq \text{diag}(\mathbf{s}[k])$ denotes a diagonal matrix holding the k -th row of the IRS phase shift matrix \mathbf{S} on its main diagonal. The matrix $\bar{\mathbf{Y}}[k]$ can be viewed as the k -th frontal matrix slice of a three-way tensor $\bar{\mathcal{Y}} \in \mathbb{C}^{L \times T \times K}$ that follows a PARAFAC decomposition, also known as canonical polyadic decomposition (CPD) [26], [41]–[44]. Each (ℓ, t, k) -th entry of the noiseless received signal tensor $\bar{\mathcal{Y}}$ can be written as:

$$\bar{y}_{\ell, t, k} = \sum_{n=1}^N g_{\ell, n} z_{t, n} s_{k, n}, \quad (18)$$

where $g_{\ell, n} \doteq [\mathbf{G}]_{\ell, n}$, $z_{t, n} \doteq [\mathbf{Z}]_{t, n}$, and $s_{k, n} \doteq [\mathbf{S}]_{k, n}$. A shorthand notation for the PARAFAC decomposition (18) is denoted as $\bar{\mathcal{Y}} = [[\mathbf{G}, \mathbf{Z}, \mathbf{S}]]$. Using the n -mode product

notation, the PARAFAC decomposition of the noiseless received signal tensor $\bar{\mathcal{Y}}$ can be represented as

$$\bar{\mathcal{Y}} = \mathcal{I}_{3, N} \times_1 \mathbf{G} \times_2 \mathbf{Z} \times_3 \mathbf{S}. \quad (19)$$

Exploiting the trilinearity of the PARAFAC decomposition, we can “unfold” received signal tensor $\bar{\mathcal{Y}}$ into the following three matrix forms [41], [42]:

$$\bar{\mathbf{Y}}_1 = \mathbf{G}(\mathbf{S} \diamond \mathbf{Z})^T \in \mathbb{C}^{L \times TK}, \quad (20)$$

$$\bar{\mathbf{Y}}_2 = \mathbf{Z}(\mathbf{S} \diamond \mathbf{G})^T \in \mathbb{C}^{T \times LK}, \quad (21)$$

$$\bar{\mathbf{Y}}_3 = \mathbf{S}(\mathbf{Z} \diamond \mathbf{G})^T \in \mathbb{C}^{K \times LT}, \quad (22)$$

where $\bar{\mathbf{Y}}_1 \doteq [\bar{\mathbf{Y}}[1], \dots, \bar{\mathbf{Y}}[K]]$, $\bar{\mathbf{Y}}_2 \doteq [\bar{\mathbf{Y}}^T[1], \dots, \bar{\mathbf{Y}}^T[K]]$, and $\bar{\mathbf{Y}}_3 \doteq [\text{vec}(\bar{\mathbf{Y}}[1]), \dots, \text{vec}(\bar{\mathbf{Y}}[K])]^T$. In the following, the algebraic structure of the PARAFAC model (18) is exploited to formulate two channel estimation methods. The PARAFAC model is powerful due to its essential factor identification uniqueness property, which has its roots on the concept of Kruskal rank (k-rank). Details can be found in [45], [46].

III. CHANNEL ESTIMATION METHODS

Our goal is to estimate the channel matrices \mathbf{H} (BS-IRS) and \mathbf{G} (IRS-UT) from the received signal tensor given in (18). Let us define $\mathcal{Y} \doteq \bar{\mathcal{Y}} + \mathcal{B}$ as the noise-corrupted received signal tensor, where $\mathcal{B} \in \mathbb{C}^{L \times T \times K}$ is the additive noise tensor. Likewise, $\mathbf{Y}_i \doteq \bar{\mathbf{Y}}_i + \mathbf{B}_i$, $i = 1, 2, 3$, are the noisy versions of the 1-mode, 2-mode and 3-mode matrix unfoldings (20)–(22) of the received signal tensor, and $\mathbf{B}_{i=1,2,3}$ the corresponding matrix unfoldings of the noise tensor.

The pilot signal matrix \mathbf{X} and the IRS phase shifts matrix \mathbf{S} can be designed as semi-unitary matrices satisfying $\mathbf{X}^H \mathbf{X} = T \mathbf{I}_M$ and $\mathbf{S}^H \mathbf{S} = K \mathbf{I}_N$, respectively. A good choice is to design both \mathbf{X} and \mathbf{S} as truncated discrete Fourier transform (DFT) matrices. The optimal design of the IRS matrix \mathbf{S} is discussed in [14] for the multiple-input single-output (MISO) case (i.e., for single-antenna users).

A. Khatri-Rao Factorization based channel estimation

First, note that we can rewrite the noise-corrupted matrix unfolding (22) as:

$$\begin{aligned} \mathbf{Y}_3 &= \mathbf{S}(\mathbf{Z} \diamond \mathbf{G})^T + \mathbf{B}_3 \\ &= \mathbf{S}(\mathbf{H}^T \diamond \mathbf{G})^T (\mathbf{X} \otimes \mathbf{I}_L)^T + \mathbf{B}_3, \end{aligned} \quad (23)$$

where we have applied the property $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \diamond \mathbf{D}) = (\mathbf{A}\mathbf{C}) \diamond (\mathbf{B}\mathbf{D})$ to the term $(\mathbf{Z} \diamond \mathbf{G}) = (\mathbf{X}\mathbf{H}^T \diamond \mathbf{G})$. A bilinear time-domain filtering is applied at the receiver by exploiting the knowledge of the IRS matrix and the pilot signal matrix, as follows

$$\boldsymbol{\Omega} \doteq (\mathbf{X}^\dagger \otimes \mathbf{I}_L) \mathbf{Y}_3^T (\mathbf{S}^T)^\dagger = \mathbf{H}^T \diamond \mathbf{G} + \tilde{\mathbf{B}}_3, \quad (24)$$

where $\tilde{\mathbf{B}}_3 = (\mathbf{X}^\dagger \otimes \mathbf{I}_L) \mathbf{B}_3^T (\mathbf{S}^T)^\dagger$ is the filtered noise term. Note that $\boldsymbol{\Omega} \in \mathbb{C}^{ML \times N}$ is a noisy version of the (Khatri-Rao structured) *virtual MIMO channel* that models the IRS-assisted MIMO transmission. Due to the semi-unitary structure of \mathbf{S} and \mathbf{X} , the correlation properties of the additive noise are not affected by the bilinear filtering step.

Algorithm 1: Khatri-Rao factorization (KRF)**Procedure****output:** $\hat{\mathbf{H}}$ and $\hat{\mathbf{G}}$ **begin***Bilinear filtering of \mathbf{Y}_3 :* $\mathbf{\Omega}^T \leftarrow \mathbf{S}^H \mathbf{Y}_3 (\mathbf{X}^* \otimes \mathbf{I}_L)$ **for** $n = 1, \dots, N$ **do** $\tilde{\mathbf{\Omega}}_n \leftarrow \text{unvec}_{L \times M}(\omega_n)$ $(\mathbf{u}_1, \sigma_1, \mathbf{v}_1) \leftarrow \text{t-SVD}(\tilde{\mathbf{\Omega}}_n)$ $\hat{\mathbf{h}}_n \leftarrow \sqrt{\sigma_1} \mathbf{v}_1^*$ $\hat{\mathbf{g}}_n \leftarrow \sqrt{\sigma_1} \mathbf{u}_1$ **end***Reconstruct $\hat{\mathbf{H}}$ and $\hat{\mathbf{G}}$:* $\hat{\mathbf{H}} \leftarrow [\hat{\mathbf{h}}_1, \dots, \hat{\mathbf{h}}_N]^T$ $\hat{\mathbf{G}} \leftarrow [\hat{\mathbf{g}}_1, \dots, \hat{\mathbf{g}}_N]$ **end****Algorithm 2: Bilinear alternating least squares (BALS)****Procedure****input :** $i = 0$; Initialize $\hat{\mathbf{H}}_{(i=0)}$ **output:** $\hat{\mathbf{H}}, \hat{\mathbf{G}}$ **begin** $i = i + 1$;**while** $\|e(i) - e(i-1)\| \geq \delta$ **do**1: Find a least squares estimate of \mathbf{G} :

$$\hat{\mathbf{G}}_{(i)} = \mathbf{Y}_1 \left[\left(\mathbf{S} \diamond \mathbf{X} \hat{\mathbf{H}}_{(i-1)}^T \right)^T \right]^\dagger$$

2: Find a least squares estimate of \mathbf{H} :

$$\hat{\mathbf{H}}_{(i)}^T = \mathbf{X}^\dagger \mathbf{Y}_2 \left[\left(\mathbf{S} \diamond \mathbf{G}_{(i)} \right)^T \right]^\dagger$$

3: Repeat steps 1 to 2 until convergence.

end**end**

From (24), we propose to estimate \mathbf{H} and \mathbf{G} by solving the following Khatri-Rao least squares approximation problem

$$\min_{\mathbf{H}, \mathbf{G}} \|\mathbf{\Omega} - \mathbf{H}^T \diamond \mathbf{G}\|_F^2. \quad (25)$$

An efficient solution to this problem is given by the Khatri-Rao factorization (KRF) algorithm [47], [48]. Note that the problem (25) can be interpreted as finding estimates of \mathbf{H} and \mathbf{G} that minimize a set of rank-1 matrix approximations, i.e.,

$$(\hat{\mathbf{H}}, \hat{\mathbf{G}}) = \arg \min_{\{\mathbf{h}_n\}, \{\mathbf{g}_n\}} \sum_{n=1}^N \|\tilde{\mathbf{\Omega}}_n - \mathbf{g}_n \mathbf{h}_n^T\|_F^2, \quad (26)$$

where $\tilde{\mathbf{\Omega}}_n \doteq \text{unvec}_{L \times M}(\omega_n) \in \mathbf{C}^{L \times M}$, while $\mathbf{g}_n \in \mathbf{C}^{L \times 1}$ and $\mathbf{h}_n^T \in \mathbf{C}^{1 \times M}$ are the n -th column of \mathbf{G} and n -th row of \mathbf{H} , respectively. The estimates of \mathbf{g}_n and \mathbf{h}_n in (26) can be obtained from the dominant left and right singular vectors of $\tilde{\mathbf{\Omega}}_n$, respectively, for $1 \leq n \leq N$. Hence, our channel estimation problem translates into solving N rank-1 matrix approximation subproblems, for which several efficient solutions exist in the literature [49]. A summary of the KRF algorithm is given in Algorithm 1, where t-SVD denotes a truncated SVD (t-SVD denotes also tensor SVD in the tensor literature) that returns the dominant singular vector and its associated singular value. Once $\hat{\mathbf{H}}$ and $\hat{\mathbf{G}}$ are found from problem (26), we can build the composite channel.

B. BALS channel estimation

From the noisy versions of the matrix unfoldings (20) and (21), we can derive an iterative solution based on a bilinear alternating least squares (BALS) algorithm. This algorithm is a simplified version of the well-known trilinear ALS algorithm for estimating the factor matrices of a PARAFAC model [50]. In our case, since \mathbf{S} is known at the receiver, it consists of

estimating the matrices \mathbf{G} and \mathbf{H} in an alternating way by iteratively optimizing the following two cost functions:

$$\hat{\mathbf{G}} = \arg \min_{\mathbf{G}} \|\mathbf{Y}_1 - \mathbf{G}(\mathbf{S} \diamond \mathbf{X} \mathbf{H}^T)^T\|_F^2, \quad (27)$$

$$\hat{\mathbf{H}} = \arg \min_{\mathbf{H}} \|\mathbf{Y}_2 - \mathbf{X} \mathbf{H}^T (\mathbf{S} \diamond \mathbf{G})^T\|_F^2, \quad (28)$$

the solutions of which are respectively given by

$$\hat{\mathbf{G}} = \mathbf{Y}_1 \left[\left(\mathbf{S} \diamond \mathbf{X} \mathbf{H}^T \right)^T \right]^\dagger, \quad (29)$$

$$\hat{\mathbf{H}}^T = \mathbf{X}^\dagger \mathbf{Y}_2 \left[\left(\mathbf{S} \diamond \mathbf{G} \right)^T \right]^\dagger. \quad (30)$$

The convergence is declared when $\|e(i) - e(i-1)\| \leq \delta$, where $e(i) = \|\mathcal{Y} - \hat{\mathcal{Y}}_{(i)}\|_F^2$ denotes the reconstruction error computed at the i -th iteration, δ is a threshold parameter, and $\hat{\mathcal{Y}}_{(i)} = [\hat{\mathbf{G}}_{(i)}, \mathbf{X} \hat{\mathbf{H}}_{(i)}^T, \hat{\mathbf{S}}]$ is the reconstructed PARAFAC model (c.f. (11), (18)) obtained from the estimated channel matrices $\hat{\mathbf{G}}_{(i)}$ and $\hat{\mathbf{H}}_{(i)}$ at the end of the i -th iteration. In this work, we adopt $\epsilon = 10^{-5}$. Despite the iterative nature of the BALS algorithm, only a few iterations are necessary for convergence (usually less than 10 iterations) thanks to the knowledge of the IRS matrix \mathbf{S} that remains fixed during the iterations.

If \mathbf{X} and \mathbf{S} are column-orthogonal (which requires $K \geq N$ and $T \geq M$), the right pseudo-inverses in (29) and (30) can be replaced by matrix products. This leads to a lower complexity implementation of the BALS algorithm with simplified estimation steps, as shown in Appendix B.

C. Computational complexity

The computational complexity is in general dominated by the (truncated) SVD steps involved in Algorithm 1 (KRF) to compute rank-1 matrix approximations, as well as in Algorithm 2 (BALS) to calculate the LS estimates of the channel matrices. First, recall that computing the SVD of a matrix $P \times Q$ has a complexity order of $\mathcal{O}(PQ \min(P, Q))$,

while computing the inner product of two matrices of dimensions $P \times F$ and $F \times Q$ has complexity $\mathcal{O}(PQF)$. The complexity of KRF is that of computing N rank-1 approximation routines from the matrix $\tilde{\mathbf{\Omega}}_n$, $n = 1, \dots, N$, which can be efficiently implemented by means of the well-known power method [49]. From these results, we find that the KRF algorithm has a complexity of order $\mathcal{O}(MLN)$ owing to the N rank-1 matrix approximations. As for the iterative BALS receiver, as discussed in Section III-B, the computational cost associated with steps 1 and 2 is that of computing two right pseudo-inverses per iteration, which is equivalent to $\mathcal{O}(NKT[2N + L])$ and $\mathcal{O}(MT[2M + LK] + N^2[LK + M] + MNLK)$, respectively. Note, however, that this computational cost is greatly reduced when \mathbf{S} and \mathbf{X} have orthogonal columns (which require $K \geq N$ and $T \geq M$, respectively). According to the steps derived in Appendix B, the cost per iteration is that of computing the matrix products in steps 1 and 2 of Algorithm 3, which corresponds to $\mathcal{O}(LN[TK + N])$ and $\mathcal{O}(M[LKT + LKN + N^2])$, respectively.

D. Dealing with IRS perturbations and blockages

In outdoor scenarios, due to the exposure of the IRS to weather and atmospheric conditions, its elements may be subject to unknown blockages, as well as time-dependent fluctuations on their phase and amplitude responses [51]. Such unknown perturbations have a random nature, and may be caused by water droplets, snowflakes, dry and damp sand particles, among others. In this case, the IRS matrix \mathbf{S} deviates from its desired structure, and the assumption of a perfect knowledge of all the phase shifts at the receiver may not hold. Otherwise stated, the receiver cannot benefit from the full knowledge of the IRS phase shifts to estimate the cascade channel, i.e., it should be able to operate in a *semi-blind* way. Adopting our tensor modeling approach, it is possible to deal with this issue by resorting to a trilinear alternating least squares (TALS) algorithm that jointly estimates \mathbf{G} , \mathbf{H} , and \mathbf{S} by fully exploiting the trilinear structure of the received signal tensor in (18)-(19). The TALS algorithm is an extension of the BALS one by adding in Algorithm 2 a third estimation step

$$\hat{\mathbf{S}}_{(i)} = \mathbf{Y}_3 \left[(\mathbf{X} \hat{\mathbf{H}}_{(i)}^T \diamond \hat{\mathbf{G}}_{(i)})^T \right]^\dagger$$

that includes the update/refinement the IRS matrix within the loop. The TALS arises as a good alternative to deal with IRS phase shift perturbations. Since the channel matrices are now estimated in a blind way, i.e., without the knowledge of the IRS matrix \mathbf{S} , more iterations are required to achieve convergence. Moreover, the complexity is also increased due to the additional LS estimation step at each iteration. TALS is a well-known algorithm for fitting a PARAFAC model [42], [43], [50]. In such a blind approach, we can resort to the Kruskal's uniqueness conditions for the PARAFAC model [45] to obtain useful system design recommendations. A simplified condition can be obtained when the channel matrices have full

rank.¹ In this case, $\min(L, N) + \min(M, N) + \min(K, N) \geq 2N + 2$ guarantees the uniqueness of \mathbf{G} , \mathbf{H} and \mathbf{S} (see [26], [46] for a deeper uniqueness discussion in the general case). It is known that TALS may suffer from slow convergence due to its sensitivity to the initialization. However, several enhancements may be used to improve its performance (see [43] and references therein).

IV. DESIGN RECOMMENDATIONS

The KRF method (Algorithm 1) has a bilinear filtering step as shown in (24) requiring that the IRS phase shift matrix \mathbf{S} and the pilot symbol matrix \mathbf{X} have full column-rank, which implies the following conditions

$$K \geq N \quad \text{and} \quad T \geq M. \quad (31)$$

As mentioned earlier, a good choice is to design \mathbf{X} and \mathbf{S} are semi-unitary (or column-orthogonal) matrices, for two reasons. First, because the semi-unitary design replaces matrix inversions in (24) by simple matrix products, simplifying the receiver processing. Second, because the correlation properties of the filtered noise term in (24) are preserved.

The BALS method (Algorithm 2) requires that the two Khatri-Rao product terms $\mathbf{M}_1 = \mathbf{S} \diamond \mathbf{X} \mathbf{H}^T \in \mathbb{C}^{KT \times N}$ and $\mathbf{M}_2 = \mathbf{S} \diamond \mathbf{G} \in \mathbb{C}^{KL \times N}$ have full column-rank, so that (29) and (30) (resp. steps 1 and 2 of Algorithm 2) admit unique solutions. This means that the conditions $KT \geq N$ and $KL \geq N$ must be satisfied. Combining these two inequalities implies $\min(KT, KL) \geq N$, or, equivalently, $K \min(T, L) \geq N$. In addition, the condition $T \geq M$ in (28) is required, since \mathbf{X} must have full column-rank to be left-invertible. Hence, the following conditions are necessary

$$K \min(T, L) \geq N. \quad \text{and} \quad T \geq M \quad (32)$$

Comparing the conditions (31) and (32), we can note that BALS has a less restrictive requirement on the minimum number K of time blocks for the channel training compared to KRF method. Note that, in the special case $L = 1$ (MISO or SISO system), the inequalities (31) and (32) are identical, i.e., BALS and KRF are subject to the same training requirements. The advantage of BALS over KRF appears in the MIMO case, since BALS can operate under $K < N$, while KRF requires $K \geq N^2$. On the other hand, KRF usually has a lower computational complexity than BALS, as will be shown later in the discussion of our numerical results.

Note that condition (32) is necessary but does not guarantee the uniqueness of the BALS estimates. Sufficient conditions can be derived by studying the rank properties of $\mathbf{M}_1 = \mathbf{S} \diamond \mathbf{X} \mathbf{H}^T \in \mathbb{C}^{KT \times N}$ and $\mathbf{M}_2 = \mathbf{S} \diamond \mathbf{G} \in \mathbb{C}^{KL \times N}$. To this end, let us invoke the following result.

¹The condition $\min(L, N) + \min(M, N) + \min(K, N) \geq 2N + 2$ usually implies more restrictive choices on the system parameters L , M , and K , compared to the conditions discussed in Section V, which are valid when considering the perfect knowledge of the IRS matrix.

²Note that if $K = 1$, KRF reduces to the conventional LS estimator. However, in this case we cannot resolve/decouple the estimation of the two channel matrices, and the performance gain obtained with such a decoupling via solving problem (25) is lost.

Lemma 1 (Rank of the Khatri-Rao product [46], [52]): For $\mathbf{A} \in \mathbb{C}^{I \times N}$ and $\mathbf{B} \in \mathbb{C}^{J \times N}$, if $\text{rank}(\mathbf{A}) \geq 1$ and $\text{rank}(\mathbf{B}) \geq 1$, then $\text{rank}(\mathbf{A} \diamond \mathbf{B}) \geq \min(\text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) - 1, N)$. A concise proof of this lemma can be found in [46], [53]. This result means that the Khatri-Rao product of \mathbf{A} and \mathbf{B} will have full column-rank if $\text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) \geq N + 1$.

The application of Lemma 1 to the Khatri-Rao structured matrices $\mathbf{M}_1 = \mathbf{S} \diamond \mathbf{X}\mathbf{H}^T$ and $\mathbf{M}_2 = \mathbf{S} \diamond \mathbf{G}$ leads to the following conditions that guarantee the uniqueness of the channel estimates via solving the problems (27) and (28)

$$\text{rank}(\mathbf{S}) + \text{rank}(\mathbf{X}\mathbf{H}^T) \geq N + 1 \quad (33)$$

$$\text{rank}(\mathbf{S}) + \text{rank}(\mathbf{G}) \geq N + 1 \quad (34)$$

Let us consider that our channel training design parameters, namely, the IRS phase shift matrix \mathbf{S} and the pilot symbols matrix \mathbf{X} are designed to have full rank. The above conditions yield useful corollaries for the system design, when BALS is used. In the following, we discuss these corollaries.

A. The BS-IRS and IRS-UT channel matrices have full rank

Assuming that both channel matrices \mathbf{H} and \mathbf{G} have full rank (e.g. i.i.d. Rayleigh fading), conditions (33)-(34) can be rewritten as

$$\min(K, N) + \min(M, N) \geq N + 1 \quad (35)$$

$$\min(K, N) + \min(L, N) \geq N + 1. \quad (36)$$

We may distinguish two cases, as follows.

- $N \geq T \geq M$ and $N \geq L$: In this scenario, the BS and the UT have small to moderate antenna array sizes, whose number of antennas are smaller than the number of IRS elements. In this case, conditions (33)-(34) reduce to

$$M + \min(K, N) \geq N + 1 \quad (37)$$

$$L + \min(K, N) \geq N + 1 \quad (38)$$

- $T \geq M \geq N$: In this scenario, the BS is assumed to be equipped with a large antenna array, which has as many antennas as the number of IRS elements (massive MIMO setup). Since condition (33) is always satisfied regardless of the value of K , the uniqueness of the channel estimates only depends on (34), which translates to

$$\min(K, N) + \min(L, N) \geq N + 1 \quad (39)$$

The conditions (37) and (38) establish a tradeoff between the time dimension (number K of IRS training blocks) and the two spatial dimensions (number M and L of transmit and receive antennas, respectively) from a channel recovery viewpoint. For instance, if $K < N$, these conditions imply $M + K \geq N + 1$ and $L + K \geq N + 1$, which is equivalent to $\min(M + K, L + K) \geq N + 1$. Hence, reducing the number of transmit (or receive) antennas should be compensated by an increase on the number of time blocks in order to ensure the uniqueness of the channel estimates via the BALS algorithm.

B. The BS-IRS and IRS-UT channel matrices are rank-deficient

In millimeter-wave MIMO systems, the large number of transmit/receive antennas combined with scattering-poor propagation may result in low rank channel matrices \mathbf{H} and \mathbf{G} . Let us assume that the signal propagating between the BS and IRS experiences R_1 clusters, while the signal propagating between the IRS and the UT experiences R_2 clusters. Moreover, assume that each cluster contributes with one ray that has a complex amplitude and an associated angle of arrival/departure. We can represent the BS-IRS and IRS-UT channels as follows [54]

$$\mathbf{H} = \mathbf{A}_{\text{IRS}} \text{diag}(\boldsymbol{\alpha}) \mathbf{A}_{\text{BS}}^H, \quad (40)$$

$$\mathbf{G} = \mathbf{B}_{\text{UT}} \text{diag}(\boldsymbol{\beta}) \mathbf{B}_{\text{IRS}}^H, \quad (41)$$

where $\mathbf{A}_{\text{BS}} \in \mathbb{C}^{M \times R_1}$, $\mathbf{A}_{\text{IRS}} \in \mathbb{C}^{N \times R_1}$, $\mathbf{B}_{\text{UT}} \in \mathbb{C}^{L \times R_2}$ and $\mathbf{B}_{\text{IRS}} \in \mathbb{C}^{N \times R_2}$ are the array response matrices, and the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ hold the complex amplitude coefficients of the BS-IRS and IRS-UT channels, respectively. More specifically, we have $\text{rank}(\mathbf{H}) = R_1$ and $\text{rank}(\mathbf{G}) = R_2$, where it is assumed that $R_1 \leq \min(M, N)$ and $R_2 \leq \min(L, N)$ (rank-deficient case).

First, note that the conditions (31) required by the KRF algorithm to solve the decoupled channel estimation problem are not affected by the rank deficiency of the channel matrices. However, this is not the case for BALS, since the uniqueness of the LS estimates of \mathbf{G} and \mathbf{H} depend on the rank of these matrices, as shown in conditions (33) and (34). Considering BALS, we can draw useful corollaries as follows.

- $T \geq M$: Conditions (33) and (34) reduce to

$$\min(K, N) + R_1 \geq N + 1 \quad (42)$$

$$\min(K, N) + R_2 \geq N + 1 \quad (43)$$

It is worth discussing the following cases. If $K \geq N$, we conclude that these conditions are always satisfied, irrespective of the ranks of the channel matrices. If $K < N$, these conditions reduce to $K + R_1 \geq N + 1$ and $K + R_2 \geq N + 1$, yielding a useful design recommendation the number K of blocks that guarantee the uniqueness of the channel estimates in the rank-deficient case.

- $K \geq N$: In this case, conditions (33) and (34) are always satisfied, irrespective of the rank of \mathbf{G} and \mathbf{H} .

Discussion: It is worth noting that the proposed channel estimation methods still work for $K = 1$. However, in this setup only the cascaded channel $\mathbf{C} = \mathbf{G} \text{diag}(\mathbf{s}) \mathbf{H}$ can be estimated. The performance enhancements that come from the decoupling of the estimates of \mathbf{H} and \mathbf{G} (via KRF or BALS) cannot be obtained. Otherwise stated, leveraging extra training time diversity by increasing the number K of IRS phase shift patterns allows us to extract additional gains in comparison to the traditional LS estimator, as will be shown in our numerical results. However, such gains come at the expense of an increase on the training resources. Therefore, here we clearly see a trade-off between training overhead and performance.

In addition, it is clear from conditions (33)-(34), or equivalently, from conditions (35) and (35), that BALS can operate under more flexible choices for K than KRF, since the latter always requires $K \geq N$. Note that, for $M \geq N$ and $L \geq N$, these conditions are satisfied even for small values of K . In practice, this means that BALS may operate with a much lower training overhead than KRF and the traditional LS methods. However, our experience shows that for a large number of IRS elements, working with small values of K results in a slower convergence speed of BALS due to the limited time diversity. Therefore, a trade-off between training overhead and complexity arises when BALS is considered.

Note also that, in terms of the required training time resources, BALS becomes equivalent to KRF in the single transmit antenna case ($M = 1$) and/or in the single receive antenna case ($L = 1$). Otherwise stated, for MISO and/or SIMO IRS-assisted systems, BALS and KRF have exactly the same requirement $K \geq N$. Thus, we can say that BALS is advantageous over KRF in terms of training overhead when considering the MIMO case. Likewise, performance gains of KRF and BALS over the baseline LS method also arise in the MIMO setup, where spatial degrees of freedom at the transmitter and the receiver are efficiently exploited to obtain more accurate channel estimates. Finally, one can note that the channel estimates $\hat{\mathbf{G}}$ and $\hat{\mathbf{H}}$ are affected by scaling factors³ satisfying $\hat{\mathbf{H}} = \Delta_H \mathbf{H}$ and $\hat{\mathbf{G}} = \mathbf{G} \Delta_G$, where $\Delta_H \Delta_G = \mathbf{I}_N$. These scaling ambiguities are irrelevant in our context since they compensate each other when building the estimate of the cascade BS-IRS-UT channel.

V. GENERALIZATIONS TO MULTI-USER SCENARIOS

Although we have focused on the single BS and single UT scenario, the proposed approach as well as the derived results can be easily generalized and adapted to IRS-assisted multiple-access/multi-user MIMO systems. Let us take the uplink case as an example. The downlink case follows exactly the same model by just inverting the roles of BS and UT. We can distinguish two scenarios, which are discussed as follows.

A. Multiple users communicate with a single BS via the IRS

Let us consider U UTs communicating with a single BS via the IRS. The direct link between the UTs and the BS is assumed to be too weak or unavailable. Assuming for simplicity that all the users have the same number L of transmit antennas, we can adapt equation (17) such that the contribution of the u -th user to the received signal at the BS is given as

$$\bar{\mathbf{Y}}_u[k] = \mathbf{H}^T \mathbf{D}_k(\mathbf{S}) \mathbf{G}_u^T \mathbf{X}_u^T \quad (44)$$

where $\mathbf{X}_u \in \mathbb{C}^{T \times L}$ and $\mathbf{G}_u \in \mathbb{C}^{L \times N}$ are respectively the u -th user pilot matrix and uplink channel matrix. Note that the IRS-BS channel \mathbf{H} is common to all the users. The total

signal received from the U users at the k -th time block, in the noiseless case, can then be expressed as

$$\begin{aligned} \bar{\mathbf{Y}}[k] &= \mathbf{H}^T \mathbf{D}_k(\mathbf{S}) (\mathbf{X}_1 \mathbf{G}_1)^T + \cdots + \mathbf{H}^T \mathbf{D}_k(\mathbf{S}) (\mathbf{X}_U \mathbf{G}_U)^T \\ &= \mathbf{H}^T \mathbf{D}_k(\mathbf{S}) \left[\sum_{u=1}^U (\mathbf{X}_u \mathbf{G}_u)^T \right]. \end{aligned} \quad (45)$$

Defining $\bar{\mathbf{X}} \doteq [\mathbf{X}_1, \dots, \mathbf{X}_U] \in \mathbb{C}^{T \times UL}$, and $\bar{\mathbf{G}} \doteq [\mathbf{G}_1^T, \dots, \mathbf{G}_U^T]^T \in \mathbb{C}^{UL \times N}$, equation (45) translates to⁴

$$\mathbf{Y}[k] = \mathbf{H}^T \mathbf{D}_k(\mathbf{S}) \bar{\mathbf{Z}}^T \quad \bar{\mathbf{Z}} \doteq \bar{\mathbf{X}} \bar{\mathbf{G}} \in \mathbb{C}^{T \times N}. \quad (46)$$

Comparing (46) with (17), we can see that the multi-user signal model has the same tensor structure as the single-user one, the essential difference being on the definition of the factor matrix $\bar{\mathbf{Z}}$ which is now given by inner product of block matrices composed of U blocks (each having L columns as in the single-user scenario). Otherwise stated (46) corresponds to a PARAFAC model of $\bar{\mathbf{Y}} \in \mathbb{C}^{M \times T \times K}$ with factor matrices $(\mathbf{H}^T, \bar{\mathbf{X}} \bar{\mathbf{G}}, \mathbf{S})$, and unfoldings 1-mode and 2-mode given as $\bar{\mathbf{Y}}_1 = \mathbf{H}^T (\mathbf{S} \diamond \bar{\mathbf{X}} \bar{\mathbf{G}})^T$ and $\bar{\mathbf{Y}}_2 = \bar{\mathbf{X}} \bar{\mathbf{G}} (\mathbf{S} \diamond \mathbf{H}^T)^T$, respectively. Since the structure of the tensor model is not changed, both KRF and BALS algorithms can be directly applied to the multi-user model (46) under more restrictive choices for T , due to the fact the full rankness of \mathbf{X} now requires $T \geq UL$. In this case, assuming that the channel matrices have full rank, the application of Lemma 1 leads to

$$\min(K, N) + \min(UL, N) \geq N + 1 \quad (47)$$

$$\min(K, N) + \min(M, N) \geq N + 1. \quad (48)$$

These conditions are analogous to (35)-(36), by exchanging the roles of M and L , while adding the factor U .

B. Multiple users communicate with multiple BSs via the IRS

We consider that P BSs receive the signals transmitted by the U users via the IRS. Without loss of generality, the BSs are assumed to be equipped with the same number M of antennas. The model (46) is only slightly modified by adding a dependency of the received signal on the index p of the receiving BS, i.e.,

$$\bar{\mathbf{Y}}_p[k] = \mathbf{H}_p^T \mathbf{D}_k(\mathbf{S}) \bar{\mathbf{Z}}^T, \quad \bar{\mathbf{Z}} \doteq \bar{\mathbf{X}} \bar{\mathbf{G}}. \quad (49)$$

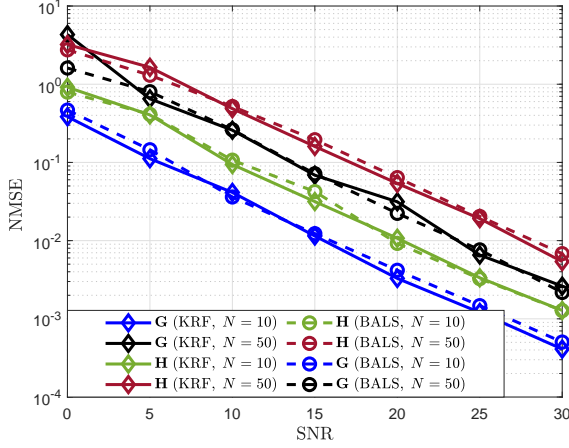
In particular, in a cooperative setting where the P BSs communicate (e.g. via a common backhauling structure), we can derive an equivalent augmented signal model as follows

$$\bar{\bar{\mathbf{Y}}}[k] = \begin{bmatrix} \bar{\mathbf{Y}}_1[k] \\ \vdots \\ \bar{\mathbf{Y}}_P[k] \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1^T \\ \vdots \\ \mathbf{H}_P^T \end{bmatrix} \mathbf{D}_k(\mathbf{S}) \bar{\mathbf{Z}}^T = \bar{\mathbf{H}}^T \mathbf{D}_k(\mathbf{S}) \bar{\mathbf{Z}}^T, \quad (50)$$

where $\bar{\mathbf{H}} \doteq [\mathbf{H}_1, \dots, \mathbf{H}_P]^T \in \mathbb{C}^{PM \times N}$ is the composite channel combining the IRS links to the P BSs. The received signal (50) corresponds to a PARAFAC model of $\bar{\bar{\mathbf{Y}}} \in \mathbb{C}^{PM \times T \times K}$ with factor matrices $(\bar{\mathbf{H}}, \bar{\mathbf{X}} \bar{\mathbf{G}}, \mathbf{S})$. Note that,

³The permutation ambiguity inherent to blind estimation is not present due to the knowledge of the IRS phase shift matrix \mathbf{S} at the receiver.

⁴Note that the positions of \mathbf{H} and \mathbf{G} in are swapped in (46) compared to (17) in addition to transposition, since channel reciprocity is assumed for the UT-IRS and IRS-BS links.

Fig. 3. NMSE of the estimated channels $\hat{\mathbf{H}}$ and $\hat{\mathbf{G}}$.

differently from the single-user single-BS model (17) and the multi-user single-BS model (46), in the multi-user multi-BS model (50) the dimensionality of the first mode of the received signal tensor has been increased by a factor P due to the assumption of cooperating BSs. In this scenario, condition (47) remains the same, while condition (48) slightly changes to $\min(K, N) + \min(PM, N) \geq N + 1$.

As a final remark, in terms of receiver processing, it is clear that both KRF and BALS have an increased computational complexity in the discussed multi-user scenarios, due to the increased dimensionality of the channel matrices $\bar{\mathbf{G}}$ and $\bar{\mathbf{H}}$.

VI. NUMERICAL RESULTS

In this section, several numerical results are presented to evaluate the performance of the proposed channel estimation methods, while comparing to competing approaches. We also evaluate the CRB as a reference for comparisons. The channel estimation accuracy is evaluated in terms of the normalized mean square error (NMSE) given by

$$\text{NMSE}(\hat{\mathbf{H}}) = \frac{1}{R} \sum_{r=1}^R \frac{\|\mathbf{H}^{(r)} - \hat{\mathbf{H}}^{(r)}\|_F^2}{\|\mathbf{H}^{(r)}\|_F^2}, \quad (51)$$

where $\hat{\mathbf{H}}^{(r)}$ is the BS-IRS channel estimated at the r -th run, and R denotes the number of Monte Carlo runs. The same definition applies to the estimated IRS-UT channel. The SNR (in dB) is defined as

$$\text{SNR} = 10 \log_{10}(\|\bar{\mathbf{Y}}\|_F^2 / \|\mathbf{B}\|_F^2), \quad (52)$$

where $\bar{\mathbf{Y}}$ is the noiseless received signal tensor generated according to (18), and \mathbf{B} is the additive noise tensor.

We assume that the entries of the BS-IRS and IRS-UT channel matrices \mathbf{H} and \mathbf{G} are independent and identically distributed zero-mean circularly-symmetric complex Gaussian random variables. The Figures 3, 5 and 6, represent an average from $R = 5000$ run Monte Carlo runs for the fixed system parameters $\{T = 4, L = 2, K = 50, M = 3\}$ and $N \in \{50, 100\}$.

Figure 3 depicts the NMSE vs. SNR curves for the KRF and the BALS algorithms. We can see that both algorithms

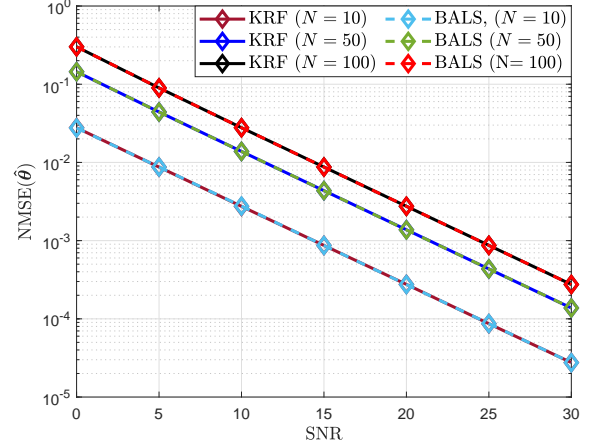
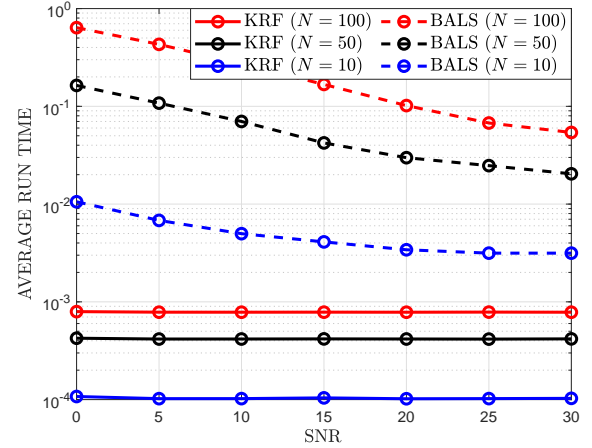
Fig. 4. NMSE for the equivalent channel $\hat{\boldsymbol{\theta}}$.

Fig. 5. Average runtime of KRF and BALS algorithms.

provide satisfactory performances. The performance degrades as the number of IRS elements is increased, which is an expected result since the number of channel coefficients in \mathbf{G} and \mathbf{H} to be estimated also increases with N . In Figure 4, the NMSE of the composite parameter vector $\boldsymbol{\theta}$ is shown. The parameters used get this figure was $K = 100$, $M = 3$, $T = 4$, $L = 20$, $N \in \{10, 50, 100\}$ and 1000 Monte Carlo runs. The results are in line with those of the previous figure, where we observe a performance degradation as N is increased, which confirms our expectations. An approach to overcome such a performance degradation is to partition the IRS into groups, and activate/deactivate each group in a sequential way in the time domain, so that the at each time, the sub-channels associated with a given group are estimated [22], [55], [56]. This approach, however, would increase the total training time by a factor corresponding to the number of groups. The following experiments compare the average runtime of KRF and BALS. The results are depicted in Figure 5, and corroborate the higher complexity of BALS compared to KRF. Note that the runtime of BALS grows faster than that of KRF with the increase on the number N of IRS elements. On the other hand, as we pointed out earlier (comparison between

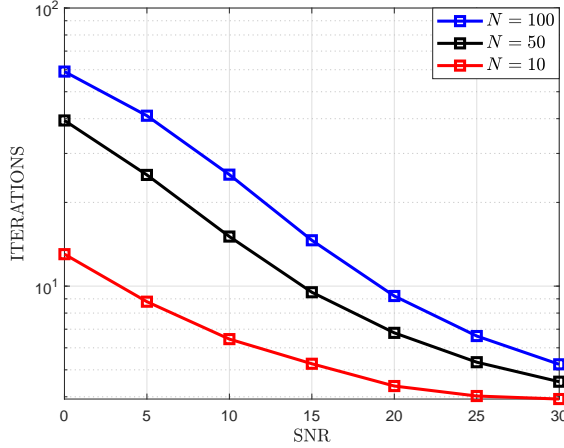


Fig. 6. Number of iterations to convergence of the BALS algorithm.

(31) and (32)), BALS can operate under less restrictive choices (smaller values) for K in comparison to KRF. Hence, there is a tradeoff between complexity and operating conditions for the two proposed channel estimation methods.

In Figure 6, we evaluate the required number of iterations of the BALS algorithm to achieve the convergence according to the criterion discussed in Section III-B. We can note that the required number of iterations grows with N , as expected. The difference in the convergence speed for different values of N is more pronounced in the low SNR range. For high SNRs, the convergence becomes less sensitive to N .

In Figure 7, we consider the uplink multi-BS scenario, which follows the signal model (50). We assume $P = 2$, $M = 1$, and $U = 1$. We compare the KRF receiver with a competing channel estimation method proposed recently in [57], which considers the single-antenna multi-user reception scenario⁵. Therein, the channel estimation requires three sequential stages, i.e., three time windows. In the first stage, the direct channel is estimated. In the second one, the equivalent channel between the first user and the BS is estimated. Finally, in the third stage, the channel associated with the remaining users are estimated. Similar to our model, in [57] the equivalent channel is obtained by stacking the contributions of the U users, i.e., $\mathbf{G} \diamond \mathbf{H} = [(\mathbf{H}\text{diag}(\mathbf{g}_1))^T, \dots, (\mathbf{H}\text{diag}(\mathbf{g}_U))^T]^T$. We can see that KRF outperforms the competing method, providing an SNR gain of nearly 5dB. Indeed, KRF jointly estimates all the involved channels in a single training stage, while in [57] the channel estimation is carried out in a sequential way, which can induce error propagation. This is a key difference that explains the performance gap in Figure 7.

In Figure 8, we compare the results of the proposed KRF method with the conventional LS method. In this experiment, we consider $K = N = 50$, $T = M = 20$, $L = 8$, and 1000 Monte Carlo runs. The CRB derived in Appendix A (equations

⁵In [57] the authors assume multiple receiving single-antenna UTs and a single multi-antenna BS in the downlink, while our model (50) assumes multiple receiving BSs and a single multi-antenna UT in the uplink. Due to the channel reciprocity assumption, the signal model of [57] is equivalent to our signal model (50). For a fair comparison, we assume $P = 2$ UTs for the channel estimation method of [57]. In this case, the dimensions of the channel matrices are exactly the same for both methods.

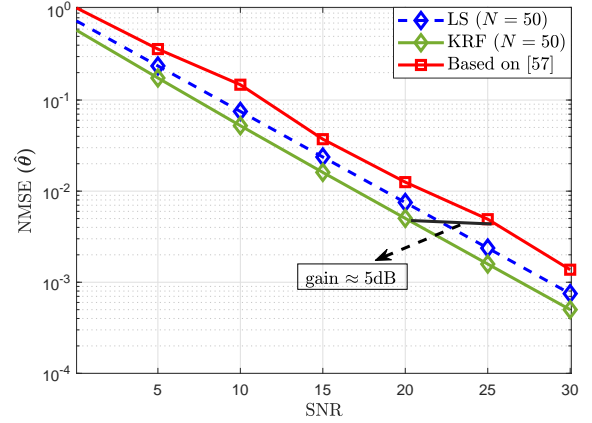
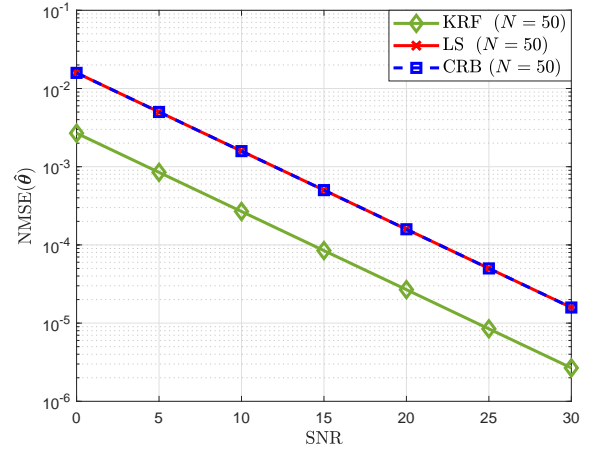


Fig. 7. NMSE of the estimated cascaded channel via the method of [57].

Fig. 8. Normalized CRB for the equivalent Khatri-Rao channel θ .

(68)-(68)) is also plotted here as a reference for comparison. Recall that the CRB considers the equivalent linear model obtained from the vectorized version of the received signal model given in (15), which we repeat here for convenience

$$\mathbf{y} = \text{vec}(\mathcal{Y}) = \mathbf{U}\boldsymbol{\theta} + \mathbf{b},$$

where $\mathbf{U} = (\mathbf{S} \otimes \tilde{\mathbf{X}})$ and $\boldsymbol{\theta} = \text{vec}(\mathbf{H}^T \diamond \mathbf{G}) \in \mathbb{C}^{MNL}$ is the parameter vector consisting of a vectorized version of the (Khatri-Rao structured) channel matrix combining the IRS-UT and the BS-IRS channel matrices. The conventional LS method plotted in the figure estimates this vectorized channel parameter as $\hat{\boldsymbol{\theta}} = \mathbf{U}^\dagger \mathbf{y}$, which ignores the Khatri-Rao structure that is lost in the vectorization of the signal model. In contrast, the proposed KRF method exploits the Khatri-Rao channel structure, and builds $\hat{\boldsymbol{\theta}}$ from the decoupled estimates $\hat{\mathbf{H}}$ and $\hat{\mathbf{G}}$ obtained according to Algorithm 1.

We can see that the LS solution attains the CRB. Furthermore, an interesting result can be noted. The proposed KRF algorithm outperforms the LS solution. The gain in terms of SNR is around 7dB. This result is explained by the fact that KRF effectively exploits the Khatri-Rao structure that is present in the equivalent channel model. Note that the KRF method solves the problem by reshaping the $ML \times N$

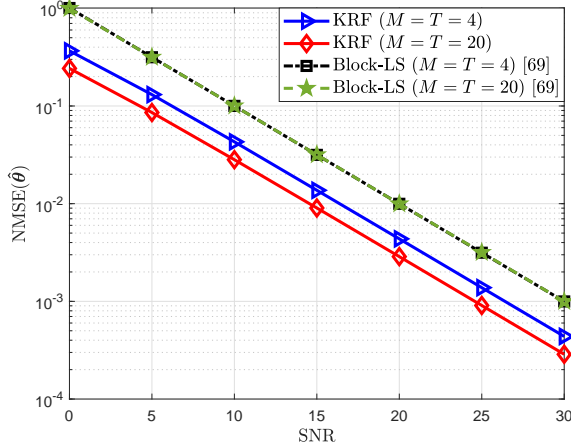


Fig. 9. NMSE of the $\hat{\theta}$ assuming that $\hat{\mathbf{H}}$ and $\hat{\mathbf{G}}$ are rank-deficient.

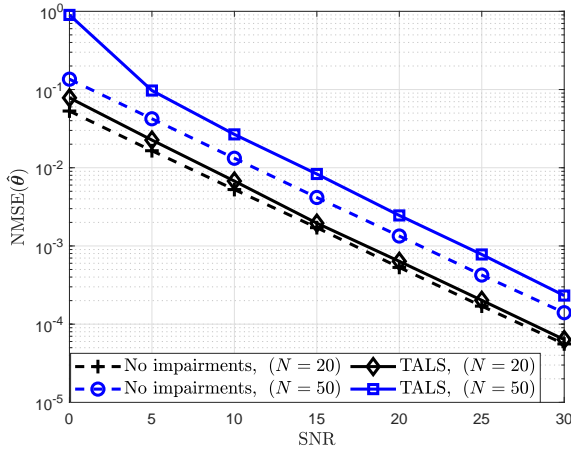


Fig. 10. TALS performance under IRS amplitude/phase perturbations.

Khatri-Rao channel in the form of N IRS subchannels of dimension $M \times L$, which provides a noise rejection gain thanks to the rank-1 approximation steps. Naturally, when M and L increase (which is the case, for instance, when assuming massive antenna arrays at the BS and UT), the larger is the spreading of the noise across the noise subspace and, consequently, higher levels of noise rejection will be achieved. This is a distinctive feature of the KRF method that is not exploited by the conventional LS channel estimator.

In Figure 9, we assume that the channel matrices \mathbf{H} and \mathbf{G} are rank-deficient. In this experiment, the channel matrices are generated according to the model given in (40)-(41). We assume uniform linear arrays at the BS and UT. The IRS has a uniform rectangular array structure. The angles of arrival (AoAs) and angles of departure (AoDs) are randomly generated according to a uniform distribution. At each Monte Carlo run, the azimuth and elevation angles are drawn within the intervals $[-\pi/2, \pi/2]$ and $[0, \pi/2]$, respectively. We consider a single path scenario, where $R_1 = R_2 = 1$, and assume $K = N = 64$, $L = 4$ and $T = M \in \{4, 20\}$. As a reference for comparison, we also plot the NMSE of the LS-based channel estimation method proposed in [58].

Therein, the time-domain pilot protocol is the same as the one considered in this work, which consists of dividing the total training time into K blocks across which the phase shift pattern of the IRS is varied. In [58], a two-stage scheme is proposed. In the first stage, the cascaded channel $\mathbf{C}_k = \mathbf{G}\mathbf{D}_k(\mathbf{S})\mathbf{H}$ associated with every time block k is individually estimated via an LS method. We refer to this approach as a “block-LS” method. The second stage extracts the AoA and AoD parameters by combining the K cascaded channel matrices. Since our method does not estimate the angular parameters of the channel matrices, we compare the proposed KRF method with the first stage of the block-LS method of [58], which also provides the unstructured estimate of the channel matrices \mathbf{H} and \mathbf{G} . We can see that KRF outperforms block-LS in the two considered system setups. Note that the performance of the block-LS method is not affected when the number M of transmit antennas (assuming $T = M$) is increased. This is in contrast to the KRF method that provides more accurate channel estimates for larger antenna arrays. In particular, the SNR gain of KRF over block-LS is nearly 3.5 dB for $M = T = 4$, and increases to 5.5 dB for $M = T = 20$. Indeed, higher values of M and/or L imply higher levels of noise rejection provided by the KRF method *via* exploiting the Khatri-Rao structure of the cascaded channel. These gains come at the expense of an increased computational complexity, as well as an increase on the length of the pilot sequences. Note that the channel ranks R_1 and R_2 do not need to be known by our channel estimation methods. Nevertheless, a performance enhancement could be obtained by exploiting the knowledge of these ranks (see, e.g. methods like [59]), or, alternatively, by means of compressed sensing methods that capitalize on sparse representations of the channel matrices \mathbf{H} and \mathbf{G} . This is an interesting topic for a future research.

In Figure 10, we assume that the IRS is affected by amplitude and phase perturbations, as well as unknown blockages, due to hardware and/or environmental-induced impairments. In this scenario, the receiver has an imperfect knowledge of the IRS phase shift matrix. The estimation/refinement of these phase shifts is carried out using the TALS algorithm, which extends BALS by including an additional LS estimation step associated with the update of \mathbf{S} within the loop, as discussed in Section III-D. To model these impairments, we assume $s_{k,n} = (a_{k,n}f_{k,n})\bar{s}_{k,n}$, where $\bar{s}_{k,n}$ is the originally designed phase shift (i.e., (k,n) -th element of a DFT matrix), $a_{k,n} \in \{0, 1\}$ models the presence or not of a blockage at the n -th element and k -th block, $f_{k,n} \sim \mathcal{CN}(0, \gamma)$ models the hardware impairments [60], [61], and γ denotes the variance of these perturbations. In the TALS algorithm, the initialization of $\hat{\mathbf{S}}$ is chosen as a DFT matrix, following its original design. This initialization always provides better results than a random one. Our simulations assume $T = M = 50$, $K = 100$, $L = 4$, $\gamma = 0.01$, 20% blocked IRS elements, and i.i.d. channel matrices. Although a performance degradation is observed in comparison with the perfectly-known IRS phase shifts case, we can see that the TALS algorithm can handle this more challenging scenario. In particular, the NMSE gap with respect to the ideal case increases as more elements are used in the IRS. Indeed, in

the imperfect IRS scenario, the total number of parameters to be estimated by the TALS algorithm is $(K + M + L)N$, in contrast to $(M + L)N$ in the perfectly-known IRS case, implying an addition of KN unknown factors.

VII. CONCLUSION AND PERSPECTIVES

We have proposed novel pilot-assisted receiver designs for IRS-assisted MIMO communication systems via a tensor modeling approach. The proposed KRF and BALS receivers effectively exploit the tensor structure that is present in the received signal. Both solutions yield decoupled estimates of the BS-IRS and IRS-UT channels at the receiver for a passive IRS. The closed-form KRF method has a lower complexity but a more restrictive requirement on the training parameter K , while the iterative BALS method, although being more computationally complex, can operate under more flexible choices for this parameter with a lower training overhead. Our design recommendations provide useful conditions on the system parameters that guarantee the uniqueness of the channel estimates. Our numerical results have demonstrated the superior performance of KRF and BALS compared to the conventional LS estimator, which ignores the Khatri-Rao structure of the combined channel matrix. In addition, the proposed tensor modeling approach allows to deal with a nonideal setup where the IRS phase shifts are not perfectly known at the receiver due to phase perturbations/fluctuations. In this more difficult setup, leveraging the trilinear structure of the received signal by means of a TALS algorithm provides us a joint estimation of the channel matrices and the IRS phase shift matrix. The proposed solutions also provide better results than recently proposed competing methods. Generalizations of our tensor modeling approach to multi-user scenarios have also been discussed, and analytical expressions for the CRB have been derived. The proposed approach can easily be extended to better deal with the millimeter wave scenario by assuming hybrid analog digital structures at the BS and UT sides. Combining the proposed algorithms with compressed sensing methods could provide further performance enhancements for low-rank (sparse) channels. In addition, leveraging to data-driven receivers capable of a joint channel estimation and symbol recovery would be desirable to save training resources. To this end, an extension of the proposed tensor modeling approach to the semi-blind case is a perspective of this work.

APPENDIX A

EXPECTED CRAMÉR RAO LOWER BOUND

In the following, we derive the closed-form CRB expressions for the channel estimation problem proposed in this work. The CRB provides the lower bound on the variance of achieved by an unbiased estimator. If $\hat{\boldsymbol{\theta}}$ is an unbiased estimate of $\boldsymbol{\theta}$, the NMSE measurements is lower bounded by the CRB such as,

$$\mathbb{E}\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2 \geq \text{Tr}\{\text{CRB}(\boldsymbol{\theta})\}, \quad (53)$$

where $\text{CRB}(\boldsymbol{\theta})$ is given as the inverse of the Fisher Information Matrix (FIM), denoted by $\mathbf{F}(\boldsymbol{\theta})$, such as

$$\text{CRB}(\boldsymbol{\theta}) \geq \mathbf{F}(\boldsymbol{\theta})^{-1}. \quad (54)$$

An extension for complex-valued parameters is derived in [62] by working on the structured parameter vector $\boldsymbol{\theta}_c = [\bar{\boldsymbol{\theta}}^T \tilde{\boldsymbol{\theta}}^T]^T$, where $\bar{\boldsymbol{\theta}} = \text{Re}(\boldsymbol{\theta})$, and $\tilde{\boldsymbol{\theta}} = \text{Im}(\boldsymbol{\theta})$. Thereby, with a nuisance parameter γ , the CRB for complex-valued random parameters is given as

$$\mathbb{E}\|\boldsymbol{\theta}_c - \hat{\boldsymbol{\theta}}_c\|^2 \geq \mathbb{E}_{\bar{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}, \gamma} \left\{ \text{Tr}\{\text{CRB}(\bar{\boldsymbol{\theta}})\} + \text{Tr}\{\text{CRB}(\tilde{\boldsymbol{\theta}})\} \right\}. \quad (55)$$

For an observation vector that follows a complex circular Gaussian distribution, $\mathbf{y} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{R})$, a useful way used to obtain the FIM is to use the Slepian-Bangs (SB) formula [63]:

$$[\mathbf{F}(\boldsymbol{\theta})]_{i,j} = 2\text{Re} \left\{ \left(\frac{\partial \boldsymbol{\mu}}{\partial [\boldsymbol{\theta}]_i} \right)^H \mathbf{R}^{-1} \left(\frac{\partial \boldsymbol{\mu}}{\partial [\boldsymbol{\theta}]_j} \right) \right\} \quad (56)$$

$$+ \text{Tr} \left\{ \left(\frac{\partial \mathbf{R}}{\partial [\boldsymbol{\theta}]_i} \right) \mathbf{R}^{-1} \left(\frac{\partial \mathbf{R}}{\partial [\boldsymbol{\theta}]_j} \right) \mathbf{R}^{-1} \right\}. \quad (57)$$

Let us recall (22):

$$\begin{aligned} [\mathcal{Y}]_{(3)} &= \mathbf{S} [(\mathbf{X} \otimes \mathbf{I}_L) (\mathbf{H}^T \diamond \mathbf{G})]^T \\ &= \mathbf{S} (\mathbf{H}^T \diamond \mathbf{G})^T (\mathbf{X} \otimes \mathbf{I}_L)^T, \end{aligned} \quad (58)$$

or, equivalently,

$$[\mathcal{Y}]_{(3)}^T = (\mathbf{X} \otimes \mathbf{I}_L) (\mathbf{H}^T \diamond \mathbf{G}) \mathbf{S}^T. \quad (59)$$

Considering the vectorized version of the 3-mode unfolding $[\mathcal{Y}]_{(3)}^T$, the following linear model with respect to the parameters of interest is obtained according to

$$\mathbf{y} = \text{vec}([\mathcal{Y}]_{(3)}^T) = \mathbf{U}\boldsymbol{\theta}, \quad (60)$$

where $\mathbf{U} = (\mathbf{S} \otimes \mathbf{X} \otimes \mathbf{I}_L)$, and

$$\boldsymbol{\theta} = \text{vec}(\mathbf{H}^T \diamond \mathbf{G}) \in \mathbb{C}^{MNL} \quad (61)$$

denotes the vectorized version of the Khatri-Rao structured channel. From the observation vector \mathbf{y} given by (60), the statistics of the noisy observation is given by

$$\mathbf{y} \sim \mathcal{CN}(\boldsymbol{\mu}_1, \mathbf{R}_1), \quad (62)$$

where,

$$\boldsymbol{\mu}_1 = \mathbf{U}\boldsymbol{\theta}, \quad (63)$$

$$\mathbf{R}_1 = \sigma^2 \mathbf{I}. \quad (64)$$

As \mathbf{R}_1 parameter-invariant, the second term of the SB formula vanishes, hence the $(2MNL) \times (2MNL)$ FIM, obtained after the calculation from (56), is given by

$$\mathbf{F}(\boldsymbol{\theta}_c) = \frac{2}{\sigma^2} \begin{bmatrix} \text{Re}\{\mathbf{U}^H \mathbf{U}\} & -\text{Im}\{\mathbf{U}^H \mathbf{U}\} \\ \text{Im}\{\mathbf{U}^H \mathbf{U}\}^T & \text{Re}\{\mathbf{U}^H \mathbf{U}\} \end{bmatrix}. \quad (65)$$

Considering the trace and the inverse of a 2×2 block matrix, we obtain

$$\text{Tr}\{\text{CRB}(\bar{\boldsymbol{\theta}})\} = \frac{\sigma^2}{2} \text{Tr} \left\{ \left(\bar{\mathbf{M}} + \bar{\mathbf{M}} \bar{\mathbf{M}}^{-1} \bar{\mathbf{M}} \right)^{-1} \right\}, \quad (66)$$

$$\begin{aligned} \text{Tr}\{\text{CRB}(\tilde{\boldsymbol{\theta}})\} &= \frac{\sigma^2}{2} \text{Tr} \left\{ \bar{\mathbf{M}}^{-1} - \bar{\mathbf{M}}^{-1} \bar{\mathbf{M}} (\bar{\mathbf{M}} + \right. \\ &\quad \left. + \bar{\mathbf{M}} \bar{\mathbf{M}}^{-1} \bar{\mathbf{M}})^{-1} \bar{\mathbf{M}} \bar{\mathbf{M}}^{-1} \right\}, \end{aligned} \quad (67)$$

where $\bar{\mathbf{M}} = \text{Re}\{\mathbf{U}^H \mathbf{U}\}$ and $\tilde{\mathbf{M}} = \text{Im}\{\mathbf{U}^H \mathbf{U}\}$. Let us recall that $\mathbf{U} = (\mathbf{S} \otimes \mathbf{X} \otimes \mathbf{I}_L)$, $\mathbf{X}^H \mathbf{X} = T \mathbf{I}_M$, and $\mathbf{S}^H \mathbf{S} = K \mathbf{I}_N$. Hence, $\mathbf{U}^H \mathbf{U} = K T \mathbf{I}_{MNL}$. This implies that $\tilde{\mathbf{M}} = \mathbf{0}$. The two above expressions can be simplified as

$$\text{CRB}(\bar{\boldsymbol{\theta}}) = \frac{\sigma^2}{2KT} \mathbf{I}_{MNL}, \quad (68)$$

$$\text{CRB}(\tilde{\boldsymbol{\theta}}) = \frac{\sigma^2}{2KT} \mathbf{I}_{MNL}. \quad (69)$$

Therefore, using definition (55):

$$\mathbb{E} \|\boldsymbol{\theta}_c - \hat{\boldsymbol{\theta}}_c\|^2 \geq \frac{\sigma^2}{KT} MNL. \quad (70)$$

It is important to note that there is no need to derive the mathematical expectation in the right-hand side of (55) over the parameters of interest and of nuisance due to the simple expression of the CRB.

APPENDIX B SIMPLIFIED VERSION OF BALS

Under the column-orthogonality assumption for \mathbf{X} and \mathbf{S} , the right pseudo-inverses in (29) and (30) can be replaced by lower complexity matrix products, leading to a faster implementation of the BALS algorithm. Defining $\mathbf{M}_1 \doteq \mathbf{S} \diamond \mathbf{X} \mathbf{H}^T$ and $\mathbf{M}_2 \doteq \mathbf{S} \diamond \mathbf{G}$, and using property (2), we have

$$\begin{aligned} \mathbf{M}_1^H \mathbf{M}_1 &= (\mathbf{S}^H \mathbf{S}) \odot (\mathbf{H} \mathbf{X}^H \mathbf{X} \mathbf{H}^H) \\ &= KT \begin{bmatrix} \|\mathbf{h}_1\|^2 & & \\ & \ddots & \\ & & \|\mathbf{h}_N\|^2 \end{bmatrix} \doteq KT \boldsymbol{\Sigma}_{\mathbf{H}} \end{aligned} \quad (71)$$

and

$$\begin{aligned} \mathbf{M}_2^H \mathbf{M}_2 &= (\mathbf{S}^H \mathbf{S}) \odot (\mathbf{G}^H \mathbf{G}) \\ &= K \begin{bmatrix} \|\mathbf{g}_1\|^2 & & \\ & \ddots & \\ & & \|\mathbf{g}_N\|^2 \end{bmatrix} \doteq K \boldsymbol{\Sigma}_{\mathbf{G}}, \end{aligned} \quad (72)$$

which implies that

$$\hat{\mathbf{G}} = (1/KT) \cdot \mathbf{Y}_1 \mathbf{M}_1^* \boldsymbol{\Sigma}_{\mathbf{H}}^{-1} \quad (73)$$

$$\hat{\mathbf{H}}^T = (1/KT) \cdot \mathbf{X}^H \mathbf{Y}_2 \mathbf{M}_2^* \boldsymbol{\Sigma}_{\mathbf{G}}^{-1}. \quad (74)$$

Due to the diagonal structure of $\boldsymbol{\Sigma}_{\mathbf{H}}$ and $\boldsymbol{\Sigma}_{\mathbf{G}}$, these expressions provide lower complexity implementations of (29) and (30), respectively, by replacing matrix inversions by simpler matrix products. In particular, each update of $\hat{\mathbf{G}}$ and $\hat{\mathbf{H}}$ can be viewed as a set of N independent processes (one for each IRS element) that can be carried out in parallel. The BALS is summarized in Algorithm 3.

REFERENCES

- [1] T. L. Singal, *Wireless Communications*. Tata McGraw Hill Education Private Limited, 2010.
- [2] E. Basar, M. D. Renzo, J. D. 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.
- [3] S. Gong, X. Lu, D. T. Hoang, D. Niyato, L. Shu, D. I. Kim, and Y. C. Liang, "Toward smart wireless communications via intelligent reflecting surfaces: A contemporary survey," *IEEE Communications Surveys Tutorials*, vol. 22, no. 4, pp. 2283–2314, 2020.

Algorithm 3: Simplified BALS

Procedure

input : $i = 0$; Initialize $\hat{\mathbf{H}}_{(i=0)}$

output: $\hat{\mathbf{H}}, \hat{\mathbf{G}}$

begin

$i = i + 1$;

while $\|e(i) - e(i-1)\| \geq \delta$ **do**

1: Compute $\hat{\mathbf{M}}_{1(i)} = \mathbf{S} \diamond \mathbf{X} \hat{\mathbf{H}}_{(i-1)}$
and find a least squares estimate of \mathbf{G} :

$$\hat{\mathbf{G}}_{(i)} = \frac{1}{KT} \mathbf{Y}_1 \hat{\mathbf{M}}_{1(i)}^* \hat{\boldsymbol{\Sigma}}_{\mathbf{H}(i-1)}^{-1}$$

2: Compute $\hat{\mathbf{M}}_{2(i)} = \mathbf{S} \diamond \mathbf{X} \hat{\mathbf{G}}_{(i)}$
and find a least squares estimate of \mathbf{H} :

$$\hat{\mathbf{H}}_{(i)}^T = \frac{1}{KT} \mathbf{X}^H \mathbf{Y}_2 \hat{\mathbf{M}}_{2(i)}^* \hat{\boldsymbol{\Sigma}}_{\mathbf{G}(i)}^{-1}$$

3: Repeat steps 1 to 2 until convergence.

end

end

- [4] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. F. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Communications Magazine*, vol. 56, pp. 162–169, 2018.
- [5] M. Jung, W. Saad, Y. Jang, G. Kong, and S. Choi, "Performance analysis of large intelligent surfaces (LISs): Asymptotic data rate and channel hardening effects," *IEEE Trans. on Wireless Commun.*, vol. 19, no. 3, pp. 2052–2065, 2020.
- [6] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, p. 4157–4170, Aug 2019.
- [7] E. Basar, "Reconfigurable intelligent surface-based index modulation: A new beyond MIMO paradigm for 6G," *IEEE Trans. on Communications*, vol. 68, no. 5, pp. 3187–3196, 2020.
- [8] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106–112, 2020.
- [9] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. on Communications*, vol. 68, no. 3, pp. 1838–1851, 2020.
- [10] Y. Song, M. R. A. Khandaker, F. Tariq, and K.-K. Wong, "Truly intelligent reflecting surface-aided secure communication using deep learning," 2020.
- [11] L. Dong and H.-M. Wang, "Secure MIMO transmission via intelligent reflecting surface," *IEEE Wireless Commun. Letters*, vol. 9, no. 6, pp. 787–790, 2020.
- [12] K. Ying, Z. Gao, S. Lyu, Y. Wu, H. Wang, and M. Alouini, "GMD-based hybrid beamforming for large reconfigurable intelligent surface assisted millimeter-wave massive MIMO," *IEEE Access*, vol. 8, pp. 19 530–19 539, 2020.
- [13] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," 2019.
- [14] T. L. Jensen and E. D. Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, pp. 5000–5004.
- [15] Z. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Commun. Letters*, vol. 9, no. 2, pp. 210–214, 2020.
- [16] B. Ning, Z. Chen, W. Chen, and Y. Du, "Channel estimation and transmission for intelligent reflecting surface assisted THz communications," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–7.
- [17] Y. Cui and H. Yin, "An efficient CSI acquisition method for intelligent reflecting surface-assisted mmWave networks," December 2019.

- [18] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, "Channel estimation for reconfigurable intelligent surface aided multi-user MIMO systems," 2019.
- [19] J. Mirza and B. Ali, "Channel estimation method and phase shift design for reconfigurable intelligent surface assisted MIMO networks," 2019.
- [20] S. Xia and Y. Shi, "Intelligent reflecting surface for massive device connectivity: Joint activity detection and channel estimation," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 5175–5179.
- [21] C. Hu and L. Dai, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," 2019.
- [22] C. You, B. Zheng, and R. Zhang, "Intelligent reflecting surface with discrete phase shifts: Channel estimation and passive beamforming," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.
- [23] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 4659–4663.
- [24] L. Wei, C. Huang, G. C. Alexandropoulos, and C. Yuen, "Parallel factor decomposition channel estimation in RIS-assisted multi-user MISO communication," in *IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Hangzhou, China, 2020.
- [25] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145–163, 2015.
- [26] 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 Trans. Signal. Process.*, vol. 65, no. 13, pp. 3551–3582, July 2017.
- [27] Y. Zniyed, R. Boyer, A. L. de Almeida, and G. Favier, "Tensor train representation of MIMO channels using the JIRAFE method," *Signal Processing, Elsevier*, vol. 171, Jun 2020.
- [28] Y. Zniyed, R. Boyer, A. L. F. Almeida, and G. Favier, "High-order tensor estimation via trains of coupled third-order CP and Tucker decompositions," *Linear Algebra and its Applications, Elsevier*, vol. 588, pp. 304–307, Mar 2020.
- [29] Y. Zniyed, R. Boyer, A. L. Almeida, and G. Favier, "A TT-based hierarchical framework for decomposing high-order tensors," *SIAM Journal on Scientific Computing*, vol. 42, no. 2, pp. A822–A848, 2020.
- [30] Y. Zniyed, R. Boyer, A. L. F. de Almeida, and G. Favier, "Tensor-train modeling for MIMO-OFDM tensor coding-and-forwarding relay systems," in *Proc. of European Signal Processing Conference (EUSIPCO)*, 2019.
- [31] A. L. F. de Almeida, X. Luciani, A. Stegeman, and P. Comon, "Confac decomposition approach to blind identification of underdetermined mixtures based on generating function derivatives," *IEEE Trans. Signal. Process.*, vol. 60, no. 11, pp. 5698–5713, 2012.
- [32] A. L. de Almeida, G. Favier, and J. C. Mota, "Space-time spreading-multiplexing for MIMO wireless communication systems using the PARATUCK-2 tensor model," *Signal Processing*, vol. 89, no. 11, pp. 2103–2116, 2009.
- [33] W. Freitas, G. Favier, and A. L. F. de Almeida, "Generalized Khatri-rao and Kronecker space-time coding for MIMO relay systems with closed-form semi-blind receivers," *Signal Processing*, vol. 151, pp. 19–31, 2018.
- [34] B. Sokal, A. L. de Almeida, and M. Haardt, "Semi-blind receivers for MIMO multi-relaying systems via rank-one tensor approximations," *Signal Processing*, vol. 166, Jan 2020.
- [35] F. E. D. Raimondi and P. Comon, "Tensor DoA estimation with directional elements," *IEEE Signal Proc. Letters*, vol. 24, no. 5, pp. 648–652, 2017.
- [36] L. N. Ribeiro, A. L. F. de Almeida, N. J. Myers, and R. W. Heath, "Tensor-based estimation of mmWave MIMO channels with carrier frequency offset," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 4155–4159.
- [37] H. Zheng, C. Zhou, Y. Gu, and Z. Shi, "Two-dimensional DOA estimation for coprime planar array: A coarray tensor-based solution," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 4562–4566.
- [38] D. C. Araújo, A. L. F. de Almeida, J. P. C. L. Da Costa, and R. T. de Sousa, "Tensor-based channel estimation for massive MIMO-OFDM systems," *IEEE Access*, vol. 7, pp. 42 133–42 147, 2019.
- [39] P. R. B. Gomes, A. L. F. de Almeida, J. P. C. L. da Costa, and R. T. de Sousa Jr, "Joint DL and UL channel estimation for millimeter wave MIMO systems using tensor modeling," *Wireless Communications and Mobile Computing*, vol. 2019, pp. 1–13, 2019.
- [40] G. T. de Araújo and A. L. F. de Almeida, "PARAFAC-based channel estimation for intelligent reflective surface assisted MIMO system," in *2020 IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2020, pp. 1–5.
- [41] 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.
- [42] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM REVIEW*, vol. 51, no. 3, pp. 455–500, 2009.
- [43] 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.
- [44] A. L. F. de Almeida, G. Favier, J. P. C. L. da Costa, and J. C. M. Mota, "Overview of tensor decompositions with applications to communications," in *Signals and Images: Advances and Results in Speech, Estimation, Compression, Recognition, Filtering, and Processing*, R. Coelho, V. Nascimento, R. de Queiroz, J. Romano, and C. Cavalcante, Eds. CRC-Press, Jan. 2016, no. Chapter 12, pp. 325–356.
- [45] J. B. Kruskal, "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra and its Applications*, vol. 18, no. 2, pp. 95 – 138, 1977.
- [46] A. Stegeman and N. D. Sidiropoulos, "On kruskal's uniqueness condition for the CANDECOMP/PARAFAC decomposition," *Linear Algebra and its Applications*, vol. 420, no. 2, pp. 540 – 552, 2007.
- [47] A. Y. Kibangou and G. Favier, "Non-iterative solution for PARAFAC with a toeplitz matrix factor," in *2009 17th European Signal Processing Conference*, Aug 2009, pp. 691–695.
- [48] F. Roemer and M. Haardt, "Tensor-based channel estimation and iterative refinements for two-way relaying with multiple antennas and spatial reuse," *IEEE Trans. Signal. Process.*, vol. 58, no. 11, pp. 5720–5735, Nov 2010.
- [49] G. H. Golub and C. F. van Loan, *Matrix Computations*, 4th ed. John Hopkins University Press, 2013.
- [50] R. Bro, "Multi- way analysis in the food industry: Models, algorithms & applications," Ph.D. dissertation, University of Amsterdam, Nov 1998.
- [51] B. Li, Z. Zhang, Z. Hu, and Y. Chen, "Joint array diagnosis and channel estimation for RIS-aided mmWave MIMO system," *IEEE Access*, vol. 8, pp. 193 992–194 006, 2020.
- [52] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of n-way arrays," *Journal of Chemometrics*, vol. 14, no. 3, pp. 229–239, 2000.
- [53] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms—part I: Lemmas for partitioned matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1022–1032, 2008.
- [54] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, 2016.
- [55] C. You, B. Zheng, and R. Zhang, "Channel estimation and passive beamforming for intelligent reflecting surface: Discrete phase shift and progressive refinement," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2604–2620, 2020.
- [56] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," *IEEE Trans. on Communications*, vol. 68, no. 7, pp. 4522–4535, 2020.
- [57] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, Seoul, Korea (South), 2020, pp. 1–6.
- [58] B. Deepak, R. S. P. Sankar, and S. P. Chepuri, "Channel estimation in reconfigurable intelligent surface assisted mmWave MIMO systems," 2020.
- [59] L. Cheng, X. Tong, S. Wang, Y. Wu, and H. V. Poor, "Learning nonnegative factors from tensor data: Probabilistic modeling and inference algorithm," *IEEE Trans. Signal. Process.*, vol. 68, pp. 1792–1806, 2020.
- [60] M. Badiu and J. P. Coon, "Communication through a large reflecting surface with phase errors," *IEEE Wireless Commun. Letters*, vol. 9, no. 2, pp. 184–188, 2020.
- [61] Y. Liu, E. Liu, and R. Wang, "Energy efficiency analysis of intelligent reflecting surface system with hardware impairments," in *2020 IEEE Global Communications Conference*, 2020, pp. 1–6.
- [62] P. M. R. de Oliveira, C. A. R. Fernandes, G. Favier, and R. Boyer, "PARATUCK semi-blind receivers for relaying multi-hop MIMO systems," *Digital Signal Processing*, vol. 92, pp. 127 – 138, 2019.
- [63] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Prentice Hall, Inc, 2005.