RESEARCH   **Open Access**

Channel estimation for MIMO multi-relay systems using a tensor approach

Xi Han[1*†], André LF de Almeida[2†] and Zhen Yang[3†]

Abstract

In this paper, we address the channel estimation problem for multiple-input multiple-output (MIMO) multi-relay systems exploiting measurements collected at the destination only. Assuming that the source, relays, and destination are multiple-antenna devices and considering a three-hop amplify-and-forward (AF)-based training scheme, new channel estimation algorithms capitalizing on a tensor modeling of the end-to-end communication channel are proposed. Our approach provides the destination with the instantaneous knowledge of all the channel matrices involved in the communication. Instead of using separate estimations for each matrix, we are interested in a joint estimation approach. Two receiver algorithms are formulated to solve the joint channel estimation problem. The first one is an iterative method based on a trilinear alternating least squares (TALS) algorithm, while the second one is a closed-form solution based on a Kronecker least squares (KRLS) factorization. A useful lower-bound on the channel training length is derived from an identifiability study. We also show the proposed tensor-based approach is applicable to two-way MIMO relaying systems. Simulation results corroborate the effectiveness of the proposed estimators and provide a comparison with existing methods in terms of channel estimation accuracy and bit error rate (BER).

Keywords: Channel estimation; Cooperative MIMO; Relaying; Tensor decomposition

1 Introduction

Cooperative communications have been considered as a promising concept to improve the link performance in modern wireless communication systems due to spatial diversity gains, enhanced coverage, and increased capacity [1-4]. In this context, relaying has been commonly accepted as a key technique to improve system performance by overcoming channel impairments, such as fading, shadowing, and path loss, in wireless fading channel environments [4-6]. By resorting to relay-assisted cooperation, multiple wireless links between mobile stations and base stations are established to create a virtual multiple-input multiple-output (MIMO) system [7]. In the simplest relay processing strategy, the relay stations amplify and forward the received data towards the base station. In this work, we adopt amplify-and-forward (AF) relaying due to its simplicity of implementation [5]. This strategy is preferable when fixed relay stations have

a limited computation capacity as opposed to the base station.

The overall link reliability of cooperative diversity schemes strongly depends on the accuracy of channel state information (CSI) associated with the multiple hops involved in the overall communication. Moreover, the use of common precoding techniques at the source and/or relays generally requires instantaneous CSI knowledge of the different channels to optimize transmission [8,9]. In practice, however, the CSI is unknown and has to be estimated with the aid of training sequences [10,11]. For two-hop relaying systems, the associated channel matrices can be estimated in separate LS estimation stages that operate sequentially at the destination [10]. When the communication involves additional hops, such a sequential LS estimation approach still applies by using additional transmission phases. The main problem is that channel estimation errors accumulate across the consecutive stages. In [11], a closed-form solution was proposed for the joint estimation of the channel matrices in a two-hop MIMO relaying system, avoiding error propagation.

*Correspondence: jinlucky333@gmail.com
†Equal contributors
[1]Information Network Center, Beijing University of Posts and Telecommunications, Beijing, China
Full list of author information is available at the end of the article

A few recent works have developed efficient receiver algorithms based on tensor analysis for channel estimation and/or symbol detection in cooperative systems [12-16]. In [12], a training sequence-based channel estimation algorithm is proposed for two-way relaying systems with multiple antennas at the relays. Recently [14], a channel estimation algorithm based on parallel factor (PARAFAC) model [17,18] was developed for two-hop MIMO relay systems. The approach allows estimation of the channel matrices associated with both hops by resorting to training sequences. Other few recent works have developed tensor-based receivers for one-way two-hop cooperative systems [13,15,16]. In particular, the approach of Ximenes et al. [16] assumes a Khatri-Rao space-time (KRST) coding [19] at the source node, and a semi-blind receiver is proposed by assuming the existence of a direct link between the source and the destination.

The approach of Roemer and Haardt and Rong et al. [12,14] allows a joint estimation of the channel matrices by resorting to training sequences. With the idea of avoiding the use of training sequences at the users' and relays' transmissions, the work [13] proposed a blind receiver for uplink multiuser cooperative diversity systems based on a PARAFAC model for the received signal. However, [13] is limited to a clustered relaying scenario, where relays belonging to the same cluster have the same spatial signature. The common feature of all these works is on the assumption of only two hops (source-to-relays and relays-to-destination). To further extend the coverage area and combat channel impairments such as path-loss and shadowing, it may be advantageous to introduce an additional hop along with an extra communication phase by means of three-hop relaying [5]. We highlight that the interest of the proposed work is on the joint channel estimation problem (i.e., joint channel and symbol estimation is not addressed here). The joint channel estimation problem was addressed in [12] for a two-way relaying system and in [14] for a one-way two-hop system. From a tensor modeling viewpoint, the common feature of both works is on the use of the PARAFAC model. Herein, we focus on a one-way three-hop multi-relay system, while resorting to a PARATUCK2 model to derive the proposed algorithms.

In this work, novel channel estimators are proposed for MIMO multi-relay systems. Assuming that the source, relays, and destination are multiple-antenna devices and considering a three-hop AF-based training scheme, new channel estimation algorithms capitalizing on a multilinear structure of the end-to-end communication channel are proposed. The proposed approach is based on a PARATUCK2 tensor model [20] of the data collected at the destination only, which allows the channel matrices to be jointly estimated at the destination. Two receiver algorithms are formulated to solve the channel estimation

problem. The first one is an iterative channel estimation method based on a trilinear alternating least squares (TALS) algorithm derived from a PARATUCK2 tensor model of the received data, while the second one is a closed-form solution based on a Kronecker least squares (KRLS) factorization. The proposed approach provides an extension of the idea recently proposed in [14] to a more general scenario with two-tier relaying using MIMO AF relays. Identifiability of the channel matrices is also examined in this work, and a useful lower-bound on the channel training length is derived. In contrast to conventional pilot-assisted LS channel estimation, where the channel matrices are estimated separately in consecutive stages, our proposed algorithms make a more efficient use of cooperative diversity by providing a joint estimation of all the channel matrices. As will be clear later, such a joint channel estimation is possible due to the use of the tensor approach to model the end-to-end system.

In comparison with conventional (multi-stage) LS channel estimation [10], the proposed tensor-based estimators have two distinguishing features: i) they avoid accumulation of channel estimation errors since all the channel matrices are estimated simultaneously (either iteratively or in closed-form), and ii) they can operate under less restrictive (and more flexible) conditions on the required number of antennas at the relays and/or destination, as will be clear from our identifiability analysis. Our approach also includes the PARAFAC-based channel estimator of [14] as a particular case. We also show that the proposed tensor modeling approach copes with a two-way MIMO multi-relaying communication system, where the TALS and KRLS channel estimators can be applied.

This paper is organized as follows. In section 2, the system model and working assumptions are described. Section 3 formulates the proposed approach. The data model is recast using tensor analysis, and the two channel estimation algorithms (TALS and KRLS) are derived. Identifiability of the channel matrices is also examined in this section. In section 4, we provide an extension of the proposed tensor-based signal model to a two-way MIMO relaying scenario. Numerical results are presented and discussed in section 5, and the conclusions are drawn in section 6.

Notation: Scalars are denoted by lowercase letters $(a, b, \ldots)$, vectors as lowercase boldface letters $(\mathrm{a}, \mathrm{b}, \ldots)$, matrices as uppercase boldface letters $(\mathbf{A}, \mathbf{B}, \ldots)$, and tensors as calligraphic letters $(\mathcal{A}, \mathcal{B}, \ldots)$. $\mathbf{A}^T$ and $\mathbf{A}^\dagger$ stand for transpose and pseudo-inverse of $\mathbf{A}$, respectively. To retrieve the element (i, j) of $\mathbf{A}$, we use $a(i, j)$. The ith row of $\mathbf{A} \in \mathbb{C}^{I \times R}$ is denoted as $\mathbf{A}_{(i,:)}$ while its rth column is denoted by $\mathbf{A}_{(:,r)}$. The operator $\mathrm{D}_i(\mathbf{A})$ forms a diagonal matrix out of the ith row of $\mathbf{A}$. The Khatri-Rao (columnwise Kronecker) product between $\mathbf{A} \in \mathbb{C}^{I \times R}$ and $\mathbf{B} \in$

$\mathbb{C}^{J \times R}$, i.e., $\mathbf{A} \diamond \mathbf{B} = \left[\mathbf{A}_{(:,1)} \otimes \mathbf{B}_{(:,1)}, \ldots, \mathbf{A}_{(:,R)} \otimes \mathbf{B}_{(:,R)} \right] \in \mathbb{C}^{IJ \times R}$.

2 System model

We consider a three-hop MIMO AF communication system where the source node transmits information to the destination node with the aid of R_1 relays in the first tier and R_2 relays in the second tier. As shown in Figure 1, the source and destination nodes are equipped with $N_s \geq 2$ and $N_d \geq 2$ antennas, respectively, and half-duplex relays are considered. The qth relay of tier 1, which receives data from the source node, is equipped with I_q antennas, $q = 1, \ldots, R_1$, while the pth relay of tier 2, which receives data from tier 1 relays, is equipped with J_p antennas, $p = 1, \ldots, R_2$. The total number of antennas that transmit in second and third phases are denoted by $N_1 = I_1 + \cdots + I_{R_1}$ and $N_2 = J_1 + \cdots + J_{R_2}$, respectively.

Some key assumptions are now given: (i) relays are synchronized at the symbol level. More specifically, the timing offset is assumed to be within one symbol period, so that timing information is acquired only through some form of (rough) coarse synchronization; (ii) fading is assumed to be frequency flat, and the data block size is smaller than the channel coherence time so that the channel is considered as time invariant; (iii) the direct links between the source (resp. tier 1 relays) and the destination node are not available[a]. This situation is evidenced in the current uplink of IEEE 802.16j.

2.1 Data model

The communication between source and destination is accomplished in three hops. In the first hop, the modulated signal vector $\mathbf{u}_s(t) \in \mathbb{C}^{N_s \times 1}$ is transmitted to R_1 relays. The received signal at the qth relay of tier 1 can be written as

$$\mathbf{y}_{sr}^{(q)}(t) = \mathbf{H}_{sr}^{(q)} \mathbf{u}_s(t) + \mathbf{v}_{sr}^{(q)}(t) \tag{1}$$

where $\mathbf{y}_{sr}^{(q)}(t) \in \mathbb{C}^{I_q \times 1}$ is the received signal vector at the qth relay of tier 1, $\mathbf{H}_{sr}^{(q)} \in \mathbb{C}^{I_q \times N_s}$ is the MIMO channel between the source and the qth tier 1 relay, and $\mathbf{v}_{sr}^{(q)}(t) \in \mathbb{C}^{I_q \times 1}$ is an additive noise vector. Noise samples are modeled as independent and identically distributed complex Gaussian random variables with zero mean and unit variance.

In the second hop, the source stops transmission and all the R_1 relays of tier 1 amplify their received signals with diagonal AF matrices $\mathbf{G}^{(1)}, \ldots, \mathbf{G}^{(R_1)}$ and simultaneously forward the resulting signals to the tier 2 relays. The received signal vector at the pth relay of tier 2 is then given by

$$\mathbf{y}_{rr}^{(p)}(t+1) = \sum_{q=1}^{R_1} \mathbf{H}_{rr}^{(p,q)} \mathbf{G}^{(q)} \mathbf{y}_{sr}^{(q)}(t) + \mathbf{v}_{rr}^{(p)}(t+1) \tag{2}$$

where $\mathbf{H}_{rr}^{(p,q)} \in \mathbb{C}^{J_p \times I_q}$ is the MIMO channel linking the R_1 tier 1 relays to R_2 tier 2 relays, while $\mathbf{v}_{rr}^{(p)}(t+1) \in \mathbb{C}^{J_p \times 1}$ denotes the corresponding noise vector. In the third hop, the source and all tier 1 relays are silent, while the tier 2 relays process the received signal vector with the diagonal AF matrices $\mathbf{J}^{(1)}, \ldots, \mathbf{J}^{(R_2)}$ and forward their amplified signals to the destination. The received signal vector at the destination is then given by

$$\mathbf{y}_{rd}(t+2) = \sum_{p=1}^{R_2} \mathbf{H}_{rd}^{(p)} \mathbf{J}^{(p)} \mathbf{y}_{rr}^{(p)}(t+1) + \mathbf{v}_{rd}(t+2), \tag{3}$$

where $\mathbf{H}_{rd}^{(p)} \in \mathbb{C}^{N_d \times J_p}$ is the MIMO channel linking the pth tier 1 relay to the destination, and $\mathbf{v}_{rd}(t+2) \in \mathbb{C}^{N_d \times 1}$ the corresponding additive noise term.

Let us define the multi-relay (block) channel matrices

$$\mathbf{H}_{rd} \doteq \left[\mathbf{H}_{rd}^{(1)}, \ldots, \mathbf{H}_{rd}^{(R_2)} \right] \in \mathbb{C}^{N_d \times N_2}, \tag{4}$$

$$\mathbf{H}_{rr} \doteq \begin{bmatrix} \mathbf{H}_{rr}^{(1,1)} & \cdots & \mathbf{H}_{rr}^{(1,R_1)} \\ \vdots & \vdots & \vdots \\ \mathbf{H}_{rr}^{(R_2,1)} & \cdots & \mathbf{H}_{rr}^{(R_2,R_1)} \end{bmatrix} \in \mathbb{C}^{N_2 \times N_1}, \tag{5}$$

$$\mathbf{H}_{sr}^{T} \doteq \left[\mathbf{H}_{sr}^{(1)T}, \ldots, \mathbf{H}_{sr}^{(R_1)T} \right] \in \mathbb{C}^{N_s \times N_1}, \tag{6}$$
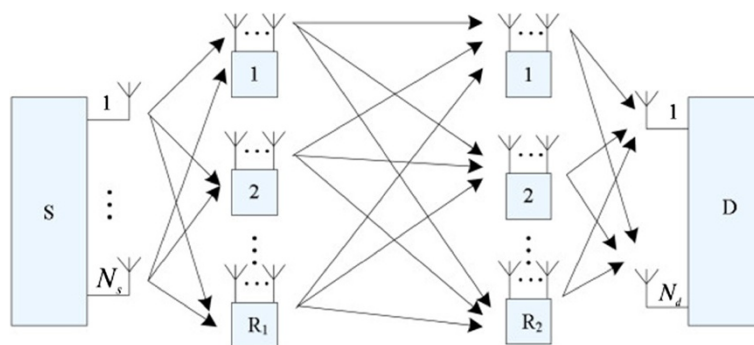

Figure 1 Block diagram of the considered MIMO multi-relay system.

and let $\mathbf{G} \doteq \mathrm{bdiag}\left[\mathbf{G}^{(1)}, \ldots, \mathbf{G}^{(R_1)}\right] \in \mathbb{C}^{N_1 \times N_1}$ and $\mathbf{J} \doteq \mathrm{bdiag}\left[\mathbf{J}^{(1)}, \ldots, \mathbf{J}^{(R_2)}\right] \in \mathbb{C}^{N_2 \times N_2}$ be the two diagonal matrices that collect the AF coefficients of the overall multi-relay system. Using these definitions, and using (1) and (2), we can rewrite (3) as follows:

$$\mathbf{y}_{\mathrm{rd}}(t+2) = \mathbf{H}_{\mathrm{rd}}\mathbf{J}\mathbf{H}_{\mathrm{rr}}\mathbf{G}\mathbf{H}_{\mathrm{sr}}\mathbf{u}_s(t) + \bar{\mathbf{v}}_{\mathrm{rd}}(t+2), \qquad (7)$$

where $\bar{\mathbf{v}}_{\mathrm{rd}}(t+2) = \bar{\mathbf{v}}_{\mathrm{sr}}(t) + \bar{\mathbf{v}}_{\mathrm{rr}}(t+1) + \mathbf{v}_{\mathrm{rd}}(t+2)$ is the total noise at the destination, which contains the filtered noise contributions from the multiple relays, with $\bar{\mathbf{v}}_{\mathrm{sr}}(t) = \mathbf{H}_{\mathrm{rd}}\mathbf{J}\mathbf{H}_{\mathrm{rr}}\mathbf{G}\mathbf{v}_{\mathrm{sr}}$, $\bar{\mathbf{v}}_{\mathrm{rr}}(t+1) = \mathbf{H}_{\mathrm{rd}}\mathbf{J}\mathbf{v}_{\mathrm{rr}}(t+1)$, $\mathbf{v}_{\mathrm{rr}}(t+1) \doteq \left[\mathbf{v}_{\mathrm{rr}}^{(1)\mathrm{T}}(t+1), \ldots, \mathbf{v}_{\mathrm{rr}}^{(R_2)\mathrm{T}}(t+1)\right]^{\mathrm{T}} \in \mathbb{C}^{N_2 \times 1}$, $\mathbf{v}_{\mathrm{sr}}(t) \doteq \left[\mathbf{v}_{\mathrm{sr}}^{(1)\mathrm{T}}(t), \ldots, \mathbf{v}_{\mathrm{sr}}^{(R_1)\mathrm{T}}(t)\right]^{\mathrm{T}} \in \mathbb{C}^{N_1 \times 1}$.

Note that, since this work is concerned with channel estimation, the AF matrices $\mathbf{G}$ and $\mathbf{J}$ cannot be optimized at the transmission (source and relays). Therefore, for simplicity, we have assumed that these matrices are diagonal. The use of non-diagonal AF matrices in the proposed approach is left for a future work. Note also that, once the channels are estimated, the design of full AF matrices can be done, e.g., based on the SVD of the channel matrices, following the idea of [9] or on the mean-square error (MSE) criterion [21]. If simplified AF schemes are used, where only power allocation is done, $\mathbf{G}$ and $\mathbf{J}$ are diagonal matrices, the coefficients of which can be designed as a function of the mean channel and noise powers [5] or optimized from power allocation strategies, as shown recently in [22].

2.2 Conventional LS estimation method
The simplest approach to estimate the effective channel $\mathbf{H}_{\mathrm{eff}} = \mathbf{H}_{\mathrm{rd}}\mathbf{J}\mathbf{H}_{\mathrm{rr}}\mathbf{G}\mathbf{H}_{\mathrm{sr}}$ (including the amplifying factors) is based on training sequences. If separate estimations of the multi-relay channels $\mathbf{H}_{\mathrm{rd}}$, $\mathbf{H}_{\mathrm{rr}}$, and $\mathbf{H}_{\mathrm{sr}}$ are required, for instance, to optimize the source precoding matrix and the relays' AF matrices, three separate LS estimation stages should operate sequentially at the destination. The method would work similarly to that of Kong and Hua [10]. Denote $\mathbf{S}_0 \in \mathbb{C}^{N_s \times L_0}$ as the training sequence matrix sent by the source node, while $\mathbf{S}_{1\mathrm{d}} \in \mathbb{C}^{N_1 \times L_1}$ and $\mathbf{S}_{2\mathrm{d}} \in \mathbb{C}^{N_2 \times L_2}$ are the training sequence matrices sent by the relays at tiers 1 and 2, respectively. Assume that orthogonal training sequences are used in all stages, which implies training sequences of length $L_0 \geq N_s, L_1 \geq N_1$ and $L_2 \geq N_2$ at the source, tier 1 and tier 2 relays, respectively. In the first stage, $\mathbf{S}_{2\mathrm{d}}$ is transmitted from all tier 2 relays to the destination. The LS estimate of $\mathbf{H}_{\mathrm{rd}}$ is obtained as

$$\widehat{\mathbf{H}}_{\mathrm{rd}} = \mathbf{Y}_1 \mathbf{S}_{2\mathrm{d}}^{\mathrm{H}}, \qquad (8)$$

where $\mathbf{Y}_1 \in \mathbb{C}^{N_d \times L_2}$ is the received signal matrix at the destination during the first training stage. In the second stage, $\mathbf{S}_{1\mathrm{d}}$ is transmitted from all tier 1 relays to the destination via AF processing at the tier 1 relays. Defining $\mathbf{Y}_2 \in \mathbb{C}^{N_d \times L_1}$ as the data received from tier 1 relays at the second training stage, an LS estimate of $\mathbf{H}_{\mathrm{rr}}$ can be obtained as

$$\widehat{\mathbf{H}}_{\mathrm{rr}} = \left(\widehat{\mathbf{H}}_{\mathrm{rd}}\mathbf{J}\right)^{\dagger} \mathbf{Y}_2 \mathbf{S}_{1\mathrm{d}}^{\mathrm{H}}. \qquad (9)$$

Finally, $\mathbf{S}_0$ is transmitted from the source to the destination via the two tiers of relays. The destination collects the received data in $\mathbf{Y}_3 \in \mathbb{C}^{N_s \times L_0}$. An estimate of $\widehat{\mathbf{H}}_{\mathrm{sr}}$ is then found as

$$\widehat{\mathbf{H}}_{\mathrm{sr}} = \left(\widehat{\mathbf{H}}_{\mathrm{rd}}\mathbf{J}\widehat{\mathbf{H}}_{\mathrm{rr}}\mathbf{G}\right)^{\dagger} \mathbf{Y}_3 \mathbf{S}_0^{\mathrm{H}}. \qquad (10)$$

This method requires 6 transmission phases to provide the destination with all the channel matrices (1 phase for estimating $\mathbf{H}_{\mathrm{rd}}$, 2 phases for estimating $\mathbf{H}_{\mathrm{rr}}$ and 3 phases for estimating $\mathbf{H}_{\mathrm{sr}}$). Note that the channel estimation errors accumulate across the consecutive stages, due to the dependency between successive channel estimates. Moreover, this method requires $N_d \geq N_2 \geq N_1$ for the uniqueness of the LS estimates of $\widehat{\mathbf{H}}_{\mathrm{rr}}$ and $\widehat{\mathbf{H}}_{\mathrm{sr}}$. In the following, we adopt a different path to solve this problem by capitalizing on tensor analysis. The idea is to provide the destination with a joint estimate of all the partial channels $\mathbf{H}_{\mathrm{rd}}$, $\mathbf{H}_{\mathrm{rr}}$, and $\mathbf{H}_{\mathrm{sr}}$ by exploiting the tensor structure of the end-to-end signal model. The proposed approach allows channel estimation to be performed under less restrictive conditions on the number N_d of receive antennas at the destination compared with the conventional LS estimator, while avoiding error accumulation.

3 Proposed approach
In order to derive the proposed channel estimators, we first recast the formulation of the system model by resorting to multi-way (tensor) analysis. First, let us divide the overall training period into K time blocks. In every time block, the same training sequence matrix $\mathbf{S}_0 \in \mathbb{C}^{N_s \times L_0}$ is transmitted by the source node. In the kth time block, the relays of tiers 1 and 2 use the AF matrices $\mathbf{G}_k$ and $\mathbf{J}_k$, respectively, $k = 1, \ldots, K$. Let us define $\mathbf{E} \in \mathbb{C}^{K \times N_1}$ and $\mathbf{F} \in \mathbb{C}^{K \times N_2}$ as channel training matrices such that $\mathrm{D}_k(\mathbf{E}) \doteq \mathbf{G}_k$ and $\mathrm{D}_k(\mathbf{F}) \doteq \mathbf{J}_k$, where $\mathrm{D}_k(\cdot)$ forms a diagonal matrix out of the kth row of its matrix argument. Otherwise stated, the rows of $\mathbf{E}$ (resp. $\mathbf{F}$) hold the AF coefficients of the R_1 (resp. R_2) relays associated with the different time blocks. Then, the signal received at the destination during the kth time block can be written as:

$$\mathbf{Y}_k = \mathbf{H}_{\mathrm{rd}}\mathrm{D}_k(\mathbf{F})\mathbf{H}_{\mathrm{rr}}\mathrm{D}_k(\mathbf{E})\mathbf{H}_{\mathrm{sr}}\mathbf{S}_0 + \mathbf{V}_k, \qquad (11)$$
$$k = 1, \ldots, K,$$

where $\mathbf{V}_k = \mathbf{H}_{\mathrm{rd}}\mathrm{D}_k\left(\mathbf{F}\right)\mathbf{H}_{\mathrm{rr}}\mathrm{D}_k\left(\mathbf{E}\right)\mathbf{V}_{\mathrm{sr},k} + \mathbf{H}_{\mathrm{rd}}\mathrm{D}_k\left(\mathbf{F}\right)\mathbf{V}_{\mathrm{rr},k}$, $\mathbf{V}_{\mathrm{sr},k} \in \mathbb{C}^{N_1 \times L_0}$ is the noise matrix at the relays during the kth time block, $\mathbf{V}_{\mathrm{rr},k} \in \mathbb{C}^{N_2 \times L_0}$ is the noise matrix at the second hop relays for the k-th time block, and $\mathbf{V}_{\mathrm{rd},k} \in \mathbb{C}^{N_d \times L_0}$ is the noise matrix at the destination for the kth time block.

Regarding the structure of the channel training matrices $\mathbf{E} \in \mathbb{C}^{K \times N_1}$ and $\mathbf{F} \in \mathbb{C}^{K \times N_2}$, unless otherwise stated, their columns are chosen as length-K random sequences following a uniform distribution between $[-1, 1]$. These sequences are defined beforehand and known at the destination node. With such a choice, the signals transmitted by the relays across the K time blocks have random phases and are subject to limited power fluctuations. Clearly, this design is not optimal for minimizing the mean square error of the channel estimation. Determining an optimum design for these matrices is a difficult problem and is not pursued in this work. Nevertheless, extensive computer simulations have demonstrated that this choice yields very good results. For convenience, we will come back later to the problem of choosing $\mathbf{E}$ and $\mathbf{F}$ from a channel identifiability viewpoint. A more elaborated design of these matrices will be then proposed.

Upon reception of the data matrix $\mathbf{Y}_k$, $k = 1, \ldots, K$, an unstructured estimate of the end-to-end channel during the kth time block is first obtained at the destination. Multiplying both sides of (11) with the known training sequence matrix $\mathbf{S}_0^{\mathrm{H}}$ yields

$$\widehat{\mathbf{H}}_k = \mathbf{Y}_k\mathbf{S}_0^{\mathrm{H}} \in \mathbb{C}^{N_d \times N_s}$$
$$= \mathbf{H}_{\mathrm{rd}}\mathrm{D}_k\left(\mathbf{F}\right)\mathbf{H}_{\mathrm{rr}}\mathrm{D}_k\left(\mathbf{E}\right)\mathbf{H}_{\mathrm{sr}} + \mathbf{V}_k\mathbf{S}_0^{\mathrm{H}}, \quad (12)$$

$k = 1, \cdots, K$. Let us introduce

$$\widehat{\mathbf{H}}_k = \mathbf{H}_k + \mathbf{V}_k\mathbf{S}_0^{\mathrm{H}}, \tag{13}$$

where

$$\mathbf{H}_k = \mathbf{H}_{\mathrm{rd}}\mathrm{D}_k\left(\mathbf{F}\right)\mathbf{H}_{\mathrm{rr}}\mathrm{D}_k\left(\mathbf{E}\right)\mathbf{H}_{\mathrm{sr}}, \quad k = 1, \ldots, K, \tag{14}$$

is the matrix-of-interest that represents the effective end-to-end channel, $\mathbf{V}_k$ is the total noise matrix, and $\widetilde{\mathbf{H}}_k$ is the noisy observation of $\mathbf{H}_k$. We can assemble the set $\{\mathbf{H}_1, \cdots, \mathbf{H}_K\}$ to form a three-way array, or a third-order tensor, $\mathcal{H} \in \mathbb{C}^{N_d \times N_s \times K}$, whose dimensions are N_d (first dimension), N_s (second dimension), and K (third dimension).

Equation (14) corresponds to a PARATUCK2 model of the (noiseless) tensor $\mathcal{H}$ [23]. The PARATUCK2 model has first appeared in [20]. A more comprehensive formulation is given in [23], which also details an alternating least squares procedure for estimating its matrix factors.

Here, we show that this tensor model can be exploited to derive novel channel estimators for a cooperative MIMO relaying system.

Now, let us define

$$\mathbf{H}_{[1]} \doteq [\mathrm{vec}\left(\mathbf{H}_1\right), \cdots, \mathrm{vec}\left(\mathbf{H}_K\right)] \in \mathbb{C}^{N_d N_s \times K} \tag{15}$$

where $\mathbf{H}_{[1]}$ is a matrix 'unfolding' for the tensor $\mathcal{H}$ obtained by stacking column-wise its K slices. Define also

$$\mathbf{W}_k = \mathrm{D}_k\left(\mathbf{F}\right)\mathbf{H}_{\mathrm{rr}}\mathrm{D}_k\left(\mathbf{E}\right) \in \mathbb{C}^{N_2 \times N_1}. \tag{16}$$

Substituting (14) into (15), and applying property $\mathrm{vec}\left(\mathbf{ACB}\right) = \left(\mathbf{B}^{\mathrm{T}} \otimes \mathbf{A}\right)\mathrm{vec}(\mathbf{C})$, we get

$$\mathbf{H}_{[1]} = \left(\mathbf{H}_{\mathrm{sr}}^{\mathrm{T}} \otimes \mathbf{H}_{\mathrm{rd}}\right)\left[\mathrm{vec}\left(\mathbf{W}_1\right), \cdots, \mathrm{vec}\left(\mathbf{W}_K\right)\right]$$
$$= \left(\mathbf{H}_{\mathrm{sr}}^{\mathrm{T}} \otimes \mathbf{H}_{\mathrm{rd}}\right)\mathrm{diag}\left(\mathrm{vec}\left(\mathbf{H}_{\mathrm{rr}}\right)\right)\left(\mathbf{E}^{\mathrm{T}} \odot \mathbf{F}^{\mathrm{T}}\right) \tag{17}$$

where

$$\mathbf{E}^{\mathrm{T}} \odot \mathbf{F}^{\mathrm{T}} = \left[\mathbf{E}_{(1,:)}^{\mathrm{T}} \otimes \mathbf{F}_{(1,:)}^{\mathrm{T}}, \cdots, \mathbf{E}_{(K,:)}^{\mathrm{T}} \otimes \mathbf{F}_{(K,:)}^{\mathrm{T}}\right] \in \mathbb{C}^{N_2 N_1 \times K}, \tag{18}$$

$\mathbf{E}_{(k,:)} \in \mathbb{C}^{1 \times N_1}$ (resp. $\mathbf{F}_{(k,:)} \in \mathbb{C}^{1 \times N_2}$) denote the kth row of $\mathbf{E}$ (resp. $\mathbf{F}$), and $\odot$ is the Khatri-Rao (columnwise Kronecker) product.

Applying property $\mathrm{vec}\left(\mathbf{A}\mathrm{diag}(\mathbf{x})\mathbf{B}\right) = \left(\mathbf{B}^{\mathrm{T}} \odot \mathbf{A}\right)\mathbf{x}$, we get from (17) the following expression:

$$\mathrm{vec}(\mathbf{H}_{[1]}) = \mathbf{\Omega}_1\mathrm{vec}\left(\mathbf{H}_{\mathrm{rr}}\right), \tag{19}$$

where

$$\mathbf{\Omega}_1 = \left[\left(\mathbf{E}^{\mathrm{T}} \odot \mathbf{F}^{\mathrm{T}}\right)^{\mathrm{T}} \odot \left(\mathbf{H}_{\mathrm{sr}}^{\mathrm{T}} \otimes \mathbf{H}_{\mathrm{rd}}\right)\right] \in \mathbb{C}^{N_D N_s K \times N_1 N_2}. \tag{20}$$

In addition to the matrix unfolding $\mathbf{H}_{[1]}$, it is useful to define two other matrix unfoldings, which collect the information of tensor $\mathcal{H}$. Therefore, let us now define

$$\mathbf{H}_{[2]} \doteq \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_K \end{bmatrix} \in \mathbb{C}^{N_d K \times N_s}, \quad \mathbf{H}_{[3]} \doteq \begin{bmatrix} \mathbf{H}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{H}_K^{\mathrm{T}} \end{bmatrix} \in \mathbb{C}^{N_s K \times N_d}. \tag{21}$$

From (14) and (16), it follows that

$$\mathbf{H}_{[2]} = \begin{bmatrix} \mathbf{H}_{\mathrm{rd}}\mathbf{W}_1 \\ \vdots \\ \mathbf{H}_{\mathrm{rd}}\mathbf{W}_K \end{bmatrix}\mathbf{H}_{\mathrm{sr}} = \begin{bmatrix} \mathbf{H}_{\mathrm{rd}} & & \\ & \ddots & \\ & & \mathbf{H}_{\mathrm{rd}} \end{bmatrix}\begin{bmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_K \end{bmatrix}\mathbf{H}_{\mathrm{sr}} \tag{22}$$

and

$$\mathbf{H}_{[3]} = \begin{bmatrix} \mathbf{H}_{sr}^T \mathbf{W}_1^T \\ \vdots \\ \mathbf{H}_{sr}^T \mathbf{W}_K^T \end{bmatrix} \mathbf{H}_{rd}^T = \begin{bmatrix} \mathbf{H}_{sr}^T & & \\ & \ddots & \\ & & \mathbf{H}_{sr}^T \end{bmatrix} \begin{bmatrix} \mathbf{W}_1^T \\ \vdots \\ \mathbf{W}_K^T \end{bmatrix} \mathbf{H}_{rd}^T \quad (23)$$

or, more compactly,

$$\mathbf{H}_{[2]} = (\mathbf{I}_K \otimes \mathbf{H}_{rd})\, \mathbf{\Omega}_2 \mathbf{H}_{sr}, \quad (24)$$

$$\mathbf{H}_{[3]} = \left(\mathbf{I}_K \otimes \mathbf{H}_{sr}^T\right) \mathbf{\Omega}_3 \mathbf{H}_{rd}^T, \quad (25)$$

where

$$\mathbf{\Omega}_2 = \begin{bmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_K \end{bmatrix} \in \mathbb{C}^{N_2 K \times N_1}, \quad \mathbf{\Omega}_3 = \begin{bmatrix} \mathbf{W}_1^T \\ \vdots \\ \mathbf{W}_k^T \end{bmatrix} \in \mathbb{C}^{N_1 K \times N_2}. \quad (26)$$

3.1 Identifiability of channel matrices

Identifiability of $\mathbf{H}_{sr}$, $\mathbf{H}_{rr}$, and $\mathbf{H}_{rd}$ in the LS sense from $\mathbf{H}_{[1]}$, $\mathbf{H}_{[2]}$, and $\mathbf{H}_{[3]}$ (see Equations (19), (24), and (25)), respectively, requires that $\mathbf{\Omega}_1 = \left[\left(\mathbf{E}^T \odot \mathbf{F}^T\right)^T \odot \left(\mathbf{H}_{sr}^T \otimes \mathbf{H}_{rd}\right)\right] \in \mathbb{C}^{N_D N_s K \times N_1 N_2}$, $\mathbf{Z}_{[2]} \doteq (\mathbf{I}_K \otimes \mathbf{H}_{rd})\,\mathbf{\Omega}_2 \in \mathbb{C}^{N_d K \times N_1}$ and $\mathbf{Z}_{[3]} \doteq \left(\mathbf{I}_K \otimes \mathbf{H}_{sr}^T\right)\mathbf{\Omega}_3 \in \mathbb{C}^{N_s K \times N_2}$ be full column-rank. These requirements come from the fact that $\mathbf{\Omega}_1$, $\mathbf{Z}_{[2]}$, and $\mathbf{Z}_{[3]}$ must be left-invertible, from which the following necessary conditions are obtained:

$$N_d N_s K \geq N_1 N_2, \quad N_d K \geq N_1, \quad N_s K \geq N_2. \quad (27)$$

From the three inequalities and from the fact that we must have $K \geq 2$, the lower bound on the number K of time blocks necessary for identifiability is given by

$$K \geq \max\left(\left\lceil \frac{N_1 N_2}{N_d N_s} \right\rceil, \left\lceil \frac{N_1}{N_d} \right\rceil, \left\lceil \frac{N_2}{N_s} \right\rceil, 2\right), \quad (28)$$

where $\lceil x \rceil$ is equal to the smallest integer that is greater than or equal to x.

Note that the identifiability of the channel matrices $\mathbf{H}_{sr}$, $\mathbf{H}_{rr}$, and $\mathbf{H}_{rd}$ from the unstructured channel tensor $\mathcal{H}$ will ensure that the compound channel $\mathbf{H}_c = \mathbf{H}_{rd}\mathbf{H}_{rr}\mathbf{H}_{sr} \in \mathbb{C}^{N_d \times N_s}$ is strictly unique. Note also that conditions $N_d N_s K \geq N_1 N_2$ and $N_d K \geq N_1$ are clearly much less restrictive in terms of the required number N_d of antennas at the destination node, in comparison with

the conventional three-step LS estimator that requires $N_d \geq N_2 \geq N_1$. Otherwise stated, estimation of the partial channels can be done even in situations where the number of receive antennas is much less than the number of relay antennas (provided that K satisfies condition (28)). This situation may arise in scenarios with denser deployments of relay stations, where the total number of relay antennas exceeds those of source and/or destination antennas. As shown by these inequalities, the possibility of affording fewer receive antennas is compensated by an increase on the number K of training blocks, which represents a trade-off.

Condition (28), although necessary, is not sufficient for identifiability. Since $\mathbf{Z}_{[2]} \doteq (\mathbf{I}_K \otimes \mathbf{H}_{rd})\,\mathbf{\Omega}_2 \in \mathbb{C}^{N_d K \times N_1}$ and $\mathbf{Z}_{[3]} \doteq \left(\mathbf{I}_K \otimes \mathbf{H}_{sr}^T\right)\mathbf{\Omega}_3 \in \mathbb{C}^{N_s K \times N_2}$, additionally, must have rank$(\mathbf{\Omega}_2) = N_1$ and rank$(\mathbf{\Omega}_3) = N_2$, i.e., both $\mathbf{\Omega}_2 \in \mathbb{C}^{N_2 K \times N_1}$ and $\mathbf{\Omega}_3 \in \mathbb{C}^{N_1 k \times N_2}$ must be full column-rank. Otherwise, $\mathbf{Z}_{[2]}$ and $\mathbf{Z}_{[3]}$ will be rank-deficient, even if (28) is respected.

Let us assume that the partial channels $\mathbf{H}_{sr}$, $\mathbf{H}_{rr}$, and $\mathbf{H}_{rd}$ are full rank matrices, which is a reasonable assumption when the wireless links are assumed to undergo scattering-rich multipath propagation. The following corollaries can then be obtained:

C1 If $N_1 = N_2$, identifiability of the partial channels is guaranteed for $N_1 \leq N_s$ and $N_2 \leq N_d$;
C2 If $N_1 = 1$, identifiability of the partial channels is guaranteed for $N_2 \leq N_d$ and $N_2 \leq K$;
C3 If $N_2 = 1$, identifiability of the partial channels is guaranteed for $N_1 \leq N_s$ and $N_1 \leq K$.

Remark: For the first corollary, we can note that if $N_1 \leq N_s$ and $N_2 \leq N_d$, then $\mathbf{H}_{sr}^T \otimes \mathbf{H}_{rd}$ is full column-rank, which ensures that $\mathbf{\Omega}_1 \in \mathbb{C}^{N_D N_s K \times N_2}$ is full column-rank due to its Khatri-Rao product structure [24]. Likewise, $\mathbf{\Omega}_2 \in \mathbb{C}^{N_2 K \times N_1}$ and $\mathbf{\Omega}_3 \in \mathbb{C}^{N_1 k \times N_2}$ are also full column-rank in this case, guaranteeing the identifiability of the channel matrices. Regarding the second corollary, it corresponds to a special case of our system model where the first relay tier reduces to a single-antenna relay. In this case, satisfying $N_2 \leq N_d$ and $N_2 \leq K$ ensures that $\mathbf{\Omega}_1$, $\mathbf{Z}_{[2]}$, and $\mathbf{Z}_{[3]}$ are all full column-rank, so that the three partial channels are identifiable. The same reasoning is valid for the third corollary, which is analogous to the second one.

3.2 Essential uniqueness

Let $\left\{\widehat{\mathbf{H}}_{sr}, \widehat{\mathbf{H}}_{rr}, \widehat{\mathbf{H}}_{rd}\right\}$ be an alternative set of matrices yielding the same unstructured channel tensor $\mathcal{H}$ satisfying the PARATUCK2 model (14). If $\mathbf{H}_{sr}$, $\mathbf{H}_{rr}$, and $\mathbf{H}_{rd}$ are full rank and the identifiability conditions (27) are satisfied, then $\widehat{\mathbf{H}}_{sr}$, $\widehat{\mathbf{H}}_{rr}$, and $\widehat{\mathbf{H}}_{rd}$ are essentially unique. In this case, we have $\widehat{\mathbf{H}}_{sr} = \mathbf{\Delta}_{sr}\mathbf{H}_{sr}$, $\widehat{\mathbf{H}}_{rd} = \mathbf{H}_{rd}\mathbf{\Delta}_{rd}$

and $\widehat{\mathbf{H}}_{rr} = \mathbf{\Delta}_{rr}^{(2)} \mathbf{H}_{rr} \mathbf{\Delta}_{rr}^{(1)}$, where the following relation holds:

$$\left(\mathbf{\Delta}_{sr} \mathbf{\Delta}_{rr}^{(1)}\right) \otimes \left(\mathbf{\Delta}_{rd} \mathbf{\Delta}_{rr}^{(2)}\right) = \mathbf{I}_{N_1 N_2}. \qquad (29)$$

Note that permutation ambiguity does not exist due to the knowledge of the training matrices $\mathbf{E}$ and $\mathbf{F}$. The relation (29) can be obtained by replacing the alternative solutions $\widehat{\mathbf{H}}_{sr}$, $\widehat{\mathbf{H}}_{rd}$, and $\widehat{\mathbf{H}}_{rr}$ into (14) and then applying some basic manipulations using properties of the Kronecker product. Equation (29) turns into the following relations: $\mathbf{\Delta}_{rd} \mathbf{\Delta}_{rr}^{(2)} = \alpha \mathbf{I}_{N_2}$ and $\mathbf{\Delta}_{sr} \mathbf{\Delta}_{rr}^{(1)} = (1/\alpha) \mathbf{I}_{N_1}$, where α is an arbitrary scalar factor. These two relations come from the fact that the Kronecker product between any two diagonal matrices is equal to the identity matrix if and only if these diagonal matrices are (scaled) identity matrices that compensate each other. Consequently, $\mathbf{H}_{sr}$, $\mathbf{H}_{rr}$, and $\mathbf{H}_{rd}$ can be recovered in an essentially unique manner up to scaling factors. The scaling ambiguity can be eliminated by normalizing the first column of $\mathbf{H}_{sr}$ or the first row of $\mathbf{H}_{rd}$ to one. Since these ambiguities compensate each other, the compound channel is strictly unique and we have $\widehat{\mathbf{H}}_c = \widehat{\mathbf{H}}_{rd} \widehat{\mathbf{H}}_{rr} \widehat{\mathbf{H}}_{sr} = \mathbf{H}_{rd} \mathbf{H}_{rr} \mathbf{H}_{sr} = \mathbf{H}_c$.

3.3 Trilinear alternating least squares algorithm

The TALS algorithm is an iterative estimation method that alternates among the LS estimations of the channel matrices $\mathbf{H}_{sr}$, $\mathbf{H}_{rr}$, and $\mathbf{H}_{rd}$ by fitting a PARATUCK2 model from the noisy matrices $\widetilde{\mathbf{H}}_{[i]} = \mathbf{H}_{[i]} + \mathbf{V}_{[i]}$, $i = 1, 2, 3$. Note that the noise term $\mathbf{V}_{[i]}$ is constructed in a way analogous to $\mathbf{H}_{[i]}$, $i = 1, 2, 3$, following Equations (15) and (21), respectively. The AF training matrices $\mathbf{E}$ and $\mathbf{F}$ are assumed to be known at the destination and are fixed during the estimation process. From (19), (24), and (25), we respectively obtain the following linear optimization problems:

$$\underset{\text{vec}(\mathbf{H}_{rr})}{\arg\min} \left\| \text{vec}\left(\widetilde{\mathbf{H}}_{[1]}\right) - \mathbf{\Omega}_1 \text{vec}(\mathbf{H}_{rr}) \right\|_F^2, \qquad (30)$$

$$\underset{\mathbf{H}_{sr}}{\arg\min} \left\| \widetilde{\mathbf{H}}_{[2]} - (\mathbf{I}_K \otimes \mathbf{H}_{rd}) \mathbf{\Omega}_2 \mathbf{H}_{sr} \right\|_F^2, \qquad (31)$$

$$\underset{\mathbf{H}_{rd}}{\arg\min} \left\| \widetilde{\mathbf{H}}_{[3]} - \left(\mathbf{I}_K \otimes \mathbf{H}_{sr}^T\right) \mathbf{\Omega}_3 \mathbf{H}_{rd}^T \right\|_F^2. \qquad (32)$$

These LS estimation problems can be solved alternately by estimating one channel matrix at each time, while fixing the other matrices to their values obtained in previous estimation steps. Therefore, each iteration of the algorithm has three estimation steps. The algorithm starts by randomly initializing two out of the three channel matrices and proceeds until convergence. In the following, a summary of the TALS algorithm is provided.

TALS algorithm (direct estimation of $\mathbf{H}_{sr}$, $\mathbf{H}_{rr}$, and $\mathbf{H}_{rd}$)

1. *Set $n = 0$;*
   Initialize randomly $\widehat{\mathbf{H}}_{rr}$ ($n = 0$) and $\widehat{\mathbf{H}}_{rd}$ ($n = 0$);
   Using (16) and (26), construct $\widehat{\mathbf{\Omega}}_2(n = 0)$ and $\widehat{\mathbf{\Omega}}_3(n = 0)$, respectively;

2. *$n \leftarrow n + 1$;*

3. *Find an estimate of $\mathbf{H}_{sr}$ using (22), by solving the LS problem*

   $$\underset{\mathbf{H}_{sr}}{\arg\min} \left\| \widetilde{\mathbf{H}}_{[2]} - \left(\mathbf{I}_K \otimes \mathbf{H}_{rd}\right) \mathbf{\Omega}_2 \mathbf{H}_{sr} \right\|_F^2$$

   the solution of which at the nth iteration is given by

   $$\widehat{\mathbf{H}}_{sr}(n) = \left[\left(\mathbf{I}_K \otimes \widehat{\mathbf{H}}_{rd}(n-1)\right) \widehat{\mathbf{\Omega}}_2(n-1) \right]^\dagger \widetilde{\mathbf{H}}_{[2]}$$

4. *Find an estimate of $\mathbf{H}_{rd}$ using (23), by solving the LS problem*

   $$\underset{\mathbf{H}_{rd}}{\arg\min} \left\| \widetilde{\mathbf{H}}_{[3]} - \left(\mathbf{I}_K \otimes \mathbf{H}_{sr}^T\right) \mathbf{\Omega}_3 \mathbf{H}_{rd}^T \right\|_F^2$$

   the solution of which at the nth iteration is given by

   $$\widehat{\mathbf{H}}_{rd}^T(n) = \left[\left(\mathbf{I}_K \otimes \widehat{\mathbf{H}}_{sr}^T(n)\right) \widehat{\mathbf{\Omega}}_3(n-1) \right]^\dagger \widetilde{\mathbf{H}}_{[3]}$$

5. *Find an estimate of $\mathbf{h}_{rr} \doteq vec(\mathbf{H}_{rr})$ using (17), by solving the LS problem*

   $$\underset{\mathbf{h}_{rr}}{\arg\min} \left\| \text{vec}\left(\widetilde{\mathbf{H}}_{[1]}\right) - \left[\left(\mathbf{E}^T \odot \mathbf{F}^T\right)^T \odot \left(\mathbf{H}_{sr}^T \otimes \mathbf{H}_{rd}\right) \right] \right.$$
   $$\left. \times \mathbf{h}_{rr} \right\|_F^2$$

   the solution of which at the nth iteration is given by

   $$\widehat{\mathbf{h}}_{rr}(n) = \left[\left(\mathbf{E}^T \odot \mathbf{F}^T\right)^T \odot \left(\widehat{\mathbf{H}}_{sr}^T(n) \otimes \widehat{\mathbf{H}}_{rd}(n)\right) \right]^\dagger$$
   $$\times \text{vec}\left(\widetilde{\mathbf{H}}_{[1]}\right)$$

6. *Rebuild $\widehat{\mathbf{\Omega}}_2(n)$ and $\widehat{\mathbf{\Omega}}_3(n)$ and repeat Steps 2 to 5 until convergence.*

Define $\mathbf{e}(n) = \text{vec}\left(\widetilde{\mathbf{H}}_{[1]}\right) - \left[\left(\mathbf{E}^T \odot \mathbf{F}^T\right)^T \odot \left(\widehat{\mathbf{H}}_{sr}^T(n) \otimes \widehat{\mathbf{H}}_{rd}(n)\right) \right] \widehat{\mathbf{h}}_{rr}(n)$. The sum of squared residuals (SSR) at the end of the nth iteration is defined as $\text{SSR}(n) = \mathbf{e}^H(n) \mathbf{e}(n)$. We declare the convergence of the algorithm when $|\text{SSR}(n) - \text{SSR}(n-1)| \leq 10^{-6}$, meaning that the model reconstruction error does not significantly change between two successive iterations.

Generally, the ALS algorithm is sensitive to the initialization, and convergence to the global minimum can be slow when all the matrix factors of the model are unknown [25]. However, in our case, we have observed that convergence to the global minimum is always achieved (e.g., within 10 to 30 iterations for medium-to-high SNRs)

due to the knowledge of the AF training matrices $\mathbf{E}$ and $\mathbf{F}$.

3.4 Kronecker least squares algorithm

We now derive a closed-form solution to our channel estimation problem by exploiting the mixed Kronecker/Khatri-Rao factorization structure of the matrix unfolding $\mathbf{H}_{[1]}$ defined in (17). Starting from (13), the noisy version of (15) is given by:

$$\widehat{\mathbf{H}}_{[1]} = \mathbf{H}_{[1]} + \left(\mathbf{S}_0^H \otimes \mathbf{I}_{N_d}\right)\mathbf{V}_{[1]}, \qquad (33)$$

where $\mathbf{V}_{[1]} = [\text{vec}(\mathbf{V}_1), \ldots, \text{vec}(\mathbf{V}_K)] \in \mathbb{C}^{N_d N_s \times K}$. Let $\mathbf{Z} = \mathbf{E}^T \odot \mathbf{F}^T \in \mathbb{C}^{N_1 N_2 \times K}$ denote the combined AF training matrix and assume that $\mathbf{Z}\mathbf{Z}^H = \mathbf{I}_{N_1 N_2}$. Multiplying both sides of (33) by $\mathbf{Z}^H$, we have:

$$\widehat{\mathcal{X}}_{[1]} \doteq \widehat{\mathbf{H}}_{[1]}\mathbf{Z}^H + \left(\mathbf{S}_0^H \otimes \mathbf{I}_{N_d}\right)\mathbf{V}_{[1]}\mathbf{Z}^H \qquad (34)$$

where $\widehat{\mathcal{X}}_{[1]} = \mathbf{X}_{[1]} + \left(\mathbf{S}_0^H \otimes \mathbf{I}_{N_d}\right)\mathbf{V}_{[1]}\mathbf{Z}^H$. From (17), we have:

$$\mathbf{X}_{[1]} = \left(\mathbf{H}_{\text{sr}}^T \otimes \mathbf{H}_{\text{rd}}\right)\text{diag}\left(\text{vec}\left(\mathbf{H}_{\text{rr}}\right)\right). \qquad (35)$$

Our goal is to directly identify the channel matrices from (35). However, let us first address the deterministic design of the AF training matrices $\mathbf{E}$ and $\mathbf{F}$ such that $\mathbf{Z}\mathbf{Z}^H \doteq \left(\mathbf{E}^T \odot \mathbf{F}^T\right)\left(\mathbf{E}^T \odot \mathbf{F}^T\right)^H = \mathbf{I}_{N_1 N_2}$. Assuming $K \geq N_1 N_2$, this condition is satisfied by designing $\mathbf{Z}$, for instance, as a discrete Fourier transform (DFT) matrix. Having fixed the structure of $\mathbf{Z}$, we are left with the problem of factorizing this matrix as the Khatri-Rao product between $\mathbf{E}^T$ and $\mathbf{F}^T$. This problem can easily be solved by means of K rank-one matrix factorizations, which admit unique solutions. Note that the kth column of $\mathbf{Z}$ can be written as

$$\mathbf{Z}(:,k) = (\mathbf{E}(k,:) \otimes \mathbf{F}(k,:))^T \in \mathbb{C}^{N_1 N_2 \times 1}, \; k = 1, \ldots, K.$$

Defining a rank-one matrix $\tilde{\mathbf{Z}}_k \doteq \text{unvec}(\mathbf{Z}(:,k)) \in \mathbb{C}^{N_2 \times N_1}$, it follows that

$$\tilde{\mathbf{Z}}_k = (\mathbf{F}(k,:))^T \mathbf{E}(k,:),$$

from which $\mathbf{E}(k,:)$ and $\mathbf{F}(k,:)$ can be determined as the unique right and left singular vectors of $\tilde{\mathbf{Z}}_k$, $k = 1, \ldots, K$. Note that the proposed design, although not optimized to minimize the mean square error of the channel estimation, ensures that the noise characteristics in (33) will not be changed when $\widehat{\mathbf{H}}_{[1]}$ is post-multiplied by $\mathbf{Z}^H$ (i.e., inverse DFT transformation).

Coming back to the channel estimation problem, from (35), let us define $\mathbf{x}_{n_1,n_2} \in \mathbb{C}^{N_s N_d \times 1}$ as the $[(n_1 - 1)N_2 + n_2]$-th column of $\mathbf{X}_{[1]} \in \mathbb{C}^{N_s N_d \times N_1 N_2}$, $n_1 = 1, \ldots, N_1$, $n_2 = 1, \ldots, N_2$. Note that

$$\mathbf{x}_{n_1,n_2} = \left(\mathbf{H}_{\text{sr}}^T(:,n_1) \otimes \mathbf{H}_{\text{rd}}(:,n_2)\right) h_{\text{rr}}(n_2, n_1) \qquad (36)$$

Defining $\tilde{\mathbf{X}}_{n_1,n_2} \doteq \text{unvec}(\mathbf{x}_{n_1,n_2}) \in \mathbb{C}^{N_d \times N_s}$ as a rank-one matrix obtaining by reshaping, we have

$$\tilde{\mathbf{X}}_{n_1,n_2} = h_{\text{rr}}(n_2, n_1)\,\mathbf{H}_{\text{rd}}(:,n_2)\mathbf{H}_{\text{sr}}(n_1,:) \qquad (37)$$

Consider the singular value decomposition (SVD) of $\tilde{\mathbf{X}}_{n_1,n_2}$:

$$\tilde{\mathbf{X}}_{n_1,n_2} = \mathbf{U}_{n_1,n_2}\mathbf{\Lambda}_{n_1,n_2}\mathbf{V}_{n_1,n_2}^H \qquad (38)$$

$$n_1 = 1, \ldots, N_1, \; n_2 = 1, \ldots, N_2. \qquad (39)$$

From the rank-one property of $\tilde{\mathbf{X}}_{n_1,n_2}$, we have:

$$\widehat{\mathbf{H}}_{\text{rd}}^{(n_2)}(:,n_2) = \mathbf{U}_{n_1,n_2}(:,1), \; n_1 = 1, \ldots, N_1, \qquad (40)$$

$$\widehat{\mathbf{H}}_{\text{sr}}^{(n_1)}(n_1,:) = (\mathbf{V}_{n_1,n_2}(:,1))^T, \; n_2 = 1, \ldots, N_2, \qquad (41)$$

$$\widehat{h}_{\text{rr}}(n_2, n_1) = \lambda_{n_1,n_2}(1,1). \qquad (42)$$

Final estimates of $\mathbf{H}_{\text{rd}}(:,n_2)$ and $\mathbf{H}_{\text{sr}}(:,n_1)$ can be obtained by averaging over the N_1 and N_2 independent estimates, respectively:

$$\widehat{\mathbf{H}}_{\text{rd}}(:,n_2) = \frac{1}{N_1}\sum_{n_1=1}^{N_1}\widehat{\mathbf{H}}_{\text{rd}}^{(n_1)}(:,n_2), \qquad (43)$$

$$\widehat{\mathbf{H}}_{\text{sr}}(n_1,:) = \frac{1}{N_2}\sum_{n_2=1}^{N_2}\widehat{\mathbf{H}}_{\text{sr}}^{(n_2)}(n_1,:), \qquad (44)$$

with

$$\widehat{\mathbf{H}}_{\text{rd}} = \left[\widehat{\mathbf{H}}_{\text{rd}}(:,1), \ldots, \widehat{\mathbf{H}}_{\text{rd}}(:,N_2)\right], \qquad (45)$$

$$\widehat{\mathbf{H}}_{\text{sr}} = \left[\widehat{\mathbf{H}}_{\text{sr}}(:,1), \ldots, \widehat{\mathbf{H}}_{\text{sr}}(:,N_1)\right]^T, \qquad (46)$$

$$\widehat{\mathbf{H}}_{\text{rr}} = \begin{bmatrix} \lambda_{1,1}(1,1) & \cdots & \lambda_{N_1,1}(1,1) \\ \vdots & \vdots & \vdots \\ \lambda_{1,N_2}(1,1) & \cdots & \lambda_{N_1,N_2}(1,1) \end{bmatrix}. \qquad (47)$$

Note that the columns of the estimated $\widehat{\mathbf{H}}_{\text{sr}}$ and $\widehat{\mathbf{H}}_{\text{rd}}$ have unit energy while each entry of $\widehat{\mathbf{H}}_{\text{rr}}$ concentrates all the energy of the wireless link connecting the source node to the destination node via a given tier 1-tier 2 relay pair. Such an interpretation is useful for designing transmit and receive spatial filters for system optimization as well as for power allocation purposes.

Discussion: The KRLS algorithm involves the computation of $N_1 N_2$ SVDs to provide rank-one approximations for the matrices $\widehat{\mathbf{X}}_{1,1}, \ldots, \widehat{\mathbf{X}}_{N_1,N_2}$, of dimensions $N_d \times N_s$, which are constructed from the $N_1 N_2$ columns $\widehat{\mathbf{X}}_{[1]}$. The distinguishing feature of the KRLS-based estimator is on the closed-form solution to the problem, as opposed to the TALS algorithm that consists of iterative LS estimation steps, which implies a higher computational complexity. However, note that the KRLS algorithm is only applicable under the condition $K \geq N_1 N_2$, which is necessary for $\mathbf{Z} = \mathbf{E}^T \odot \mathbf{F}^T$ to have orthogonal rows, leading to (35). In contrast, the TALS algorithm can operate under a

much lower bound on K, as discussed in Section 3.1. This is clearly a trade-off between both estimators in terms of identifiability conditions and computational complexity. As will be shown in the next section, both estimators provide satisfactory performances, and the choice of the best estimator is rather dependent on the design constraints of the system. For instance, we can say that the TALS estimator is preferable if processing power at the receiver is not too limited, as is often the case with base station reception in outdoor micro- or macro-cells. The KRLS solution would be more likely chosen in indoor scenarios, where channel coherence time is long enough to allow for higher values of K.

4 Extension to two-way MIMO relaying systems

In the previous sections, we have focused on a multi-relay cooperative scheme, where transmission is directed in one direction, i.e., from a specific source to a specific destination via two tiers of multiple relays. In this section, we show that the same modeling approach can be extended to a two-way MIMO relaying scenario, where pilot/data transmission takes place in both directions. In the first phase, two sources simultaneously transmit their data to the multiple relays. Note that, in the two-way case, the relays of each tier receive a superposition of $N_{s_1} + N_{s_2}$ signals coming from sources 1 and 2. In the second and third phases, inter-relay communication takes place. More specifically, in phase two, tier 1 relays transmit signals towards tier 2 relays, while tier 1 relays stay silent. In phase three, the opposite happens. Finally, in the fourth communication phase, all the relays transmit to the two sources, and each one of them receives a superposition of $N_1 + N_2$ signals.

In the first transmission phase, we assume that training symbol matrices $\mathbf{S}_1 \in \mathbb{C}^{N_{s_1} \times L}$ and $\mathbf{S}_2 \in \mathbb{C}^{N_{s_2} \times L}$ are transmitted from sources 1 and 2, respectively. We omit the additive noise terms for convenience of presentation. The signal received at the ith relay tier is given by:

$$\mathbf{X}^{(i)} = \mathbf{H}_{s_1 r_i} \mathbf{S}_1 + \mathbf{H}_{s_2 r_i} \mathbf{S}_2 = \mathbf{H}^{(i)} \mathbf{S}, \quad i = 1, 2, \quad (48)$$

where $\mathbf{H}^{(i)} \doteq \left[\mathbf{H}_{s_1 r_i} \, \mathbf{H}_{s_2 r_i}\right] \in \mathbb{C}^{N_i \times (N_{s_1} + N_{s_2})}$, and $\mathbf{S} \doteq [\mathbf{S}_1^T \, \mathbf{S}_2^T]^T \in \mathbb{C}^{(N_{s_1} + N_{s_2}) \times L}$. The training sequence $\mathbf{S}_i$ chosen by source i, is designed to satisfy the following conditions:

(i) $\mathbf{S}_i \mathbf{S}_i^H = \mathbf{I}_{N_i}, i = 1, 2,$
(ii) $\mathbf{S}_1 \mathbf{S}_2^H = \mathbf{0}_{N_1 \times N_2}.$

A possible construction satisfying these two conditions is based on the normalized DFT matrix of size $L \times (N_{s_1} + N_{s_2})$, with $L \geq N_{s_1} + N_{s_2}$. This design allows the sources to eliminate the self-interference generated by their own transmission, when receiving the signal back from the relays.

In the second and third phases, where inter-relay communications happen, the signal received at the relays of tier i from the relays of tier j, $(i,j) = \{(1,2),(2,1)\}$, can be written as:

$$\begin{aligned}
\mathbf{Z}_k^{(i)} &= \mathbf{H}_{r_j r_i} D_k(\mathbf{E}_j) \mathbf{X}^{(j)} \\
&= \mathbf{H}_{r_j r_i} D_k(\mathbf{E}_j) \mathbf{H}^{(j)} \mathbf{S}, \quad (49)
\end{aligned}$$

$k = 1, \ldots, K$, where $\mathbf{H}_{r_j r_i} \in \mathbb{C}^{N_i \times N_j}$ is the MIMO channel linking the relays of tier j at transmission to the relays of tier i at reception, $(i,j) = \{(1,2),(2,1)\}$. Note that channel reciprocity in the inter-relay communications is not a necessary assumption which means that we may have $\mathbf{H}_{r_1 r_2} \neq \mathbf{H}_{r_2 r_1}$.

Finally, in the fourth transmission phase, the signals received at sources 1 and 2 are post-multiplied by $\mathbf{S}_2^H$ and $\mathbf{S}_1^H$, respectively, to accomplish self-interference elimination, yielding

$$\begin{aligned}
\mathbf{Y}_k^{(1)} &= \left(\mathbf{H}_{r_1 s_1} D_k(\mathbf{F}_1) \mathbf{Z}_k^{(1)}\right) \mathbf{S}_2^H + \left(\mathbf{H}_{r_2 s_1} D_k(\mathbf{F}_2) \mathbf{Z}_k^{(2)}\right) \mathbf{S}_2^H \\
&= \underbrace{\mathbf{H}_{r_1 s_1} D_k(\mathbf{F}_1) \mathbf{H}_{r_2 r_1} D_k(\mathbf{E}_2) \mathbf{H}_{s_2 r_2}}_{\text{tier 2} \to \text{tier 1 relay path}} \\
&\quad + \underbrace{\mathbf{H}_{r_2 s_1} D_k(\mathbf{F}_2) \mathbf{H}_{r_1 r_2} D_k(\mathbf{E}_1) \mathbf{H}_{s_2 r_1}}_{\text{tier 1} \to \text{tier 2 relay path}} \\
&= \bar{\mathbf{H}}^{(1,1)} D_k(\bar{\mathbf{F}}_{1,2}) \mathbf{G}_{rr}^{(1)} D_k(\bar{\mathbf{E}}_{2,1}) \bar{\mathbf{H}}^{(1,2)}, \quad k = 1, \ldots, K,
\end{aligned}$$
$$(50)$$

and

$$\begin{aligned}
\mathbf{Y}_k^{(2)} &= \left(\mathbf{H}_{r_1 s_2} D_k(\mathbf{F}_1) \mathbf{Z}_k^{(1)}\right) \mathbf{S}_1^H + \left(\mathbf{H}_{r_2 s_2} D_k(\mathbf{F}_2) \mathbf{Z}_k^{(2)}\right) \mathbf{S}_1^H \\
&= \underbrace{\mathbf{H}_{r_2 s_2} D_k(\mathbf{F}_2) \mathbf{H}_{r_1 r_2} D_k(\mathbf{E}_1) \mathbf{H}_{s_1 r_1}}_{\text{tier 1} \to \text{tier 2 relay path}} \\
&\quad + \underbrace{\mathbf{H}_{r_1 s_2} D_k(\mathbf{F}_1) \mathbf{H}_{r_2 r_1} D_k(\mathbf{E}_2) \mathbf{H}_{s_1 r_2}}_{\text{tier 2} \to \text{tier 1 relay path}} \\
&= \bar{\mathbf{H}}^{(2,1)} D_k(\bar{\mathbf{F}}_{2,1}) \mathbf{G}_{rr}^{(2)} D_k(\bar{\mathbf{E}}_{1,2}) \bar{\mathbf{H}}^{(2,2)}, \quad k = 1, \ldots, K,
\end{aligned}$$
$$(51)$$

where

$$\bar{\mathbf{H}}^{(1,1)} \doteq [\mathbf{H}_{r_1 s_1} \, \mathbf{H}_{r_2 s_1}] \in \mathbb{C}^{N_{s_1} \times (N_1 + N_2)} \quad (52)$$

$$\bar{\mathbf{H}}^{(1,2)} \doteq [\mathbf{H}_{s_2 r_2}^T \, \mathbf{H}_{s_2 r_1}^T]^T \in \mathbb{C}^{(N_1 + N_2) \times N_{s_2}} \quad (53)$$

$$\bar{\mathbf{H}}^{(2,1)} \doteq [\mathbf{H}_{r_2 s_2} \, \mathbf{H}_{r_1 s_2}] \in \mathbb{C}^{N_{s_2} \times (N_1 + N_2)} \quad (54)$$

$$\bar{\mathbf{H}}^{(2,2)} \doteq [\mathbf{H}_{s_1 r_1}^T \, \mathbf{H}_{s_1 r_2}^T]^T \in \mathbb{C}^{(N_1 + N_2) \times N_{s_1}} \quad (55)$$

$$\mathbf{G}_{rr}^{(1)} \doteq \text{blockdiag}\left(\mathbf{H}_{r_2 r_1} \, \mathbf{H}_{r_1 r_2}\right) \in \mathbb{C}^{(N_1 + N_2) \times (N_1 + N_2)} \quad (56)$$

$$\mathbf{G}_{rr}^{(2)} \doteq \text{blockdiag}\left(\mathbf{H}_{r_1 r_2} \, \mathbf{H}_{r_2 r_1}\right) \in \mathbb{C}^{(N_1 + N_2) \times (N_1 + N_2)} \quad (57)$$

$$\bar{\mathbf{F}}_{i,j} \doteq [\mathbf{F}_i \, \mathbf{F}_j], \quad \bar{\mathbf{E}}_{i,j} \doteq [\mathbf{E}_i \, \mathbf{E}_j], \quad (i,j) = \{(1,2),(2,1)\}. \quad (58)$$

Therefore, we can conclude that the signals received at sources 1 and 2 in the considered two-way MIMO relaying scenario (Equations (50) and (51)) follows a PARATUCK2 model. By analogy with the noiseless part of the one-way

signal model (14), we have the following correspondences between the factor matrices:

$$\left(\mathbf{H}_{\mathrm{rd}}, \mathbf{H}_{\mathrm{sr}}, \mathbf{H}_{\mathrm{rr}}\right) \longleftrightarrow \left(\bar{\mathbf{H}}^{(1,1)}, \bar{\mathbf{H}}^{(1,2)}, \mathbf{G}_{\mathrm{rr}}^{(1)}\right)$$

$$(\mathbf{E}, \mathbf{F}) \longleftrightarrow \left(\bar{\mathbf{F}}_{1,2}, \bar{\mathbf{E}}_{2,1}\right) \qquad \text{(source 1)} \qquad (59)$$

$$\left(\mathbf{H}_{\mathrm{rd}}, \mathbf{H}_{\mathrm{sr}}, \mathbf{H}_{\mathrm{rr}}\right) \longleftrightarrow \left(\bar{\mathbf{H}}^{(2,1)}, \bar{\mathbf{H}}^{(2,2)}, \mathbf{G}_{\mathrm{rr}}^{(2)}\right)$$

$$(\mathbf{E}, \mathbf{F}) \longleftrightarrow \left(\bar{\mathbf{F}}_{2,1}, \bar{\mathbf{E}}_{1,2}\right) \qquad \text{(source 2)} \qquad (60)$$

Consequently, the tensor-based channel estimation algorithms proposed in the previous section can be equally applied at each source to estimate the channels $\bar{\mathbf{H}}^{(i,1)}$, $\bar{\mathbf{H}}^{(i,2)}$ and $\mathbf{G}_{\mathrm{rr}}^{(i)}$, $i = 1, 2$, from Equations (50) and (51), respectively. If reciprocity is assumed in the two-way relay channels, we have:

$$\mathbf{H}_{\mathrm{r}_i\mathrm{s}_i} = \mathbf{H}_{\mathrm{s}_i\mathrm{r}_i}^T, \quad i = 1, 2 \qquad (61)$$

$$\mathbf{H}_{\mathrm{r}_i\mathrm{s}_j} = \mathbf{H}_{\mathrm{s}_j\mathrm{r}_i}^T, \quad (i, j) = \{(1, 2), (2, 1)\}, \qquad (62)$$

$$\mathbf{H}_{\mathrm{r}_i\mathrm{r}_j} = \mathbf{H}_{\mathrm{r}_j\mathrm{r}_i}^T, \quad (i, j) = \{(1, 2), (2, 1)\}, \qquad (63)$$

which in turn implies $\bar{\mathbf{H}}^{(1,1)} = \left(\bar{\mathbf{H}}^{(2,2)}\right)^T = \mathbf{H}_{s_1}$, $\bar{\mathbf{H}}^{(1,2)} = \left(\bar{\mathbf{H}}^{(2,1)}\right)^T = \mathbf{H}_{s_2}$, and $\mathbf{G}_{\mathrm{rr}}^{(1)} = \left(\mathbf{G}_{\mathrm{rr}}^{(2)}\right)^T = \mathbf{G}$. In this particular case, the PARATUCK2 models (50) and (51) become essentially equal, i.e., they depend on the same unknown channel matrices $\mathbf{H}_{s_1}$, $\mathbf{H}_{s_2}$, and $\mathbf{G}$ to be estimated. Note, however, that such a reciprocity is not a necessary assumption of our modeling approach, which can be used in the general case of non-symmetrical two-way MIMO relay channels.

5 Numerical results

We now present computer simulation results for assessing the performance of the proposed channel estimator in selected system configurations. The estimator's performance is evaluated in terms of the normalized mean square error (NMSE) of the estimated channel matrices. From the estimated channels, the performance in terms of bit error rate (BER) is calculated by assuming a linear receive filter. The BER and NMSE curves are plotted as a function of the overall signal-to-noise ratio (SNR) at the destination. This SNR is given by the ratio between the powers of the useful signal component and the noise component in Equation (11). For each simulated SNR value, the results represent an average over $L = 5,000$ Monte Carlo runs. At each run, the channel coefficients are drawn from a circularly symmetric complex-valued Gaussian distribution with zero-mean and unit variance, while the transmitted symbols are drawn from a BPSK sequence. The SNR level at the tier 1 and tier 2 relays are assumed to be 30 dB above the SNR level at the destination.

For purposes of performance evaluation, the scaling ambiguities affecting the estimates of the channel matrices are removed by assuming the first column of $\mathbf{H}_{\mathrm{sr}}$ and first row of $\mathbf{H}_{\mathrm{rd}}$ contain all one elements, similarly to [11,14]. These scaling ambiguities can be determined as follows. First, we find $\boldsymbol{\Delta}_{\mathrm{sr}} = \mathrm{D}_1\left(\widehat{\mathbf{H}}_{\mathrm{sr}}^T\right)$ and $\boldsymbol{\Delta}_{\mathrm{rd}} = \mathrm{D}_1\left(\widehat{\mathbf{H}}_{\mathrm{rd}}\right)$. Then, applying property $(\mathbf{AB}) \otimes (\mathbf{CD}) = (\mathbf{A} \otimes \mathbf{C})(\mathbf{B} \otimes \mathbf{D})$ yields $\left(\boldsymbol{\Delta}_{\mathrm{sr}} \otimes \boldsymbol{\Delta}_{\mathrm{rd}}\right)\left(\boldsymbol{\Delta}_{\mathrm{rr}}^{(1)} \otimes \boldsymbol{\Delta}_{\mathrm{rr}}^{(2)}\right) = \mathbf{I}_{N_1 N_2}$, from which we obtain $\boldsymbol{\Delta}_{\mathrm{rr}}^{(1)} \otimes \boldsymbol{\Delta}_{\mathrm{rr}}^{(2)} = \boldsymbol{\Delta}_{\mathrm{sr}}^{-1} \otimes \boldsymbol{\Delta}_{\mathrm{rd}}^{-1}$. A solution to this relation is then found as $\boldsymbol{\Delta}_{\mathrm{rr}}^{(1)} = \left[\mathrm{D}_1\left(\widehat{\mathbf{H}}_{\mathrm{sr}}^T\right)\right]^{-1}$ and $\boldsymbol{\Delta}_{\mathrm{rr}}^{(2)} = \left[\mathrm{D}_1\left(\widehat{\mathbf{H}}_{\mathrm{rd}}\right)\right]^{-1}$.

In Figure 2, we depict the NMSE performance for the compound channel of our proposed estimators in comparison with the conventional LS estimator. The parameters are $N_s = 2$, $N_1 = 4$, $N_2 = 4$, $N_d = 6$, $K = 16$, $L_0 = 30$, and the number of transmitted data symbols is $N = 1000$. We can see that TALS and KRLS have similar performances, which are considerably better than the conventional (three-stage) LS estimator. The worst performance of the LS estimator comes from the error accumulation across successive channel estimation stages, which degrades its overall NMSE performance.

Figure 3 shows the NMSE performance of our proposed estimators in comparison with the two-hop bilinear alternating least squares (BALS) estimator of Rong et al. [14]. This estimator is a special case of the proposed one, where only one tier of relays is used. In this case, model (11) reduces to

$$\mathbf{Y}_k = \mathbf{H}_{\mathrm{rd}} \mathrm{D}_k(\mathbf{E})\, \mathbf{H}_{\mathrm{sr}} \mathbf{S}_0 + \mathbf{V}_k, \qquad (64)$$
$$k = 1, \ldots, K,$$

and the channel matrices $\mathbf{H}_{\mathrm{sr}}$ and $\mathbf{H}_{\mathrm{rd}}$ are estimated by means of a BALS algorithm. The parameter setting is the same as that of Figure 2. It can be seen the proposed estimator operates satisfactorily, being able to effectively estimate the three channel matrices. Figure 3 also indicates the proposed estimator performs close to the BALS estimator operating in a two-hop system. A small performance degradation is observed, which is due to the presence of an additional AF transmission phase of our three-hop system, resulting in a higher overall noise contribution at the destination. Note also that the TALS estimator involves three estimation steps while the BALS one has two estimation steps only.

Figure 4 shows the BER performance of a linear zero forcing (ZF) receiver designed from the estimated channel matrices, which are obtained from the TALS, KRLS, or the conventional LS estimators. The ZF receiver operates on data block collected in the received data matrix $\mathbf{Y} \in \mathbb{C}^{KN_d \times N}$. The length of the data block is $N = 100$ symbols, and the remaining system parameters are the
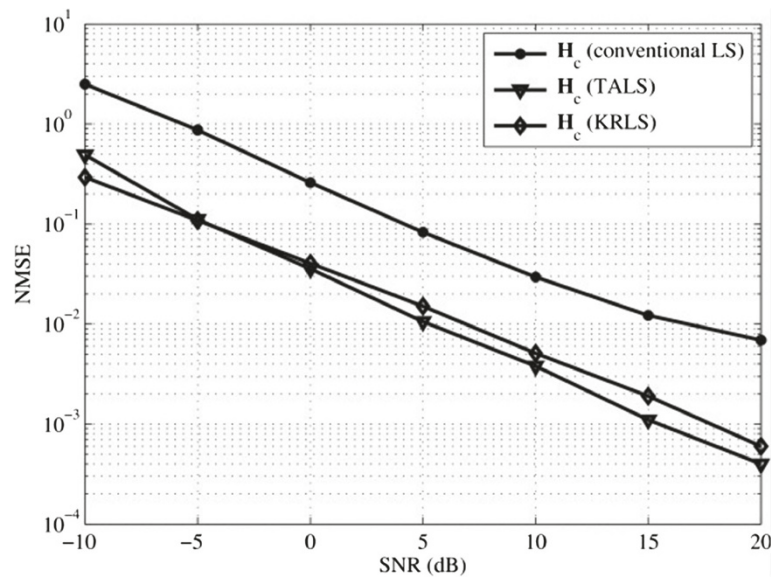
Figure 2 NMSE of estimated compound channel H$_c$. Proposed estimators (TALS and KRLS) vs. conventional LS estimator.

same as those of the previous experiment. The ZF filter output is given by:

$$
\widehat{\mathbf{S}}_{\mathrm{ZF}} = \begin{bmatrix} \mathbf{H}_{\mathrm{rd}}\mathbf{D}_1\,(\mathbf{F})\,\mathbf{H}_{\mathrm{rr}}\mathbf{D}_1\,(\mathbf{E})\,\mathbf{H}_{\mathrm{sr}} \\ \vdots \\ \mathbf{H}_{\mathrm{rd}}\mathbf{D}_K\,(\mathbf{F})\,\mathbf{H}_{\mathrm{rr}}\mathbf{D}_K\,(\mathbf{E})\,\mathbf{H}_{\mathrm{sr}} \end{bmatrix}^{\dagger} \mathbf{Y}. \tag{65}
$$

This figure shows similar BER performances for TALS and KRLS, which are better than that of the conventional LS algorithm. This result corroborates the effectiveness of our channel estimators when used with linear receiver for symbol detection. In Figure 5, we evaluate the impact of the number of relay antennas on the BER performance of a linear ZF detector using the proposed TALS channel estimator. The fixed system parameters are $N_s = 2$, $N_d = 6$, $L_0 = 30$, and $K = 10$. It can be seen that the BER performance is considerably improved as the number of relay antennas is increased, corroborating the expected gains of cooperative diversity. Although not plotted in this figure, the BER curves of the KRLS estimator are similar to those obtained with the TALS one.
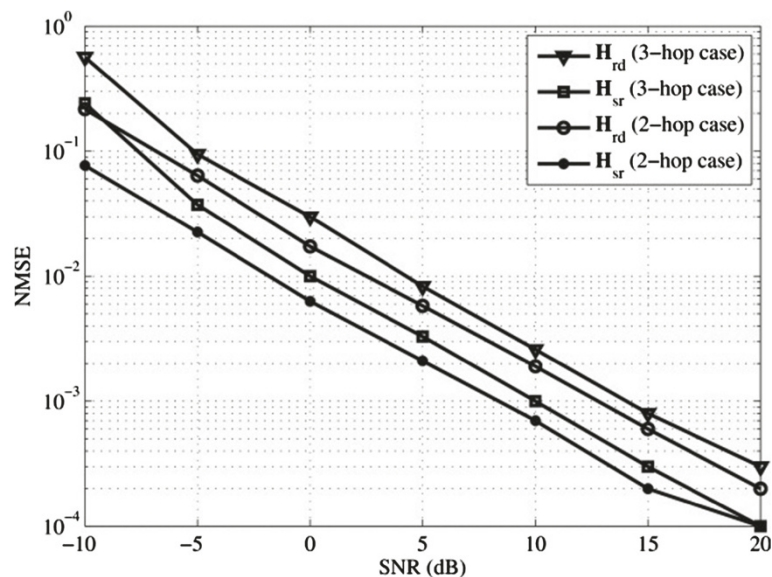
Figure 3 NMSE of estimated H$_{sr}$ and H$_{rd}$. Proposed estimators (TALS and KRLS) vs. conventional LS estimator.
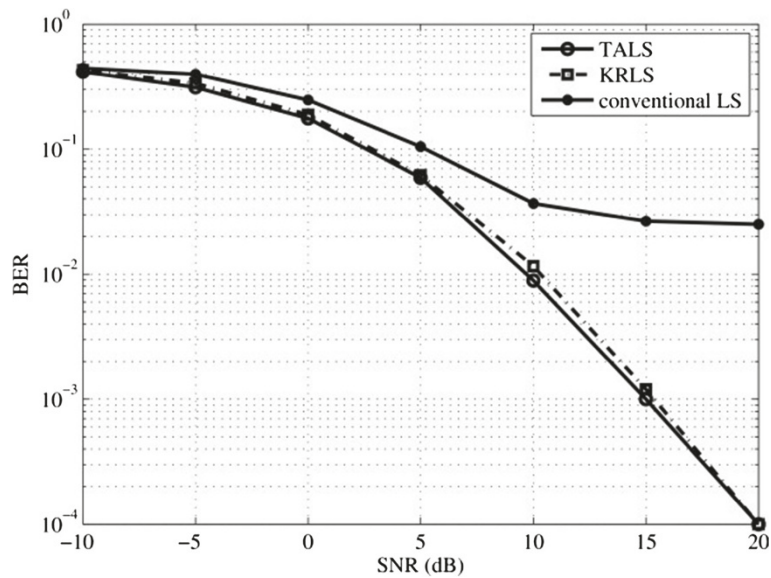
Figure 4 BER performance of linear ZF detector (TALS and KRLS). BER performance of a linear ZF detector at the destination designed from the estimated channel matrices (TALS and KRLS).

Figure 6 depicts the performance of the ZF receiver designed from the perfect CSI for all channel matrices. Two parameter settings are considered, where $N_d = 2$ and 4, respectively. The other system parameters are fixed to $N_s = 2$, $N_1 = N_2 = 3$, $L_0 = 6$, and $K = 9$. First, it can be seen that the BER performances are considerably improved as the number of antennas at the destination is increased, owing to the higher spatial diversity, as expected. From these results, we also find that the

TALS and KRLS estimators provide similar results and, more interestingly, their performances are close to that of the perfect CSI case. For instance, for a target BER of 10^{-1}, the SNR gap with respect to the perfect CSI case is less than 2 dB.

6 Conclusions
We have proposed channel estimation algorithms for MIMO AF multi-relay systems. The proposed estimators
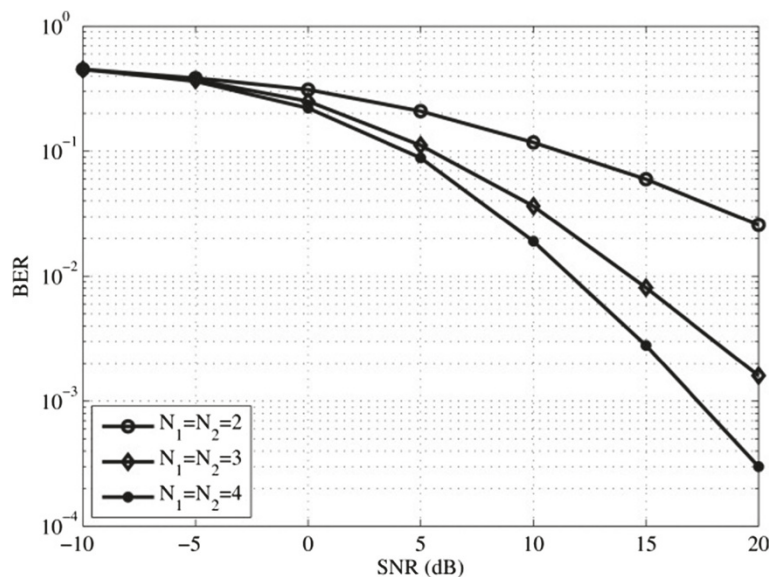
Figure 5 BER performance of linear ZF detector (TALS) versus the number of relay antennas.
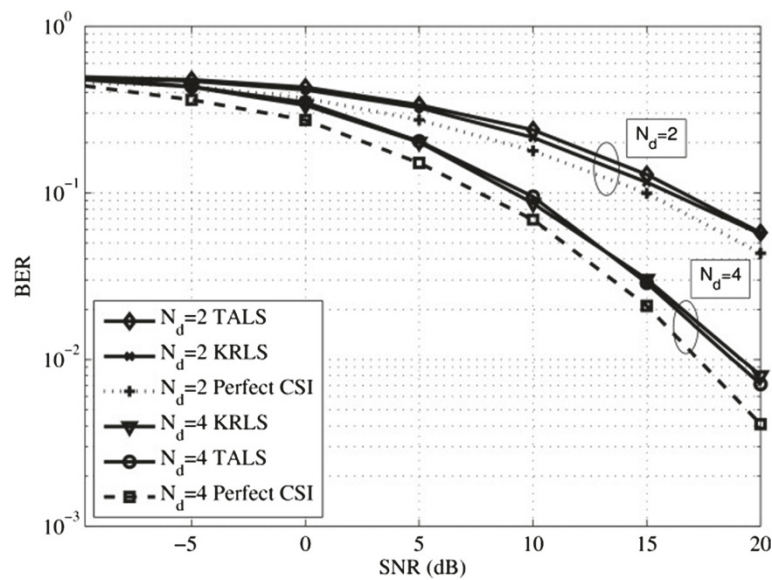
Figure 6 BER performance of linear ZF detector versus the number of antennas at the destination. BER performance of a linear ZF detector as a function of the number of antennas at the destination (TALS and KRLS).

are designed to provide the destination (base station) with the instantaneous CSI of all the channels involved in the communication. In contrast to conventional pilot-assisted channel estimation, the proposed algorithms make a more efficient use of cooperative diversity by providing a joint estimation of all the channel matrices thanks to the use of a tensor modeling of the end-to-end system. Such a joint estimation can be accomplished either iteratively (using TALS) or in closed-form (using KRLS). Our numerical results corroborate the effectiveness of the proposed algorithms. The TALS estimator has a higher computational complexity than the KRLS one due to its iterative nature. On the other hand, the minimum condition for operation of KRLS ($K \geq N_1 N_2$) is more restrictive than the identifiability conditions of TALS, which implies more training (i.e., higher number of time blocks) to carry out the joint channel estimation. Both algorithms are suitable to the joint channel estimation problem, and a particular choice is mostly dictated by practical system requirements. We have also provided an extension of the proposed approach to two-way MIMO multi-relay system and verified that such an extension results in the same tensor model as the one-way scenario. Consequently, the proposed algorithms can be applied to one- and two-way multi-relay MIMO schemes.

Endnote

[a]Since our focus is on the relay channel, direct links are not considered for simplicity. However, the idea proposed in this work can be easily extended to include direct links.

Author details
[1]Information Network Center, Beijing University of Posts and Telecommunications, Beijing, China. [2]Department of Teleinformatics Engineering, Federal University of Ceará, Campus do Pici, B. 725, 60455-970 Fortaleza, Brazil. [3]School of Computer Science, Beijing University of Posts and Telecommunications, 100876 Beijing, China.

References
1. A Sendonaris, E Erkip, B Aazhang, User cooperation diversity - part I: system description. IEEE Trans. Commun. **51**(11), 1927–1938 (2003)
2. A Sendonaris, E Erkip, B Aazhang, User cooperation diversity - part II: implementation aspects and performance analysis. IEEE Trans. Commun. **51**(11), 1939–1948 (2003)
3. JN Laneman, DNC Tse, GW Wornell, Cooperative diversity in wireless networks: efficient protocols and outage behavior. IEEE Trans. Inform. Theor. **50**(12), 3062–3080 (2004)
4. L Cao, J Zhang, N Kanno, in *Proc. IEEE GLOBECOM'09*. Multi-user cooperative communications with relay-coding for uplink IMT-advanced 4G systems, vol. 1 no. 1 (Honolulu, HI, November 2009), pp. 1–6
5. KJR Liu, AK Sadek, W Su, A Kwasinski, *Cooperative Communications and Networking* (Cambridge University Press, New York, USA, 2009)
6. R Pabst, BH Walke, DC Schultz, P Herhold, H Yanikomeroglu, S Mukherjee, H Viswanathan, M Lott, W Zirwas, M Dohler, H Aghvami, DD Falconer, GP Fettweis, Relay-based deployment concepts for wireless and mobile broadband radio. IEEE Comm. Mag. **42**(9), 80–89 (2004)
7. M Dohler, Y Li, *Cooperative Communications: Hardware, Channel and PHY* (John Wiley & Sons, West Sussex, United Kingdom, 2010)

8. Y Rong, X Tang, Y Hua, A unified framework for optimizing linear nonregenerative multicarrier MIMO relay communication systems. IEEE Trans. Signal Process. **57**(12), 4837–4851 (2009)
9. A Toding, MRA Khandaker, Y Rong, Joint source and relay optimization for parallel MIMO relay networks. EURASIP J. Adv. Signal Process. **174**, 1–7 (2012)
10. T Kong, Y Hua, Optimal design of source and relay pilots for mimo relay channel estimation. IEEE Trans. Signal Process. **59**(9), 4438–4446 (2011)
11. P Lioliou, M Viberg, M Coldrey, Efficient channel estimation techniques for amplify and forward relaying systems. IEEE Trans. Comm. **60**(11), 3150–3155 (2012)
12. F Roemer, M Haardt, Tensor-based channel estimation and iterative refinements for two-way relaying with multiple antennas and spatial reuse. IEEE Trans. Signal Process. **58**(11), 5720–5735 (2010)
13. CAR Fernandes, ALF de Almeida, DB Costa, Unified tensor modeling for blind receivers in multiuser uplink cooperative systems. IEEE Signal Process. Lett. **19**(5), 247–250 (2012)
14. Y Rong, MRA Khandaker, Y Xiang, Channel estimation of dual-hop MIMO relay system via parallel factor analysis. IEEE Trans. Wireless Comm. **11**(6), 2224–2233 (2012)
15. ALF de Almeida, CAR Fernandes, D Benevides da Costa, Multiuser detection for uplink ds-cdma amplify-and-forward relaying systems. IEEE Signal Process. Lett. **20**(7), 697–700 (2013)
16. LR Ximenes, G Favier, ALF Almeida, YCB Silva, PARAFAC-PARATUCK semi-blind receivers for two-hop cooperative MIMO relay systems. IEEE Trans. Signal Process. **62**(14), 3604–3615 (2014)
17. RA Harshman, Foundations of the PARAFAC procedure: model and conditions for an 'explanatory' multi-mode factor analysis. UCLA Working Papers Phonetics **16**, 1–84 (1970)
18. JD Carroll, J-J Chang, Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. Psychometrika **35**(3), 283–319 (1970)
19. ND Sidiropoulos, RS Budampati, Khatri-Rao space-time codes. IEEE Trans. Signal Process. **50**(10), 2396–2407 (2002)
20. RA Harshman, ME Lundy, Uniqueness proof for a family of models sharing features of Tucker's three-mode factor analysis and PARAFAC/CANDECOMP. Psychometrika **61**(1), 133–154 (1996)
21. BK Chalise, YD Zhang, MG Amin, in *Proc. of SPIE'12*. Joint Optimization of Source Beamformer and Relay Coefficients Using MSE Criterion, vol. 8404, no. 1, (May 2012)
22. M Mohammadi, M Ardebilipour, Z Mobini, R-A Zadeh, Performance analysis and power allocation for multi-hop multi-branch amplify-and-forward cooperative networks over generalized fading channels. EURASIP J. Wireless Commun. Networking **2013**(1), 1–13 (2013)
23. R Bro, Multi-way analysis in the food industry: Models, algorithms and applications. PhD thesis, University of Amsterdam, Amsterdam, 1998
24. ND Sidiropoulos, X Liu, Identifiability results for blind beamforming in incoherent multipath with small delay spread. IEEE Trans. Signal Process. **49**(1), 228–236 (2001)
25. A Smilde, R Bro, P Geladi, *Multi-way Analysis: Applications in the Chemical Sciences* (John Wiley & Sons, West Sussex, England, 2004)