RESEARCH

CrossMark

A new blind algorithm for channel estimation in OFDM-based amplify-and-forward two-way relay networks

Tzu-Chiao Lin[*] and See-May Phoong

Abstract

In this paper, we propose a blind channel estimation algorithm for the amplify-and-forward (AF) two-way relay network (TWRN) which consists of two terminal nodes and one relay node. The orthogonal frequency division multiplexing (OFDM) modulation is adopted for frequency selective channel. Both cyclic prefix (CP) and zero padding (ZP) are considered. The two cascaded channels are estimated in two steps. First, the cascaded channel causing the self-interference is estimated using a proposed power reduction method. Then, the other cascaded channel from source to destination is estimated by subspace method. Closed-form formulas for channel estimates are derived. In addition, we also carry out the theoretical mean square error analysis and derive the approximated Cramer-Rao bounds.

Keywords: Blind channel estimation, Orthogonal frequency division multiplexing (OFDM), Two-way relay network (TWRN)

1 Introduction

Research on wireless relay networks became popular since the pioneering work [1] developed low-complexity cooperative diversity strategies. In [1], data streams flow unidirectionally from the source to the relay and then to the destination. This network structure is known as the one-way relay network (OWRN). However, since most communication systems are bidirectional, it is necessary to consider the situation when the source node and the destination node exchange their roles. Such a relay network is known as the two-way relay network (TWRN). In TWRN, the relay treats the received signals in a "network coding"-like manner [2], and the terminals can recover the signal collision since they know their own transmitted signals. As a result, the overall communication rate between two source terminals in TWRN is approximately twice that achieved in OWRN [3].

Despite its throughput advantage, TWRN faces more challenges in terms of transceiver design, relay processing optimization, and transmission protocol development. In

[4], the capacity analysis and the achievable rate region for amplify-and-forward (AF) and decode-and-forward (DF) TWRN are explored. In [5], the authors point out that the throughput of AF-TWRN is 1.5 times of DF-TWRN. The distributed space-time code (STC) at relays for both AF-TWRN and DF-TWRN has been developed in [6]. Moreover, the optimal beamforming with full channel knowledge at the multi-antenna relay that maximizes the overall system capacity of AF-TWRN is derived in [7]. In [8], the authors address the problem of robust linear relay precoder and destination equalizer design for multiple-input multiple-output relay systems. In [9], the authors compare several network-coding AF-TWRN and consider imperfect time synchronization. Most existing works on TWRN [2–9] have assumed perfect channel state information (CSI) at the relay node and/or the source terminals. While traditional channel estimation methods can be applied to DF-TWRN, the channel estimation problem for AF-TWRN is more challenging due to the self-interfering signals.

In traditional channel estimation methods for point-to-point systems, they can be divided into two groups: data-aided (DA) [10–17] and non data-aided (blind) [18–26]. In general, DA channel estimation methods differ in the

*Correspondence: f99942054@ntu.edu.tw
Graduate Institute of Communication Engineering and Department of EE, National Taiwan University, Taipei, Taiwan

Springer Open

way they interpolate or filter punctual DA least square (DA-LS) channel estimates over data subcarriers. This can be accomplished using time-frequency Wiener filtering [10, 11], which is optimal in the minimum mean square error (MMSE) sense if knowledge of the channel statistics (KCS) is available. On the other hand, channel estimation can be accomplished by elaborating raw estimates in the time domain using a discrete Fourier transform (DFT)-based scheme. In [12], the MMSE channel estimator working in the time domain has been proposed. In order to reduce computational complexity, using the singular value decomposition and several low-rank approximations to the MMSE estimator has been proposed in [13] and [14]. Li et al. [12–14] also require complete KCS. In [15], the authors compare the MMSE approach with maximum likelihood (ML) channel estimation, where complete KCS is not required. This latter approach works well with dense multipath channels and quasi-uniform profiles. In practice, after the inverse DFT (IDFT), not all the channel impulse response (CIR) samples are significant because many may correspond to delays where no propagation channel paths are actually present. Therefore, the authors in [16] exploit this idea to estimate channel. In [17], the authors propose a method to approach the MMSE channel estimation performance, while avoiding the need for a priori KCS.

For blind channel estimation methods, earlier works require either higher order statistics (HOS) of the received data [18] or over-sampling at the receiver [19]. By exploiting linear redundant precoding , only second-order statistics (SOS) of the received data is required and these methods are robust to channel order overestimation [20, 21]. Another popular blind algorithm is the so-called subspace-based algorithm which was originally developed in [19]. The subspace method has simple structure and achieves good performance. In [22], a blind channel identification method by exploiting virtual carriers (VC) is derived. In [23], a generalization in cyclic prefix (CP) systems is proposed. By arranging the received data appropriately, [23] generates a rank-deduction matrix, and thus, subspace method can work. In [24], the authors propose another simpler arrangement of the received data. Pan and Phoong [25] and [26] utilize the repetition method to reduce the number of required received data and consider the existence of VCs.

As in the traditional point-to-point systems, study of channel estimation algorithm is also demanded for AF-TWRN systems [27–34]. DA channel estimation methods for AF-TWRN are proposed in [27–30]. Gao et al. [27] develops an optimal training design for flat-fading environment. The authors also combine their algorithm with orthogonal frequency division multiplexing (OFDM) to estimate the channel impulse responses for frequency selective environment in [28]. The case of multiple-input

multiple-output is considered in [29], and [30] provides two channel training algorithms for channel estimation.

On the other hand, [31–34] are blind channel estimation methods. In [31], the authors propose a ML approach to estimate the flat-fading channels blindly, but the transmitted signals are limited to constant modulus modulation. Zhao et al. [32] find a closed-form solution and thus provides a low-complexity ML algorithm. For non-constant modulus modulation, [33] gives an iterative algorithm, which is based on the maximum a posteriori (MAP) approach, and it requires a large number of received blocks. In [34], the authors consider the frequency selective environment. They apply a non-unitary linear precoding at both terminals and derive a blind channel estimation algorithm from SOS of the received signals. However, the use of non-unitary linear precoding leads to degradation in bit error rate (BER) performance.

In this paper, we develop a blind channel estimation algorithm for AF-TWRN under OFDM modulation. Our method consists of two steps. The first step is to estimate the cascaded channel causing the self-interference. Since the terminal knows its own transmitted signal, we choose the method based on power reduction to estimate the channel, which is also named LS method. The self-interference signal can be removed by using the estimated channel. The second step is to estimate the cascaded channel from source to destination. We utilize the rank reduction method, which is also known as subspace-based algorithm [23–26]. This is because subspace methods do not require complete KCS, work well with all multipath channels, and achieve good performance. Closed-form formulas for these two cascaded channel estimates are derived. The theoretical performance analysis and approximated Cramer-Rao bounds (ACRB) are given as well. The proposed method can be applied to both CP-based and zero padding (ZP)-based OFDM systems. Simulation results will be provided to show the performance of the proposed method.

The rest of this paper is organized as follows. The system model for CP-OFDM AF-TWRN is introduced in Section 2. Section 3 describes the proposed algorithm for blind channel estimation. In Section 4, we analyze the performance of the proposed channel estimation methods and the ACRBs. Simulation results are presented in Section 5, and concluding remarks are made in Section 6. The results in Section 3.1 and 3.2 of this paper have appeared in a conference paper [35].

Notation In this paper, $E\{x\}$ stands for the statistical expectation of the random variable x. The symbols $\mathbf{A}^T$, $\mathbf{A}^*$, and $\mathbf{A}^\dagger$ denote the transpose, the complex conjugate, and the conjugate-transpose of matrix $\mathbf{A}$, respectively. $\|\mathbf{A}\|_F$ is the Frobenius norm of matrix $\mathbf{A}$. If $\mathbf{A}$ is a square, $tr(\mathbf{A})$ denotes the trace of matrix $\mathbf{A}$. $\mathbf{I}_m$ is the $m \times m$

identity matrix, whereas **0** represents an all-zero matrix with appropriate dimension. $J = \sqrt{-1}$ is the imaginary unit. $\mathbf{T}_m(\mathbf{c})$ and $\tilde{\mathbf{T}}_m(\mathbf{c})$ are two Toeplitz matrices respectively defined as

$$\mathbf{T}_m(\mathbf{c}) \triangleq \left.\begin{bmatrix} c_n & \cdots & c_1 & 0 & \cdots & 0 \\ 0 & c_n & \cdots & c_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & c_n & \cdots & c_1 \end{bmatrix}\right\} m \text{ rows} \quad (1)$$

and

$$\tilde{\mathbf{T}}_m(\mathbf{c}) \triangleq \underbrace{\begin{bmatrix} c_1 & 0 & \cdots & 0 \\ \vdots & c_1 & \ddots & \vdots \\ c_n & \vdots & \ddots & 0 \\ 0 & c_n & \ddots & c_1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & c_n \end{bmatrix}}_{m \text{ columns}}, \quad (2)$$

where $\mathbf{c} = [c_1, c_2, \ldots, c_n]^T$ is an arbitrary vector.

2 System model

Consider a TWRN with two terminal nodes $\mathbb{T}_1$ and $\mathbb{T}_2$, and one relay node $\mathbb{R}$, as shown in Fig. 1. Each node has one antenna which cannot transmit and receive simultaneously. The channel from $\mathbb{T}_i$ to $\mathbb{R}$ is denoted as $\mathbf{f}_i = [f_{i,0}, f_{i,1}, \ldots, f_{i,L}]^T$, whereas the one from $\mathbb{R}$ back to $\mathbb{T}_i$ is denoted as $\mathbf{g}_i = [g_{i,0}, g_{i,1}, \ldots, g_{i,L}]^T$ for $i = 1$ and 2. For notational simplicity, we assume that the lengths of $\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{g}_1$, and $\mathbf{g}_2$ do not exceed $L+1$.[1] Similar to most other algorithms, we assume that the channels do not change when the channel estimation is performed.

2.1 OFDM modulation at terminals

Denote the kth OFDM block from $\mathbb{T}_i$ as $\mathbf{s}_k^{(i)} = \left[s_{k,0}^{(i)}, s_{k,1}^{(i)}, \ldots, s_{k,N-1}^{(i)}\right]^T$, where N is the OFDM block length. The corresponding time domain signal block is obtained from the normalized IDFT as

$$\mathbf{x}_k^{(i)} = \mathbf{W}^\dagger \mathbf{s}_k^{(i)} = \left[x_{k,0}^{(i)} \ x_{k,1}^{(i)} \ \cdots \ x_{k,N-1}^{(i)} \right]^T, \quad (3)$$

where $\mathbf{W}$ is the $N \times N$ normalized DFT matrix with the (m, n)th entry given by $\frac{1}{\sqrt{N}} e^{-J 2\pi mn/N}$. To maintain the subcarrier orthogonality during the overall transmission, we propose to add a CP of length $2L$.[2] This implicitly requires $N \geq 2L$ which is nevertheless satisfied by most OFDM systems. Define $\mathbf{x}_{k,cp}^{(i)} = [x_{k,N-2L}^{(i)}, \ldots, x_{k,N-1}^{(i)}]^T$. The signal sent out from $\mathbb{T}_i$ is expressed as $\left[\mathbf{x}_{k,cp}^{(i)T} \ \mathbf{x}_k^{(i)T} \right]^T$ for $i = 1$ and 2.

2.2 Relay processing

The relay $\mathbb{R}$ receives the signal [34]

$$\mathbf{r}_k = \begin{bmatrix} r_{k,0} \\ r_{k,1} \\ \vdots \\ r_{k,N+2L-1} \end{bmatrix} = \sum_{i=1}^{2} \mathbf{T}_{N+2L}(\mathbf{f}_i) \begin{bmatrix} \mathbf{x}_{k-1,isi}^{(i)} \\ \mathbf{x}_{k,cp}^{(i)} \\ \mathbf{x}_k^{(i)} \end{bmatrix} + \mathbf{n}_{k,r}, \quad (4)$$

where $\mathbf{x}_{k-1,isi}^{(i)}$ is the term which causes the inter-symbol interference (ISI):

$$\mathbf{x}_{k-1,isi}^{(i)} = \begin{bmatrix} x_{k-1,N-L}^{(i)} \\ \vdots \\ x_{k-1,N-1}^{(i)} \end{bmatrix}. \quad (5)$$

Moreover, each element in the noise vector $\mathbf{n}_{k,r}$ is assumed to be independent and identically distributed (i.i.d.) zero-mean complex white Gaussian.

We assume that the relay $\mathbb{R}$ employs the amplify-and-forward scheme. It scales $\mathbf{r}_k$ by the factor of

$$\alpha = \sqrt{\frac{P_r}{\mathrm{E}\left\{\|\mathbf{r}_k\|_F^2\right\}}} = \sqrt{\frac{P_r}{\|\mathbf{f}_1\|_F^2 \sigma_1^2 + \|\mathbf{f}_2\|_F^2 \sigma_2^2 + \sigma_{n_r}^2}}, \quad (6)$$

where P_r is the average transmission power of $\mathbb{R}$. In the second equality, we have made the assumptions that the transmitted signals $\mathbf{x}_k^{(1)}$, $\mathbf{x}_k^{(2)}$, and the received noise $\mathbf{n}_{k,r}$ are uncorrelated with variances σ_1^2, σ_2^2, and $\sigma_{n_r}^2$, respectively. Then, the relay broadcasts $\alpha \mathbf{r}_k$ to both terminals.
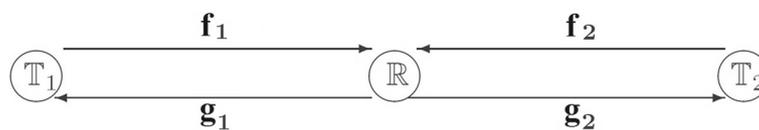
Fig. 1 System configuration for two-way relay network. It shows a two-way relay network with two terminal nodes $\mathbb{T}_1$ and $\mathbb{T}_2$, and one relay node $\mathbb{R}$. Each node has one antenna which cannot transmit and receive simultaneously. The channel from $\mathbb{T}_i$ to $\mathbb{R}$ is denoted as $\mathbf{f}_i$, whereas the one from $\mathbb{R}$ back to $\mathbb{T}_i$ is denoted as $\mathbf{g}_i$ for $i = 1$ and 2

2.3 Signal reformulation at terminals

Due to symmetry, we only illustrate the processing at $\mathbb{T}_1$. The $(N+2L) \times 1$ vector received at $\mathbb{T}_1$ can be expressed as

$$\mathbf{y}_k = \begin{bmatrix} y_{k,0} \\ y_{k,1} \\ \vdots \\ y_{k,N+2L-1} \end{bmatrix} = \mathbf{T}_{N+2L}(\mathbf{g}_1) \begin{bmatrix} \alpha \mathbf{r}_{k-1,isi} \\ \alpha \mathbf{r}_k \end{bmatrix} + \mathbf{n}_{k,t}, \tag{7}$$

where $\mathbf{r}_{k-1,isi}$ is similar to (5)

$$\mathbf{r}_{k-1,isi} = \begin{bmatrix} r_{k-1,N+L} \\ \vdots \\ r_{k-1,N+2L-1} \end{bmatrix}, \tag{8}$$

and each element in the noise vector $\mathbf{n}_{k,t}$ is assumed to be i.i.d. zero-mean complex white Gaussian, with variance $\sigma_{n_t}^2$. Substituting (4) into (7), we have

$$\mathbf{y}_k = \mathbf{T}_{N+2L}(\mathbf{h}_1) \begin{bmatrix} \mathbf{x}_{k-1,cp}^{(1)} \\ \mathbf{x}_{k,cp}^{(1)} \\ \mathbf{x}_k^{(1)} \end{bmatrix} + \mathbf{T}_{N+2L}(\mathbf{h}_2) \begin{bmatrix} \mathbf{x}_{k-1,cp}^{(2)} \\ \mathbf{x}_{k,cp}^{(2)} \\ \mathbf{x}_k^{(2)} \end{bmatrix} + \mathbf{n}_{k,e}, \tag{9}$$

where $\mathbf{h}_1 = \alpha(\mathbf{g}_1 * \mathbf{f}_1)$ and $\mathbf{h}_2 = \alpha(\mathbf{g}_1 * \mathbf{f}_2)$ with $*$ being the linear convolution between two vectors by the fact that the multiplication of two Toeplitz matrices is still a Toeplitz matrix. The last term $\mathbf{n}_{k,e}$ denotes the equivalent noise

$$\mathbf{n}_{k,e} = \alpha \mathbf{T}_{N+2L}(\mathbf{g}_1) \begin{bmatrix} n_{k-1,r}(N+L) \\ \vdots \\ n_{k-1,r}(N+2L-1) \\ \mathbf{n}_{k,r} \end{bmatrix} + \mathbf{n}_{k,t}. \tag{10}$$

When $N \gg L$, $\mathbf{n}_{k,e}$ can be approximated as white noise.

2.4 Data detection at terminals

After removing the first $2L$ elements of $\mathbf{y}_k$ in (9), we obtain a vector of size N:

$$\bar{\mathbf{y}}_k = \mathbf{T}_N(\mathbf{h}_1) \begin{bmatrix} \mathbf{x}_{k,cp}^{(1)} \\ \mathbf{x}_k^{(1)} \end{bmatrix} + \mathbf{T}_N(\mathbf{h}_2) \begin{bmatrix} \mathbf{x}_{k,cp}^{(2)} \\ \mathbf{x}_k^{(2)} \end{bmatrix} + \bar{\mathbf{n}}_{k,e}, \tag{11}$$

where $\bar{\mathbf{n}}_{k,e}$ is the last N elements of $\mathbf{n}_{k,e}$. If the cascaded channel $\mathbf{h}_1$ is known to $\mathbb{T}_1$, then the first term on the right-hand side of (11) can be removed since $\mathbb{T}_1$ knows its own signal $\mathbf{x}_k^{(1)}$. If $\mathbf{h}_2$ is known, the regular OFDM detection can be efficiently performed using fast Fourier transform. So $\mathbb{T}_1$ can recover the data from $\mathbb{T}_2$ if both $\mathbf{h}_1$ and $\mathbf{h}_2$ are available. Hence, our goal is to estimate $\mathbf{h}_1$ and $\mathbf{h}_2$. Below, we will show how to blindly estimate these two cascaded channels from the received signal $\mathbf{y}_k$.

3 Proposed method for channel estimation

In this paper, we assume that $\mathbf{x}_k^{(1)}$ and $\mathbf{x}_k^{(2)}$ are uncorrelated. Moreover, the transmitted signals and the noises are uncorrelated as well. Under these two assumptions, we propose an algorithm to estimate $\mathbf{h}_1$ and $\mathbf{h}_2$ blindly. Though our derivations are based on CP-OFDM system, the results can be also extended to ZP-OFDM system. The details will be discussed later.

3.1 The estimation of $\mathbf{h}_1$

Let us look at the received vector $\bar{\mathbf{y}}_k$ in (11). Notice that $\mathbf{x}_k^{(1)}$ is known at $\mathbb{T}_1$. If we have a perfect estimate of $\mathbf{h}_1$, then the first term at the right-hand side of (11) can be eliminated completely from $\bar{\mathbf{y}}_k$. Due to uncorrelatedness of $\mathbf{x}_k^{(1)}$, $\mathbf{x}_k^{(2)}$, and $\bar{\mathbf{n}}_{k,e}$, the power of $\bar{\mathbf{y}}_k$ will be reduced when $\mathbf{x}_k^{(1)}$ is eliminated from $\bar{\mathbf{y}}_k$. Based on this power reduction, we are able to derive a closed-form formula for an estimate of the $(2L + 1) \times 1$ vector $\mathbf{h}_1$, as shown below.

Define a cost function

$$J\left(\hat{\mathbf{h}}_1\right) = \mathrm{E}\left\{ \left\| \bar{\mathbf{y}}_k - \mathbf{T}_N\left(\hat{\mathbf{h}}_1\right) \begin{bmatrix} \mathbf{x}_{k,cp}^{(1)} \\ \mathbf{x}_k^{(1)} \end{bmatrix} \right\|_F^2 \right\}, \tag{12}$$

where $\bar{\mathbf{y}}_k$ is the $N \times 1$ vector in (11) and $\hat{\mathbf{h}}_1$ is an estimate of $\mathbf{h}_1$. Substituting (11) into (12), we get

$$
\begin{aligned}
J\left(\hat{\mathbf{h}}_1\right) &= \mathrm{E}\left\{ \left\| \left(\mathbf{T}_N(\mathbf{h}_1) - \mathbf{T}_N\left(\hat{\mathbf{h}}_1\right)\right) \begin{bmatrix} \mathbf{x}_{k,cp}^{(1)} \\ \mathbf{x}_k^{(1)} \end{bmatrix} \right. \right. \\
&\quad \left. \left. + \mathbf{T}_N(\mathbf{h}_2) \begin{bmatrix} \mathbf{x}_{k,cp}^{(2)} \\ \mathbf{x}_k^{(2)} \end{bmatrix} + \bar{\mathbf{n}}_{k,e} \right\|_F^2 \right\} \\
&= \mathrm{E}\left\{ \left\| \mathbf{T}_N\left(\mathbf{h}_1 - \hat{\mathbf{h}}_1\right) \begin{bmatrix} \mathbf{x}_{k,cp}^{(1)} \\ \mathbf{x}_k^{(1)} \end{bmatrix} \right\|_F^2 \right\} \\
&\quad + \mathrm{E}\left\{ \left\| \mathbf{T}_N(\mathbf{h}_2) \begin{bmatrix} \mathbf{x}_{k,cp}^{(2)} \\ \mathbf{x}_k^{(2)} \end{bmatrix} \right\|_F^2 \right\} + \mathrm{E}\left\{ \left\| \bar{\mathbf{n}}_{k,e} \right\|_F^2 \right\}.
\end{aligned}
\tag{13}
$$

Using the assumptions mentioned above to simplify the expression, we have

$$
\begin{aligned}
J\left(\hat{\mathbf{h}}_1\right) &= N\left(\sigma_1^2 \left\| \mathbf{h}_1 - \hat{\mathbf{h}}_1 \right\|_F^2 + \sigma_2^2 \left\| \mathbf{h}_2 \right\|_F^2 + |\alpha|^2 \sigma_{n_r}^2 \left\| \mathbf{g}_1 \right\|_F^2 + \sigma_{n_t}^2 \right) \\
&\geq N\left(\sigma_2^2 \left\| \mathbf{h}_2 \right\|_F^2 + |\alpha|^2 \sigma_{n_r}^2 \left\| \mathbf{g}_1 \right\|_F^2 + \sigma_{n_t}^2 \right).
\end{aligned}
\tag{14}
$$

Obviously, the cost function has the minimum if and only if $\|\mathbf{h}_1 - \hat{\mathbf{h}}_1\|_F^2 = 0$, or equivalently, $\hat{\mathbf{h}}_1 = \mathbf{h}_1$. Assume that $\mathbb{T}_1$ has collected K blocks. For mean-ergodic processes, the ensemble average (or statistical average) can be well approximated by the time average:

$$\bar{J}\left(\hat{\mathbf{h}}_1\right) = \frac{1}{K}\sum_{k=0}^{K-1}\left\|\bar{\mathbf{y}}_k - \mathbf{T}_N\left(\hat{\mathbf{h}}_1\right)\begin{bmatrix}\mathbf{x}_{k,cp}^{(1)}\\\mathbf{x}_k^{(1)}\end{bmatrix}\right\|_F^2$$

$$= \frac{1}{K}\sum_{k=0}^{K-1}\left\|\bar{\mathbf{y}}_k - \sqrt{N}\mathbf{W}^\dagger\mathbf{D}\left(\mathbf{s}_k^{(1)}\right)\mathbf{W}_{2L+1}\hat{\mathbf{h}}_1\right\|_F^2, \tag{15}$$

where $\mathbf{D}\left(\mathbf{s}_k^{(1)}\right)$ is a diagonal matrix with the elements of $\mathbf{s}_k^{(1)}$ on the main diagonal, and $\mathbf{W}_{2L+1}$ is the first $2L+1$ columns of the DFT matrix $\mathbf{W}$. Let

$$\mathbf{y} = \begin{bmatrix}\bar{\mathbf{y}}_0^T & \bar{\mathbf{y}}_1^T & \cdots & \bar{\mathbf{y}}_{K-1}^T\end{bmatrix}^T \tag{16}$$

and

$$\mathbf{S} = \begin{bmatrix}\mathbf{D}\left(\mathbf{s}_0^{(1)}\right) & \mathbf{D}\left(\mathbf{s}_1^{(1)}\right) & \cdots & \mathbf{D}\left(\mathbf{s}_{K-1}^{(1)}\right)\end{bmatrix}^T. \tag{17}$$

Then, (15) can be rewritten as

$$\bar{J}\left(\hat{\mathbf{h}}_1\right) = \frac{1}{K}\left\|\mathbf{y} - \sqrt{N}\left(\mathbf{I}_K\otimes\mathbf{W}^\dagger\right)\mathbf{S}\mathbf{W}_{2L+1}\hat{\mathbf{h}}_1\right\|_F^2, \tag{18}$$

where the symbol $\otimes$ denotes the Kronecker product. The least squares solution of (18) can be calculated as

$$\hat{\mathbf{h}}_1 = \frac{1}{\sqrt{N}}\left(\mathbf{W}_{2L+1}^\dagger\mathbf{S}^\dagger\mathbf{S}\mathbf{W}_{2L+1}\right)^{-1}\mathbf{W}_{2L+1}^\dagger\mathbf{S}^\dagger\left(\mathbf{I}_K\otimes\mathbf{W}\right)\mathbf{y}. \tag{19}$$

When $K >> 1$, we have $\mathbf{S}^\dagger\mathbf{S}\approx K\sigma_1^2\mathbf{I}_N$ as the modulation symbols are statistically independent. In this case, (19) can be approximated as

$$\hat{\mathbf{h}}_1 \approx \frac{1}{\sqrt{N}K\sigma_1^2}\mathbf{W}_{2L+1}^\dagger\sum_{k=0}^{K-1}\left(\mathbf{s}_k^{(1)}\right)^*\odot\left(\mathbf{W}\bar{\mathbf{y}}_k\right), \tag{20}$$

where the symbol $\odot$ denotes the Hadamard product. Notice that there is no scalar ambiguity in the estimation of $\mathbf{h}_1$ since $\mathbf{s}_k^{(1)}$ and $\bar{\mathbf{y}}_k$ are known at $\mathbb{T}_1$.

3.2 The estimation of $\mathbf{h}_2$

In order to estimate the $(2L+1)\times 1$ vector $\mathbf{h}_2$, we first remove the self-interfering signal from the received vector. Define

$$\mathbf{z}_k = \mathbf{y}_k - \mathbf{T}_{N+2L}\left(\hat{\mathbf{h}}_1\right)\begin{bmatrix}\mathbf{x}_{k-1,cp}^{(1)}\\\mathbf{x}_{k,cp}^{(1)}\\\mathbf{x}_k^{(1)}\end{bmatrix}. \tag{21}$$

Assuming that the estimation of $\mathbf{h}_1$ is perfect (i.e., $\hat{\mathbf{h}}_1 = \mathbf{h}_1$), from (9) and (21), we have

$$\mathbf{z}_k = \mathbf{T}_{N+2L}(\mathbf{h}_2)\begin{bmatrix}\mathbf{x}_{k-1,cp}^{(2)}\\\mathbf{x}_{k,cp}^{(2)}\\\mathbf{x}_k^{(2)}\end{bmatrix} + \mathbf{n}_{k,e}. \tag{22}$$

Note that the vector $\mathbf{z}_k$ is simply the received vector in a usual CP-OFDM system with channel $\mathbf{h}_2$ and transmitted vector $\mathbf{x}_k^{(2)}$. Many blind estimation methods have been proposed for the estimation of $\mathbf{h}_2$ from $\mathbf{z}_k$. Below, we will adopt the subspace-based algorithm in [24]. Define the re-modulated vector

$$\tilde{\mathbf{z}}_k = \begin{bmatrix}z_{k-1,2L}\\\vdots\\z_{k-1,N+2L-1}\\z_{k,0}\\\vdots\\z_{k,2L-1}\end{bmatrix}, \tag{23}$$

where $z_{k,i}$ is the ith entry of $\mathbf{z}_k$. That is, $\tilde{\mathbf{z}}_k$ is a $(N+2L)\times 1$ vector formed by the last N entries of $\mathbf{z}_{k-1}$ and the first $2L$ entries of $\mathbf{z}_k$. Next, we construct the vector

$$\mathbf{v}_k = \mathbf{z}_k - \tilde{\mathbf{z}}_k. \tag{24}$$

Substituting (9) and (21) into (24), we have

$$\mathbf{v}_k = \tilde{\mathbf{T}}_N(\mathbf{h}_2)\underbrace{\left(\begin{bmatrix}\mathbf{x}_{k,cp}^{(2)}\\x_{k,0}^{(2)}\\\vdots\\x_{k,N-2L-1}^{(2)}\end{bmatrix} - \mathbf{x}_{k-1}^{(2)}\right)}_{\triangleq\mathbf{d}_k} + \boldsymbol{\eta}_k, \tag{25}$$

where $\tilde{\mathbf{T}}_N(\cdot)$ is defined in (2) and $\boldsymbol{\eta}_k$ is colored noise. The covariance matrix of $\boldsymbol{\eta}_k$ is [24]

$$\mathrm{E}\{\boldsymbol{\eta}_k\boldsymbol{\eta}_k^\dagger\} = \sigma_{n_e}^2\underbrace{\begin{bmatrix}2\mathbf{I}_{2L} & \mathbf{0} & -\mathbf{I}_{2L}\\\mathbf{0} & 2\mathbf{I}_{N-2L} & \mathbf{0}\\-\mathbf{I}_{2L} & \mathbf{0} & 2\mathbf{I}_{2L}\end{bmatrix}}_{\triangleq\mathbf{R}_w},$$

where $\sigma_{n_e}^2$ is the average power of $\mathbf{n}_{k,e}$. It can be verified that

$$\mathbf{R}_w^{-1/2} = \begin{bmatrix}c_1\mathbf{I}_{2L} & \mathbf{0} & c_2\mathbf{I}_{2L}\\\mathbf{0} & \frac{1}{\sqrt{2}}\mathbf{I}_{N-2L} & \mathbf{0}\\c_2\mathbf{I}_{2L} & \mathbf{0} & c_1\mathbf{I}_{2L}\end{bmatrix} \tag{26}$$

with

$$c_1 = \sqrt{\frac{2/3+\sqrt{1/3}}{2}} \text{ and } c_2 = \sqrt{\frac{2/3-\sqrt{1/3}}{2}}.$$

Carrying out the whitening process on $\mathbf{v}_k$, we get the whitened vector $\mathbf{v}_k^{(w)} = \mathbf{R}_w^{-1/2}\mathbf{v}_k$ and its covariance matrix is

$$\mathbf{R}_v^{(w)} = \mathbf{R}_w^{-1/2}\tilde{\mathbf{T}}_N(\mathbf{h}_2)\mathbf{R}_d\tilde{\mathbf{T}}_N^\dagger(\mathbf{h}_2)\mathbf{R}_w^{-1/2} + \sigma_{n_e}^2\mathbf{I}_{N+2L}, \tag{27}$$

where $\mathbf{R}_d = \mathrm{E}\left\{\mathbf{d}_k\mathbf{d}_k^\dagger\right\}$ is the covariance matrix of $\mathbf{d}_k$ defined in (25). A necessary condition that $\mathbf{R}_d$ has full rank

is that $\mathbb{T}_1$ collects $K \geq N$ blocks. Utilizing eigenvalue decomposition, (27) can be computed as

$$\mathbf{R}_v^{(w)} = \mathbf{U}_s \Sigma \mathbf{U}_s^\dagger + \sigma_{n_e}^2 \mathbf{U}_o \mathbf{U}_o^\dagger, \qquad (28)$$

where Σ is an $N \times N$ diagonal matrix and the $(N+2L) \times N$ matrix $\mathbf{U}_s$ spans the signal subspace. On the other hand, the $(N + 2L) \times 2L$ matrix $\mathbf{U}_o$ spans the noise subspace. That is,

$$\mathbf{U}_o^\dagger \mathbf{R}_w^{-1/2} \tilde{\mathbf{T}}_N(\mathbf{h}_2) = \mathbf{0}. \qquad (29)$$

Let

$$\mathbf{J}_i = \begin{bmatrix} \mathbf{0}_{i \times (2L+1)} \\ \mathbf{I}_{2L+1} \\ \mathbf{0}_{(N-1-i) \times (2L+1)} \end{bmatrix} \text{ for } i = 0, 1, \ldots, N-1. \qquad (30)$$

Then, (29) can be rewritten as

$$\underbrace{\begin{bmatrix} \mathbf{U}_o^\dagger \mathbf{R}_w^{-1/2} \mathbf{J}_0 \\ \vdots \\ \mathbf{U}_o^\dagger \mathbf{R}_w^{-1/2} \mathbf{J}_{N-1} \end{bmatrix}}_{\triangleq \mathbf{U}} \mathbf{h}_2 = \mathbf{0}. \qquad (31)$$

Hence, we can estimate $\mathbf{h}_2$ (up to a scalar ambiguity) by calculating the eigenvector corresponding to the smallest eigenvalue of $\mathbf{U}^\dagger \mathbf{U}$.

In summary, our algorithm is as follows.

1. Estimate $\mathbf{h}_1$ by (20).
2. Eliminate the interference from $\mathbb{T}_1$ by (21).
3. Calculate $\mathbf{v}_k^{(w)} = \mathbf{R}_w^{-1/2} \mathbf{v}_k$ by (24) and (26) and obtain the $(N + 2L) \times 2L$ matrix $\mathbf{U}_o$ spanning the noise subspace by eigenvalue decomposition.
4. Estimate $\mathbf{h}_2$ (up to a scalar ambiguity) by calculating the eigenvector corresponding to the smallest eigenvalue of $\mathbf{U}^\dagger \mathbf{U}$.

3.3 A note on the identifiability issue
Note that the estimate of $\mathbf{h}_1$ is unique because the cost function in (14) has a unique minimum at $\hat{\mathbf{h}}_1 = \mathbf{h}_1$. The second channel $\mathbf{h}_2$ is estimated by the subspace method. Let us look at the vector $\mathbf{z}_k$ in (22). When the self-interfering signal is completely eliminated, the remaining part $\mathbf{z}_k$ is identical to the case of single-input single-output (SISO) CP-OFDM system in [24]. The identifiability issue of this method has been studied in [24]. It has been shown that if $h_{2,0} \neq 0$, then the vector $\mathbf{h}_2$ is uniquely determined (up to a scalar ambiguity).

3.4 Comparison with an existing work
A blind channel estimation algorithm in OFDM-based TWRN was proposed in [34]. Comparing our method with that in [34], there are two major differences. One is that [34] requires a precoding matrix $\mathbf{P}$, where

$$\mathbf{P}\mathbf{P}^\dagger = \begin{bmatrix} 1 & \theta & \cdots & \theta \\ \theta & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \theta \\ \theta & \cdots & \theta & 1 \end{bmatrix}.$$

A necessary condition on θ is $-\frac{1}{N-1} \leq \theta \leq 1$. In other words, the kth transmitted vector from $\mathbb{T}_i$ is the precoded vector $\mathbf{P}\mathbf{s}_k^{(i)}$ instead of $\mathbf{s}_k^{(i)}$. Notice that for $\theta \neq 0$, $\mathbf{P}$ is not a unitary matrix. The channel noise can be amplified when the receiver performs the operation $\mathbf{P}^{-1}$. It was shown in [34] that when θ increases from 0 to 1, the mean square error (MSE) of channel estimate decreases. Due to noise amplification, larger θ does not necessarily yield smaller BER, so there exists a compromise between channel estimation error and BER. Another difference between our method and [34] is that there is a 2×2 ambiguity matrix in [34], or equivalently, there are four ambiguity scalars. On the other hand, there is only one ambiguity scalar in our algorithm. In terms of complexity, we can see that the main complexity of our method is the computation of the eigenvalue decomposition of an $N \times N$ matrix in (28), whereas the eigenvalue decomposition in [34] is for a $(2L + 1) \times (2L + 1)$ matrix. Hence, our method is more complicated than [34].

3.5 Repeated use of the remodulated vector $\mathbf{v}_k$
To obtain $\mathbf{U}_o$ in (28), $\mathbb{T}_1$ has to collect $K \geq N$ blocks. In OFDM systems, N is usually large. The number of blocks, K, needed for the channel estimation is large. In order to reduce the required block number K, we can use the repetition method proposed in [23, 25, 26]. Define the repetition parameter Q and form the matrix $\tilde{\mathbf{T}}_Q(\mathbf{v}_k)$, where $\mathbf{v}_k$ is defined in (24). According to (25), $\tilde{\mathbf{T}}_Q(\mathbf{v}_k)$ can be represented as

$$\tilde{\mathbf{T}}_Q(\mathbf{v}_k) = \tilde{\mathbf{T}}_{N+Q-1}(\mathbf{h}_2)\tilde{\mathbf{T}}_Q(\mathbf{d}_k) + \tilde{\mathbf{T}}_Q(\boldsymbol{\eta}_k). \qquad (32)$$

It was shown in [25] that $\tilde{\mathbf{T}}_Q(\boldsymbol{\eta}_k)$ is colored noise, and its covariance matrix can be calculated as

$$\mathrm{E}\left\{\tilde{\mathbf{T}}_Q\left(\boldsymbol{\eta}_k\right)\tilde{\mathbf{T}}_Q^\dagger(\boldsymbol{\eta}_k)\right\} = \sigma_{n_e}^2 \sum_{q=1}^{Q} \begin{bmatrix} \mathbf{0}_{(q-1)\times(q-1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_w & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_{(Q-q)\times(Q-q)} \end{bmatrix}$$
$$= \sigma_{n_e}^2 \mathbf{E}\Lambda\mathbf{E}^\dagger, \qquad (33)$$

where we have applied the eigenvalue decomposition in the second equality. Therefore, we need to whiten the matrix $\tilde{\mathbf{T}}_Q(\mathbf{v}_k)$ by $\mathbf{E}\Lambda^{-1/2}\mathbf{E}^\dagger$. Since each vector $\mathbf{v}_k$ is repeated Q times in (32), the required number of blocks becomes $K \geq \frac{N-1}{Q} + 1$ blocks [23]. Collecting these K blocks, we can follow the procedure in (28)–(31) to estimate $\mathbf{h}_2$ (up to a scalar ambiguity).

3.6 Multiple relay nodes

The extension to the case of multiple relay nodes is straight forward as shown in Fig. 2. Suppose that we have M relay nodes $\mathbb{R}_1, \mathbb{R}_2, \ldots, \mathbb{R}_M$. Let the channels from $\mathbb{T}_i$ to $\mathbb{R}_m$ be denoted as $\mathbf{f}_i^{(m)}$ and the channels from $\mathbb{R}_m$ to $\mathbb{T}_i$ be denoted as $\mathbf{g}_i^{(m)}$. Then, (4) becomes

$$\mathbf{r}_k^{(m)} = \sum_{i=1}^{2} \mathbf{T}_{N+2L}\left(\mathbf{f}_i^{(m)}\right) \begin{bmatrix} \mathbf{x}_{k-1,isi}^{(i)} \\ \mathbf{x}_{k,cp}^{(i)} \\ \mathbf{x}_k^{(i)} \end{bmatrix} + \mathbf{n}_{k,r}^{(m)}, \qquad (34)$$

where $\mathbf{r}_k^{(m)}$ is the signal received by relay node $\mathbb{R}_m$ and $\mathbf{n}_{k,r}^{(m)}$ is the noise at $\mathbb{R}_m$. When $\mathbb{T}_1$ receives the signal, (7) becomes

$$\mathbf{y}_k = \sum_{m=1}^{M} \mathbf{T}_{N+2L}\left(\mathbf{g}_1^{(m)}\right) \begin{bmatrix} \alpha_m \mathbf{r}_{k-1,isi}^{(m)} \\ \alpha_m \mathbf{r}_k^{(m)} \end{bmatrix} + \mathbf{n}_{k,t}, \qquad (35)$$

where α_m is the amplification scalar in the relay node $\mathbb{R}_m$. Combining (34) with (35), the received vector at $\mathbb{T}_1$ continues to have the form given in (9), but now the cascaded channels are $\mathbf{h}_1 = \sum_{m=1}^{M} \alpha_m \left(\mathbf{g}_1^{(m)} * \mathbf{f}_1^{(m)}\right)$ and $\mathbf{h}_2 = \sum_{m=1}^{M} \alpha_m \left(\mathbf{g}_1^{(m)} * \mathbf{f}_2^{(m)}\right)$, and the equivalent noise $\mathbf{n}_{k,e}$ becomes

$$\mathbf{n}_{k,e} = \sum_{m=1}^{M} \alpha_m \mathbf{T}_{N+2L}\left(\mathbf{g}_1^{(m)}\right) \begin{bmatrix} n_{k-1,r}^{(m)}(N+L) \\ \vdots \\ n_{k-1,r}^{(m)}(N+2L-1) \\ \mathbf{n}_{k,r}^{(m)} \end{bmatrix} + \mathbf{n}_{k,t}.$$

Hence, the above methods can be applied to the case of multiple relay nodes.

3.7 The case of ZP-OFDM systems

The proposed method can be also applied to TWRN ZP-OFDM system. In this case, $2L$ zeros are padded at the end of $\mathbf{x}_k^{(i)}$ in (3) instead of adding the cyclic prefix of length

$2L$. Due to the padded zeros, the received vector does not suffer from ISI. Therefore, (9) can be rewritten as

$$\mathbf{y}_k = \tilde{\mathbf{T}}_N(\mathbf{h}_1)\mathbf{x}_k^{(1)} + \tilde{\mathbf{T}}_N(\mathbf{h}_2)\mathbf{x}_k^{(2)} + \mathbf{n}_{k,e}. \qquad (36)$$

To estimate $\mathbf{h}_1$, we modify the cost function in (12) as

$$J\left(\hat{\mathbf{h}}_1\right) = \mathrm{E}\left\{\left\|\mathbf{y}_k - \tilde{\mathbf{T}}_N\left(\hat{\mathbf{h}}_1\right)\mathbf{x}_k^{(1)}\right\|_F^2\right\}. \qquad (37)$$

Following a procedure similar to (12)–(20), an estimate of $\mathbf{h}_1$ can be obtained by

$$\hat{\mathbf{h}}_1 = \frac{1}{\frac{N}{\sqrt{N+2L}}K\sigma_1^2} \tilde{\mathbf{W}}_{2L+1}^{\dagger} \sum_{k=0}^{K-1} \left(\tilde{\mathbf{W}}_N \mathbf{x}_k^{(1)}\right)^* \odot \left(\tilde{\mathbf{W}}\mathbf{y}_k\right), \ (38)$$

where $\tilde{\mathbf{W}}$ is the $(N+2L)\times(N+2L)$ normalized DFT matrix with the (m, n)th entry given by $\frac{1}{\sqrt{N+2L}}e^{-J2\pi mn/(N+2L)}$, whereas $\tilde{\mathbf{W}}_{2L+1}$ and $\tilde{\mathbf{W}}_N$ are respectively the first $2L+1$ and N columns of $\tilde{\mathbf{W}}$.

Assume that the estimation of $\mathbf{h}_1$ is perfect so that we can eliminate the interference from $\mathbb{T}_1$. Similar to (21), define

$$\mathbf{z}_k = \mathbf{y}_k - \tilde{\mathbf{T}}_N\left(\hat{\mathbf{h}}_1\right)\mathbf{x}_k^{(1)}. \qquad (39)$$

Substituting (36) into (39), we have

$$\mathbf{z}_k = \tilde{\mathbf{T}}_N(\mathbf{h}_2)\mathbf{x}_k^{(2)} + \mathbf{n}_{k,e}. \qquad (40)$$

This form is similar to (25), so we can follow the procedure in (25)–(31) to estimate $\mathbf{h}_2$. Note that the noise $\mathbf{n}_{k,e}$ is (almost) white. Similar to the previous discussion, a necessary condition is $K \geq N$. To reduce the limitation of a large K, we exploit the repetition method in [26]. That is, we utilize $\tilde{\mathbf{T}}_Q(\mathbf{z}_k)$ instead of $\mathbf{z}_k$ to estimate $\mathbf{h}_2$, and the necessary condition becomes $K \geq \frac{N-1}{Q} + 1$. In this case, the noise term $\tilde{\mathbf{T}}_Q(\mathbf{n}_{k,e})$ is colored (though $\mathbf{n}_{k,e}$ is white) and the covariance matrix is [26]
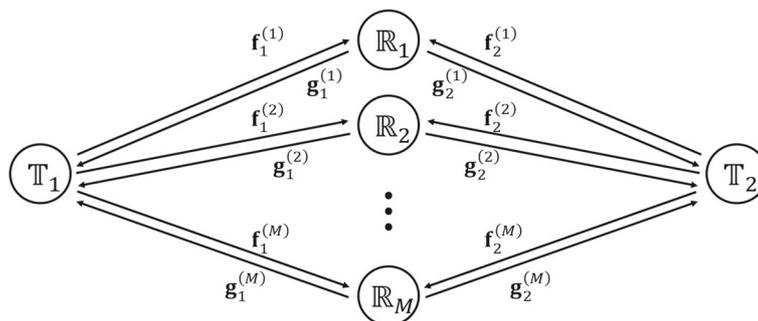
Fig. 2 System configuration for multiple relay network. It shows the case of multiple relay nodes. Suppose that we have M relay nodes $\mathbb{R}_1, \mathbb{R}_2, \ldots, \mathbb{R}_M$. Let the channels from $\mathbb{T}_i$ to $\mathbb{R}_m$ be denoted as $\mathbf{f}_i^{(m)}$ and the channels from $\mathbb{R}_m$ to $\mathbb{T}_i$ be denoted as $\mathbf{g}_i^{(m)}$

$$\mathrm{E}\left\{\tilde{\mathbf{T}}_Q\left(\mathbf{n}_{k,e}\right)\tilde{\mathbf{T}}_Q^\dagger\left(\mathbf{n}_{k,e}\right)\right\}=\sigma_{n_e}^2\mathbf{D}\left(\left[1,2,\ldots,\underbrace{Q',\ldots,Q'}_{|N+2L-Q|+1},\ldots,2,1\right]\right),$$

where $\mathbf{D}(\cdot)$ is defined in (15) and $Q'=\min\{Q,N+2L\}$. Therefore, we need to whiten the matrix $\tilde{\mathbf{T}}_Q(\mathbf{z}_k)$ by

$$\mathbf{D}\left(\left[1,\frac{1}{\sqrt{2}},\ldots,\underbrace{\frac{1}{\sqrt{Q'}},\ldots,\frac{1}{\sqrt{Q'}}}_{|N+2L-Q|+1},\ldots,\frac{1}{\sqrt{2}},1\right]\right).$$

Following a procedure similar to Section 3.5, one can obtain a blind estimate of $\mathbf{h}_2$ (up to a scalar ambiguity).

4 Analysis of MSE performance and Cramer-Rao bound

In this section, we will derive the theoretical MSE about channel estimation for $\mathbf{h}_1$ and $\mathbf{h}_2$ respectively. In the following analysis, we assume that the channel taps are uncorrelated and the transmitted vectors $\mathbf{x}_k^{(i)}$ are also uncorrelated for different k or i.

4.1 The analysis of $\mathbf{h}_1$ estimate

In the estimation of $\mathbf{h}_1$, we regard the signal from $\mathbb{T}_2$ as interference. Since (20) is the least squares solution of (18), the difference between $\hat{\mathbf{h}}_1$ and $\mathbf{h}_1$ can be calculated as

$$\Delta\mathbf{h}_1\triangleq\hat{\mathbf{h}}_1-\mathbf{h}_1$$
$$=\frac{1}{\sqrt{N}K\sigma_1^2}\mathbf{W}_{2L+1}^\dagger\sum_{k=0}^{K-1}\left(\mathbf{s}_k^{(1)}\right)^*\odot\left(\mathbf{W}\boldsymbol{\xi}_k\right),\quad(41)$$

where $\boldsymbol{\xi}_k$ denotes the interference and noise. From (11), we have

$$\boldsymbol{\xi}_k=\mathbf{C}(\mathbf{h}_2)\mathbf{x}_k^{(2)}+\bar{\mathbf{n}}_{k,e},\qquad(42)$$

where $\mathbf{C}(\mathbf{h}_2)$ is an $N\times N$ circulant matrix having $\left[\mathbf{h}_2^T\ \mathbf{0}_{1\times(N-2L-1)}\right]^T$ as its first column. Assuming that $\mathbf{x}_k^{(2)}$ and $\mathbf{n}_{k,e}$ are uncorrelated, the covariance matrix of $\boldsymbol{\xi}_k$ can be computed as

$$\mathrm{E}\left\{\boldsymbol{\xi}_k\boldsymbol{\xi}_k^\dagger\right\}=\sigma_2^2\mathbf{C}\left(\mathbf{h}_2\right)\mathbf{C}^\dagger\left(\mathbf{h}_2\right)+\sigma_{n_e}^2\mathbf{I}_N$$
$$=\mathbf{W}^\dagger\left(\sigma_2^2\mathbf{D}\left(\mathbf{h}_{2f}\right)\mathbf{D}^\dagger\left(\mathbf{h}_{2f}\right)+\sigma_{n_e}^2\mathbf{I}_N\right)\mathbf{W},\quad(43)$$

where $\mathbf{D}\left(\mathbf{h}_{2f}\right)$ is the $N\times N$ diagonal matrix with diagonal entries from the $N\times1$ frequency response vector $\mathbf{h}_{2f}=\sqrt{N}\mathbf{W}_{2L+1}\mathbf{h}_2$. Then, the covariance matrix of $\Delta\mathbf{h}_1$ can be computed as

$$\mathbf{R}_{\Delta\mathbf{h}_1}\triangleq\mathrm{E}\left\{\Delta\mathbf{h}_1\Delta\mathbf{h}_1^\dagger\right\}$$
$$=\frac{1}{NK\sigma_1^2}\mathbf{W}_{2L+1}^\dagger\left(\sigma_2^2\mathbf{D}\left(\mathbf{h}_{2f}\right)\mathbf{D}^\dagger\left(\mathbf{h}_{2f}\right)+\sigma_{n_e}^2\mathbf{I}_N\right)\mathbf{W}_{2L+1}\quad(44)$$
$$=\frac{1}{NK\sigma_1^2}\left(\sigma_2^2\mathbf{A}+\sigma_{n_e}^2\mathbf{I}_{2L+1}\right),$$

where $\mathbf{A}$ is a $(2L+1)\times(2L+1)$ Toeplitz and Hermitian matrix with $A_{m,n}=\sum_{l=0}^{2L+m-n}h_{2,l}h_{2,l-m+n}^*$ if $m\le n$ and $A_{m,n}=\sum_{l=m-n}^{2L}h_{2,l}h_{2,l-m+n}^*$ if $m\ge n$. Therefore, the theoretical MSE can be calculated as

$$\mathrm{E}\left\{\|\Delta\mathbf{h}_1\|_F^2\right\}=tr\left\{\mathbf{R}_{\Delta\mathbf{h}_1}\right\}=\frac{2L+1}{NK}\frac{\sigma_2^2\|\mathbf{h}_2\|_F^2+\sigma_{n_e}^2}{\sigma_1^2},\quad(45)$$

where $tr\left\{\mathbf{R}_{\Delta\mathbf{h}_1}\right\}$ is the sum of the diagonal elements of $\mathbf{R}_{\Delta\mathbf{h}_1}$. Define the signal-to-noise ratio (SNR) as

$$\mathrm{SNR}\triangleq\frac{\sigma_2^2}{\sigma_{n_e}^2}=\frac{\sigma_2^2}{\alpha^2\sigma_{n_r}^2\|\mathbf{g}_1\|_F^2+\sigma_{n_t}^2},\qquad(46)$$

where the second equality is obtained by using (10). Then, (45) can be written as

$$\mathrm{E}\left\{\|\Delta\mathbf{h}_1\|_F^2\right\}=\frac{2L+1}{NK}\frac{\sigma_2^2}{\sigma_1^2}\left(\|\mathbf{h}_2\|_F^2+\frac{1}{\mathrm{SNR}}\right).\quad(47)$$

Note from the above equation that the MSE is proportional to the signal power from $\mathbb{T}_2$ but inversely proportional to the signal power from $\mathbb{T}_1$ and the number of the received signal blocks. Moreover, for high SNR, the MSE floors at the value of $\frac{2L+1}{NK}\frac{\sigma_2^2}{\sigma_1^2}\|\mathbf{h}_2\|_F^2$.

4.2 The analysis of $\mathbf{h}_2$ estimate

During the estimation of $\mathbf{h}_2$ in Section 3.2, it is assumed that the estimate of $\mathbf{h}_1$ is perfect. However, the estimation error $\Delta\mathbf{h}_1$ will affect the accuracy of the estimation of $\mathbf{h}_2$. From (21), if $\Delta\mathbf{h}_1\ne0$, the interference and noise terms can be written as

$$\mathbf{T}_{N+2L}(\Delta\mathbf{h}_1)\begin{bmatrix}\mathbf{x}_{k-1,cp}^{(1)}\\\mathbf{x}_{k,cp}^{(1)}\\\mathbf{x}_k^{(1)}\end{bmatrix}+\mathbf{n}_{k,e}.\quad(48)$$

Next, we look at $\mathbf{v}_k$ in (25). Following the procedure (21)–(26), the whitened vector $\mathbf{v}_k^{(w)}$ now becomes

$$\mathbf{v}_k^{(w)}=\mathbf{R}_w^{-1/2}\mathbf{v}_k=\mathbf{R}_w^{-1/2}\tilde{\mathbf{T}}_N(\mathbf{h}_2)\mathbf{d}_k+\boldsymbol{\zeta}_k,$$

where

$$\boldsymbol{\zeta}_k=\mathbf{R}_w^{-1/2}\tilde{\mathbf{T}}_N(\Delta\mathbf{h}_1)1\left(\begin{bmatrix}x_{k,cp}^{(1)}\\x_{k,0}^{(1)}\\\vdots\\x_{k,N-2L-1}^{(1)}\end{bmatrix}-\mathbf{x}_{k-1}^{(1)}\right)+\mathbf{R}_w^{-1/2}\boldsymbol{\eta}_k.$$
$$(49)$$

Recall from the subspace method in Section 3.2 that the estimate $\hat{\mathbf{h}}_2$ is obtained from the noise subspace $\mathbf{U}_o$ in (28).

Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{N+2L}$ be the eigenvalues of $\mathbf{R}_\nu^{(w)}$. The noise subspace $\mathbf{U}_o$ is the eigenspace corresponding to the smallest $2L$ eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{2L}$. The error vector $\boldsymbol{\zeta}_k$ can cause two effects: (i) it perturbs the noise subspace $\mathbf{U}_o$ and (ii) it also perturbs the eigenvalues, i.e., $\lambda_1 + \Delta\lambda_1, \lambda_2 + \Delta\lambda_2, \ldots, \lambda_{N+2L} + \Delta\lambda_{N+2L}$. Note that λ_{2L} belongs to the noise subspace $\mathbf{U}_o$ and λ_{2L+1} belongs to the signal subspace $\mathbf{U}_s$. Their difference $\lambda_{2L+1} - \lambda_{2L}$ is usually large. Nevertheless, the perturbation on eigenvalues may lead to the case $\lambda_{2L} + \Delta\lambda_{2L} > \lambda_{2L+1} + \Delta\lambda_{2L+1}$, especially when the SNR is low. In this case, the noise subspace will be polluted by the signal subspace and this will cause a large error in the estimation of $\mathbf{h}_2$. Below, we derive the MSE by studying the following two cases separately.

Case I: $\lambda_{2L} + \Delta\lambda_{2L} < \lambda_{2L+1} + \Delta\lambda_{2L+1}$

In this case, we can exploit the first-order approximation of the perturbation to $\mathbf{U}_o$. In [24], the channel estimation error has been derived. However, the theoretical MSE derived in [24] is based on white noise. As the noise $\boldsymbol{\zeta}_k$ is colored, the formula derived in [24] is not applicable. For the case of colored noise $\boldsymbol{\zeta}_k$, we have derived a new formula and the theoretical MSE of the $\mathbf{h}_2$ estimate can be calculated as

$$\mathrm{E}\left\{\|\Delta\mathbf{h}_2\|_F^2\right\} = \frac{1}{2K\sigma_2^2} tr\left\{\mathbf{U}^\sharp \left(\mathbf{I}_N \otimes \mathbf{U}_o^\dagger \mathbf{R}_\zeta \mathbf{U}_o\right)\left(\mathbf{U}^\dagger\right)^\sharp\right\}, \quad (50)$$

where $\Delta\mathbf{h}_2 \triangleq \hat{\mathbf{h}}_2 - \mathbf{h}_2$, $\mathbf{U}^\sharp$ is the Moore-Penrose pseudoinverse matrix of $\mathbf{U}$ defined in (31), and $\mathbf{R}_\zeta \triangleq \mathrm{E}\{\boldsymbol{\zeta}_k \boldsymbol{\zeta}_k^\dagger\}$ is the covariance matrix of $\boldsymbol{\zeta}_k$ and it can be written as

$$\mathbf{R}_\zeta = 2\sigma_1^2 \mathbf{R}_w^{-1/2} \mathrm{E}\left\{\tilde{\mathbf{T}}_N(\Delta\mathbf{h}_1)\tilde{\mathbf{T}}_N^\dagger(\Delta\mathbf{h}_1)\right\} \mathbf{R}_w^{-1/2} + \sigma_{n_e}^2 \mathbf{I}_{N+2L}. \quad (51)$$

Notice that $\tilde{\mathbf{T}}_N(\Delta\mathbf{h}_1)$ can be rewritten as

$$\tilde{\mathbf{T}}_N(\Delta\mathbf{h}_1) = \begin{bmatrix} \mathbf{J}_0\Delta\mathbf{h}_1 & \mathbf{J}_1\Delta\mathbf{h}_1 & \cdots & \mathbf{J}_{N-1}\Delta\mathbf{h}_1 \end{bmatrix},$$

where $\mathbf{J}_i$ is defined in (30). Hence, (51) can be rewritten as

$$\mathbf{R}_\zeta = 2\sigma_1^2 \mathbf{R}_w^{-1/2}\left(\sum_{i=0}^{N-1} \mathbf{J}_i \mathbf{R}_{\Delta\mathbf{h}_1}\mathbf{J}_i^T\right)\mathbf{R}_w^{-1/2} + \sigma_{n_e}^2\mathbf{I}_{N+2L}, \quad (52)$$

where $\mathbf{R}_{\Delta\mathbf{h}_1}$ is defined in (44).

Case II: $\lambda_{2L} + \Delta\lambda_{2L} \geq \lambda_{2L+1} + \Delta\lambda_{2L+1}$

In this case, our algorithm cannot find the accurate noise subspace $\mathbf{U}_o$ because it has been polluted by signal subspace $\mathbf{U}_s$. Thus, we assume that the eigenvector of $\mathbf{U}$ corresponding to the smallest eigenvalue is random and uncorrelated to the true cascaded channel $\mathbf{h}_2$. Define this unit-norm eigenvector as $\tilde{\mathbf{h}}_2$. The scalar ambiguity can be calculated as $\alpha = \left(\tilde{\mathbf{h}}_2^\dagger \mathbf{h}_2\right)/\left(\tilde{\mathbf{h}}_2^\dagger\tilde{\mathbf{h}}_2\right) = \tilde{\mathbf{h}}_2^\dagger \mathbf{h}_2$. That is,

$$\hat{\mathbf{h}}_2 = \tilde{\mathbf{h}}_2\alpha = \tilde{\mathbf{h}}_2\left(\tilde{\mathbf{h}}_2^\dagger \mathbf{h}_2\right). \quad (53)$$

The estimation error is $\Delta\mathbf{h}_2 = \hat{\mathbf{h}}_2 - \mathbf{h}_2 = \left(\tilde{\mathbf{h}}_2\tilde{\mathbf{h}}_2^\dagger - \mathbf{I}_{2L+1}\right)\mathbf{h}_2$. Hence, the theoretical MSE of the $\mathbf{h}_2$ estimate can be calculated as

$$\mathrm{E}\left\{\|\Delta\mathbf{h}_2\|_F^2\right\} = \mathbf{h}_2^\dagger \mathrm{E}\left\{\left(\tilde{\mathbf{h}}_2\tilde{\mathbf{h}}_2^\dagger - \mathbf{I}_{2L+1}\right)^2\right\}\mathbf{h}_2. \quad (54)$$

Since the unit-norm vector $\tilde{\mathbf{h}}_2$ is assumed to be random, $\mathrm{E}\left\{\tilde{\mathbf{h}}_2\tilde{\mathbf{h}}_2^\dagger\right\}$ can be approximated as $\frac{1}{2L+1}\mathbf{I}_{2L+1}$. Therefore, (54) can be written as

$$\mathrm{E}\left\{\|\Delta\mathbf{h}_2\|_F^2\right\} = \frac{2L}{2L+1}\|\mathbf{h}_2\|_F^2. \quad (55)$$

Overall MSE: Utilizing Bayes' theorem, the theoretical MSE of the $\mathbf{h}_2$ estimate can be written as

$$\begin{aligned} \mathrm{E}\left\{\|\Delta\mathbf{h}_2\|_F^2\right\} = &\ P_{err}\mathrm{E}\{\Delta\mathbf{h}_2\|_F^2\big| \lambda_{2L} + \Delta\lambda_{2L} \\ &\geq \lambda_{2L+1} + \Delta\lambda_{2L+1}\} \\ &+ (1 - P_{err})\mathrm{E}\{\|\Delta\mathbf{h}_2\|_F^2\big| \lambda_{2L} + \Delta\lambda_{2L} \\ &< \lambda_{2L+1} + \Delta\lambda_{2L+1}\}, \end{aligned} \quad (56)$$

where P_{err} is the probability of $\lambda_{2L} + \Delta\lambda_{2L} \geq \lambda_{2L+1} + \Delta\lambda_{2L+1}$ and it can be expressed by

$$P_{err} = Q\left(\sqrt{\frac{K}{2}}\frac{\lambda_{2L+1} - \lambda_{2L}}{\sigma_{n_e}^2}\right), \quad (57)$$

where $Q(\cdot)$ is the Q-function:

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{u^2}{2}}\, du.$$

The derivation of P_{err} is given in Appendix A. Substituting (50) and (55) into (56), the theoretical MSE of the $\mathbf{h}_2$ estimate can be represented as

$$\begin{aligned} \mathrm{E}\left\{\|\Delta\mathbf{h}_2\|_F^2\right\} = &\ P_{err}\frac{2L}{2L+1}\|\mathbf{h}_2\|_F^2 \\ &+ (1 - P_{err})\frac{1}{2K\sigma_2^2}tr\left\{\mathbf{U}^\sharp\left(\mathbf{I}_N \otimes \mathbf{U}_o^\dagger\mathbf{R}_\zeta\mathbf{U}_o\right)\left(\mathbf{U}^\dagger\right)^\sharp\right\}, \end{aligned} \quad (58)$$

where $\mathbf{R}_\zeta$ is given in (52).

4.3 Approximated Cramer-Rao bound

When we estimate $\mathbf{h}_1$, the signal from $\mathbb{T}_2$ can be viewed as interference, and the signal from $\mathbb{T}_1$ can be seen as pilot. To simplify the derivation, we assume that $\boldsymbol{\xi}_k$ in (42) is white. Hence, an ACRB of $\mathbf{h}_1$ estimation is [36]

$$\mathrm{ACRB}_1 = \frac{\sigma_\xi^2}{N}tr\left\{\left((\mathbf{SW}_{2L+1})^\dagger \mathbf{SW}_{2L+1}\right)^{-1}\right\}, \quad (59)$$

where $\mathbf{W}_{2L+1}$ and $\mathbf{S}$ are defined in (15) and (17), respectively, and σ_ξ^2 is the average power of $\boldsymbol{\xi}_k$. From (43),

we have $\sigma_\xi^2 = \sigma_2^2 \|\mathbf{h}_2\|_F^2 + \sigma_{n_e}^2$. Therefore, (59) can be simplified as

$$\text{ACRB}_1 = \frac{2L+1}{NK} \frac{\sigma_2^2 \|\mathbf{h}_2\|_F^2 + \sigma_{n_e}^2}{\sigma_1^2}. \tag{60}$$

Notice that this form is the same as (45).

Next, we consider the ACRB of $\mathbf{h}_2$. In [24], the authors have derived an ACRB and concluded that the ACRB is the same as the channel estimation MSE. Hence, from (50), an ACRB of $\mathbf{h}_2$ estimation is

$$\text{ACRB}_2 = \frac{1}{2K\sigma_2^2} tr \left\{ \mathbf{U}^\sharp \left(\mathbf{I}_N \otimes \mathbf{U}_o^\dagger \mathbf{R}_\zeta \mathbf{U}_o \right) \left(\mathbf{U}^\dagger \right)^\sharp \right\}. \tag{61}$$

In the derivations of the ACRBs, the noises are assumed to be white even though they are actually colored. Therefore, the ACRBs in (60) and (61) are in general larger than or equal to the true Cramer-Rao bounds.

5 Simulation results

In the simulation, we consider a TWRN with one relay node. The channel taps $f_{i,l}$ and $g_{i,l}$ are generated as independent and identically distributed zero-mean complex Gaussian random variables with variances equal to 1/9. The order of these channels is $L = 8$, so the order of the cascaded channels is $2L = 16$. The channels are normalized so that $\|\mathbf{f}_1\|_F^2 = \|\mathbf{f}_2\|_F^2 = \|\mathbf{g}_1\|_F^2 = \|\mathbf{g}_2\|_F^2 = 1$. The channel does not change while the channel estimation is performed. The channel noise is additive white

Gaussian noise (AWGN), and the transmission symbols are 16-QAM with gray code. The size of the DFT matrix is $N = 64$, and the length of CP is $2L = 16$. In all plots, we set $\sigma_1^2 = \sigma_2^2$ and $\sigma_{n_r}^2 = \sigma_{n_t}^2$. The SNR is defined in (46), and the normalized MSE is defined as

$$\frac{1}{M_c} \sum_{m=1}^{M_c} \frac{\|\hat{\mathbf{h}}_i^{(m)} - \mathbf{h}_i\|_F^2}{\|\mathbf{h}_i\|_F^2} \text{ for } i = 1 \text{ and } 2,$$

where $\hat{\mathbf{h}}_i^{(m)}$ represents the estimated $\mathbf{h}_i$ in the mth trial. $M_c = 2000$ denotes the total number of Monte-Carlo trials.

First, we look at the MSE performance of the proposed methods. The number of received blocks is $K = 500$. In Fig. 3, we plot the normalized MSEs for $\mathbf{h}_1$ and $\mathbf{h}_2$. The "simulation" curves of $\mathbf{h}_1$ and $\mathbf{h}_2$ are obtained by (20) and (31), respectively, whereas the "theory" curves of $\mathbf{h}_1$ and $\mathbf{h}_2$ are calculated by (47) and (58), respectively. Moreover, we also display the ACRBs of $\mathbf{h}_1$ and $\mathbf{h}_2$ according to (60) and (61). From Fig. 3, it can be seen that the simulated result, the theoretical MSE, and the ACRB of $\mathbf{h}_1$ is close. Moreover, the proposed method can give a good estimate of $\mathbf{h}_1$, even at very low SNR of 0 dB. One can see that the MSE floors at $\frac{2L+1}{NK} = 5.3 \times 10^{-4}$ at high SNR, and this confirms our analysis in (47). For $\mathbf{h}_2$, the MSE performance is worse than that of $\mathbf{h}_1$ for SNR < 25 dB, but the MSE of $\mathbf{h}_2$ floors at a much smaller value of 1.2×10^{-5}. This flooring happens at very high SNR, and the estimation error of $\mathbf{h}_1$ affects the accuracy of $\mathbf{h}_2$ estimate. The
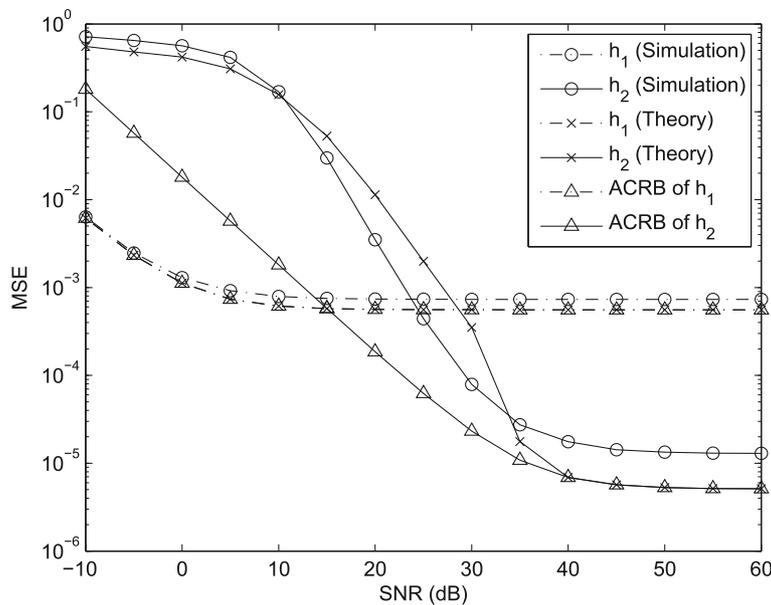

Fig. 3 Comparison of the numerical and theoretical normalized MSE for $K = 500$. We plot the normalized MSEs for $\mathbf{h}_1$ and $\mathbf{h}_2$. The "simulation" curves of $\mathbf{h}_1$ and $\mathbf{h}_2$ are obtained by (20) and (31), respectively, whereas the "Theory" curves of $\mathbf{h}_1$ and $\mathbf{h}_2$ are calculated by (47) and (58), respectively. Moreover, we also display the ACRBs of $\mathbf{h}_1$ and $\mathbf{h}_2$ according to (60) and (61). The number of received blocks is $K = 500$

gap between numerical and theoretical results is small at high SNR. At low SNR, the gap between simulation result and ACRB becomes very large, but the theoretical curve is still close to the numerical curve. Recall that the theoretical MSE value is a combination of two cases in Section 4.2, and the ACRB of $\mathbf{h}_2$ in (61) is equal to the theoretical MSE when we do not consider the perturbation on eigenvalues, i.e., case II in Section 4.2 (case II usually happens at low SNR). Therefore, the difference between the theoretical MSE and ACRB of $\mathbf{h}_2$ at low SNR is caused by the perturbation of eigenvalues. From Fig. 3, we conclude that the change of eigenvalue sequence dominates the performance degradation at low SNR. In addition, the assumption of white noise in the derivation of ACRB also affects the accuracy, especially when the SNR is low.
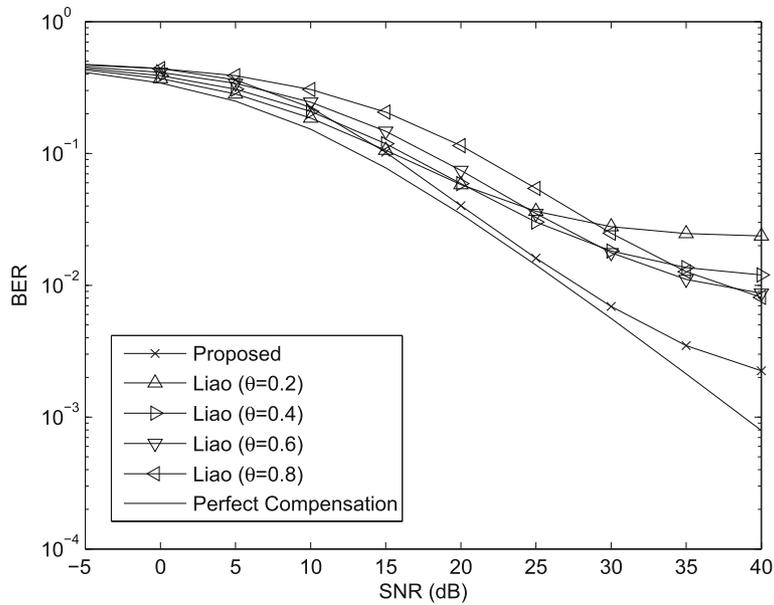
Next, we compare the performances of our method with the method proposed by Liao et al. in [34]. As mentioned in Section 3.4, Liao's algorithm has a compromise between channel estimation error and BER. The parameter θ in Liao's algorithm is set to 0.2, 0.4, 0.6, and 0.8. From [34], it is found that $\theta = 0.4$ yields a good BER performance when SNR = 25 dB. In Figs. 4 and 5, the number of received blocks is $K = 500$. Figure 4 shows the MSE performances. Since the MSEs of $\mathbf{h}_1$ and $\mathbf{h}_2$ by Liao's algorithm are the same, we plot one MSE curve only. From the figure, we see that as θ increases from 0.2 to 0.8, the MSE of Liao's algorithm decreases. For the estimation of $\mathbf{h}_1$, our method is better than Liao's methods for $\theta = 0.2$ and 0.4, but worse than that for $\theta = 0.6$ and $\theta = 0.8$. As we will see in Fig. 5, the BER performance

for $\theta = 0.8$ is not good due to severe noise amplification. For $\mathbf{h}_2$, Liao's method is better at low SNR whereas our method is better at high SNR. In Fig. 5, we show BER performances. Zero-forcing equalizers are used at the receiver. The "perfect compensation" represents the case that the channel taps are perfectly known at the receiver. It is seen that among the four curves of $\theta = 0.2, 0.4, 0.6, 0.8$, Liao's method has the best BER performance when θ is set as 0.4 for SNR = 25 dB. Though the MSE of Liao's method is the smallest when $\theta = 0.8$, its BER performance is not good due to the noise amplification problem of the precoding matrix. These results are matched with [34]. From Fig. 5, we see that the proposed algorithm outperforms Liao's methods when SNR $\geq$ 15 dB, and the performance of our method is close to the perfect compensation.

Figures 6 and 7 show the simulation results when the number of blocks is $K = 50$. In this case, $K < N$, and thus, the estimation of $\mathbf{h}_2$ by (31) does not work. We exploit the repetition method discussed in Section 3.5 to solve this issue. We set the repetition parameter $Q = 10$, and the necessary condition $K \geq \frac{N-1}{Q} + 1$ is satisfied. In Fig. 6, the MSE performance is shown. We can observe that the repetition method is extremely useful when the terminal receives few blocks. On the other hand, $\mathbf{h}_1$ estimation by (20) and Liao's algorithm are based on the power reduction, so there is no limitation on the number of blocks K. From the figure, we see that the proposed method outperforms Liao's algorithms with $\theta = 0.2$ and 0.4 for all SNR. In Fig. 7, the performance is measured by
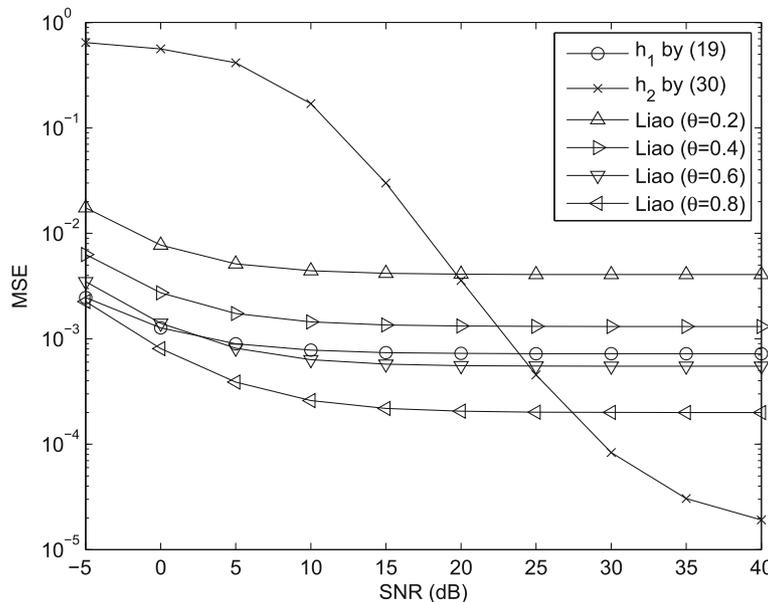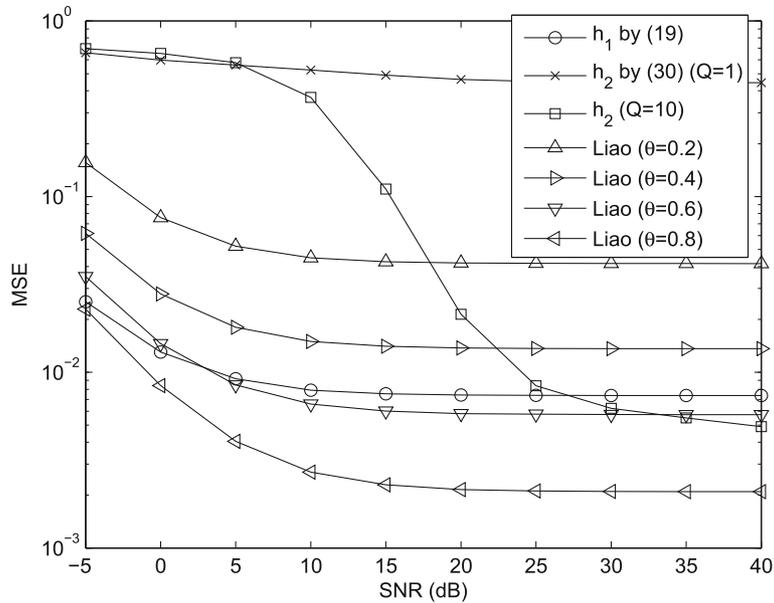

Fig. 4 Comparison of the normalized MSE for $K = 500$. We compare the normalized MSE performances of our method with the method proposed by Liao et al. in [34]. The parameter θ in Liao's algorithm is set to 0.2, 0.4, 0.6, and 0.8. The number of received blocks is $K = 500$
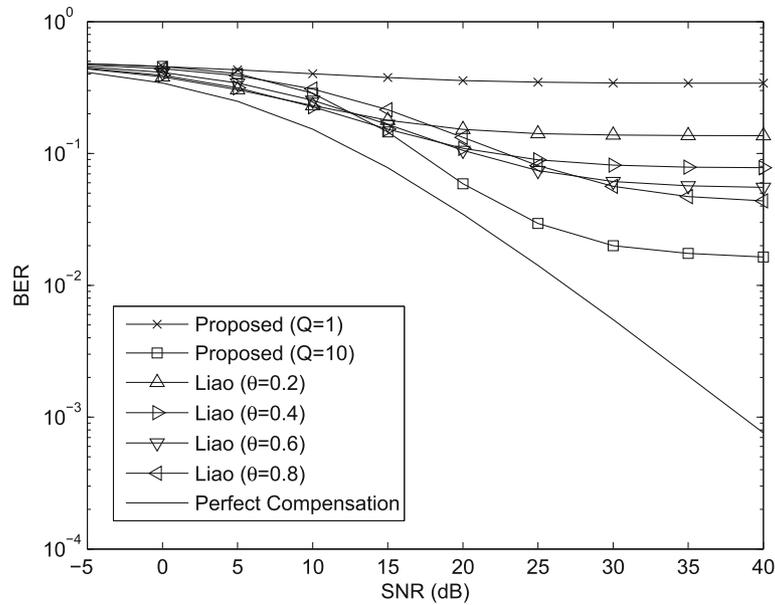
Fig. 5 Comparison of the BER for $K = 500$. We compare the BER performances of our method with the method proposed by Liao et al. in [34]. The parameter θ in Liao's algorithm is set to 0.2, 0.4, 0.6, and 0.8. The number of received blocks is $K = 500$

BER. It can be seen that the proposed algorithm performs better than Liao's method when SNR $\geq$ 15 dB. Comparing Fig. 7 with Fig. 5, we find that the BER performance degrades when K reduces from 500 to 50. This is due to the larger channel estimation errors for $K = 50$ and imperfect interference cancelation by $\mathbf{h}_1$ using (21).

Finally, we compare the proposed algorithm for CP-OFDM and ZP-OFDM systems. In Fig. 8, the solid curves and the dashed curves represent the MSEs for CP-OFDM and ZP-OFDM, respectively. We can find that the performances are almost the same. In other words, our method works well for both CP and ZP systems.

Fig. 6 Comparison of the normalized MSE for $K = 50$. We compare the normalized MSE performances of our method with the method proposed by Liao et al. in [34]. The parameter θ in Liao's algorithm is set to 0.2, 0.4, 0.6, and 0.8. The number of received blocks is $K = 50$
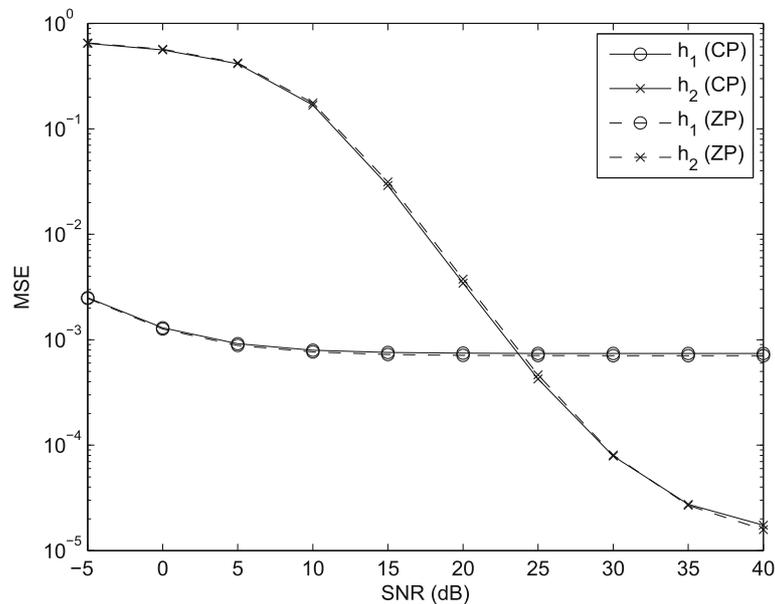
Fig. 7 Comparison of the BER for $K = 50$. We compare the BER performances of our method with the method proposed by Liao et al. in [34]. The parameter θ in Liao's algorithm is set to 0.2, 0.4, 0.6, and 0.8. The number of received blocks is $K = 50$

6 Conclusions

In this paper, we propose a blind channel estimation method in OFDM-based amplify-and-forward two-way relay networks. The first cascaded channel $\mathbf{h}_1$ is estimated by the power reduction method whereas the second cascaded channel $\mathbf{h}_2$ is estimated by the subspace method.

Close-form formulas are derived. We also analyze the theoretical performance and derive the ACRBs for channel estimation. Our algorithm can be applied to both CP-OFDM and ZP-OFDM systems, and it can use repetition method to handle the case of few received blocks. Simulation results verify our analysis.

Fig. 8 Comparison of the normalized MSE for CP-OFDM and ZP-OFDM systems for $K = 500$. We compare the proposed algorithm for CP-OFDM and ZP-OFDM systems. The solid curves and the dashed curves represent the MSEs for CP-OFDM and ZP-OFDM, respectively. The number of received blocks is $K = 500$

Endnotes

^{1} The proposed method can be applied to the more general case of different channel lengths by simply using an appropriate cyclic prefix length.

^{2} If $\mathbb{T}_1$ and $\mathbb{T}_2$ add CP of length L, then the relay needs to carry out the operations of OFDM symbol timing synchronization, CP removal, and CP insertion. In order to simplify the tasks of the relay, $\mathbb{T}_1$ and $\mathbb{T}_2$ add CP of length $2L$.

Appendix A

A proof of (57)

To simplify our derivation, we utilize the fact that this condition usually occurs at low SNR. From (47) and the simulation in Section 5, it can be seen that the estimate of $\mathbf{h}_1$ is still quite accurate at low SNR, so the second term $\mathbf{n}_{k,e}$ in (48) is dominant. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{N+2L}$ be the eigenvalues of $\mathbf{R}_v^{(w)}$ and the corresponding unit-norm eigenvectors are respectively $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{N+2L}$. By (27) and (52), $\mathbf{R}_v^{(w)}$ can be expressed by

$$
\begin{aligned}
\mathbf{R}_v^{(w)} = {} & 2\sigma_2^2 \mathbf{R}_w^{-1/2} \tilde{\mathbf{T}}_N(\mathbf{h}_2) \tilde{\mathbf{T}}_N^\dagger(\mathbf{h}_2) \mathbf{R}_w^{-1/2} \\
& + 2\sigma_1^2 \mathbf{R}_w^{-1/2} \left(\sum_{i=0}^{N-1} \mathbf{J}_i \mathbf{R}_{\Delta \mathbf{h}_1} \mathbf{J}_i^T \right) \mathbf{R}_w^{-1/2} + \sigma_{n_e}^2 \mathbf{I}_{N+2L}.
\end{aligned}
$$

(62)

Since the received signals are finite and the second term in (48) is dominant, we have the following approximation:

$$
\begin{aligned}
\frac{1}{K} \sum_{k=1}^K \mathbf{v}_k^{(w)} \left(\mathbf{v}_k^{(w)} \right)^\dagger \approx {} & E\{ \mathbf{R}_w^{-1/2} \mathbf{v}_k \mathbf{v}_k^\dagger \mathbf{R}_w^{-1/2} \} \\
& + \underbrace{ \frac{1}{K} \sum_{k=1}^K \mathbf{R}_w^{-1/2} \boldsymbol{\eta}_k \boldsymbol{\eta}_k^\dagger \mathbf{R}_w^{-1/2} - \sigma_{n_e}^2 \mathbf{I}_{N+2L} }_{\triangleq \mathbf{N}},
\end{aligned}
$$

(63)

and the corresponding eigenvalues become $\lambda_1 + \Delta\lambda_1, \lambda_2 + \Delta\lambda_2, \ldots, \lambda_{N+2L} + \Delta\lambda_{N+2L}$. Notice that $\mathbf{N}$ is a Hermitian matrix with mean $\mathbf{0}$. For large K, the central limit theorem indicates that the diagonal entries of $\mathbf{N}$ are real normal distributed and the other entries are circularly symmetric complex normal distributed [37]. According to the result in [38], all entries of $\mathbf{N}$ have the same variance $\frac{1}{K}\sigma_{n_e}^4$.

From matrix theory [39], the eigenvalue perturbation $\Delta\lambda_i$ can be approximated as $\mathbf{b}_i^\dagger \mathbf{N} \mathbf{b}_i$. Then, the mean is

$$
E\{\Delta\lambda_i\} = E\left\{ \mathbf{b}_i^\dagger \mathbf{N} \mathbf{b}_i \right\} = \mathbf{b}_i^\dagger E\{\mathbf{N}\} \mathbf{b}_i = 0,
$$

(64)

and the variance is

$$
\begin{aligned}
E\left\{ |\Delta\lambda_i|^2 \right\} = {} & E\left\{ \mathbf{b}_i^\dagger \mathbf{N} \mathbf{b}_i \mathbf{b}_i^\dagger \mathbf{N}^\dagger \mathbf{b}_i \right\} \\
= {} & E\left\{ \sum_j \sum_l \sum_m \sum_n b_i^*(j) N(j,l) b_i(l) b_i(m) N^*(m,n) b_i^*(n) \right\} \\
= {} & \sum_j \sum_l \sum_m \sum_n b_i^*(j) b_i(l) b_i(m) b_i^*(n) E\{N(j,l) N^*(m,n)\}.
\end{aligned}
$$

(65)

Because $\mathbf{R}_w^{-1/2} \boldsymbol{\eta}_k$ is white, all entries of $\mathbf{N}$ are uncorrelated, so the last term $E\left\{ N(j,l) N^*(m,n) \right\}$ is equal to $\frac{1}{K}\sigma_{n_e}^4 \delta(j-m)\delta(l-n)$, where $\delta(\cdot)$ is the Kronecker delta function. Thus, (65) can be rewritten as

$$
\begin{aligned}
E\left\{ |\Delta\lambda_i|^2 \right\} = {} & \frac{1}{K}\sigma_{n_e}^4 \sum_j \sum_l b_i^*(j) b_i(l) b_i(j) b_i^*(l) \\
= {} & \frac{1}{K}\sigma_{n_e}^4 \left| \mathbf{b}_i^\dagger \mathbf{b}_i \right|^2 = \frac{1}{K}\sigma_{n_e}^4.
\end{aligned}
$$

(66)

The last equality holds since $\mathbf{b}_i^\dagger \mathbf{b}_i = 1$ for $i = 1, 2, \ldots, N+2L$.

Notice that $\mathbf{N}$ is normal distributed and $\mathbf{b}_i$ is constant, so the random variable $\Delta\lambda_i$ is normal distributed as well. It means that the probability of $\lambda_{2L} + \Delta\lambda_{2L} \geq \lambda_{2L+1} + \Delta\lambda_{2L+1}$ can be computed as

$$
\begin{aligned}
P_{err} &\triangleq \Pr\{\lambda_{2L} + \Delta\lambda_{2L} \geq \lambda_{2L+1} + \Delta\lambda_{2L+1}\} \\
&= Q\left(\frac{\lambda_{2L+1} - \lambda_{2L}}{\sqrt{E\{|\Delta\lambda_{2L+1}|^2\} + E\{|\Delta\lambda_{2L}|^2\}}} \right).
\end{aligned}
$$

(67)

Substituting (66) into (67), we obtain (57).

Abbreviations
ACRB: Approximated cramer-rao bound; AF: Amplify-and-forward; AWGN: Additive white Gaussian noise; BER: Bit error rate; CIR: Channel impulse response; CP: Cyclic prefix; CSI: Channel state information; DA: Data-aided; DF: Decode-and-forward; DFT: Discrete fourier transform; HOS: Higher order statistics; IDFT: Inverse discrete fourier transform; i.i.d.: Independent and identically distributed; ISI: Inter-symbol interference; KCS: Knowledge of the channel statistics; LS: Least squares; MAP: Maximum a posteriori; ML: Maximum likelihood; MMSE: Minimum mean square error; MSE: Mean square error; OFDM: Orthogonal frequency division multiplexing; OWRN: One-way relay network; SISO: Single-input single-output; SNR: Signal-to-noise ratio; SOS: Second-order statistics; STC: Space-time code; TWRN: Two-way relay network; VC: Virtual carrier; ZP: Zero padding

Authors' contributions
T-CL and S-MP constructed the theory. T-CL performed simulations and wrote a draft. S-MP modified the paper. Both authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. JN Laneman, DNC Tse, GW Wornell, Cooperative diversity in wireless networks: efficient protocols and outage behavior. IEEE Trans. Inf. Theory. **50**(12), 3062–3080 (2004)
2. S Katti, S Gollakota, D Katabi, Embracing wireless interference: analog network coding. Comput. Sci. Artif. Intell. Lab. Tech. Rep (2007)
3. B Rankov, A Wittneben, Spectral efficient signaling for half-duplex relay channels. Annual Conference on Signals, Systems, and Computers, 1066–1071 (2005)
4. B Rankov, A Wittneben, Achievable rate regions for the two-way relay channel. International Symposium on Information Theory (ISIT), 1668–1672 (2006)
5. P Popovski, H Yomo, Wireless network coding by amplify-and-forward for bi-directional traffic flows. IEEE Commun. Lett. **11**(1), 16–18 (2007)
6. T Cui, F Gao, T Ho, A Nallanathan, Distributed space-time coding for two-way wireless relay networks. International Conference on Communications (ICC), 3888–3892 (2008)
7. R Zhang, Y-C Liang, CC Chai, S Cui, Optimal beamforming for two-way multi-antenna relay channel with analogue network coding. IEEE J. Sel. Areas Commun. **27**(5), 699–712 (2009)
8. C Xing, S Ma, Y-C Wu, Robust joint design of linear relay precoder and destination equalizer for dual-hop amplify-and-forward MIMO relay systems. IEEE Trans. Signal Process. **58**(4), 2273–2283 (2010)
9. MW Baidas, AB MacKenzie, RM Buehrer, Network-coded bi-directional relaying for amplify-and-forward cooperative networks: a comparative study. IEEE Trans. Wirel. Commun. **12**(7), 3238–3252 (2013)
10. P Hoeher, S Kaiser, P Robertson, Two-dimensional pilot symbol aided channel estimation by Wiener filtering. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP). **3**, 1845–1848 (1997)
11. P Hoeher, S Kaiser, P Robertson, Pilot-symbol-aided channel estimation in time and frequency. IEEE Global Telecommunications Conference, 90–96 (1997)
12. Y Li, LJ Cimini, NR Sollenberger, Robust channel estimation for OFDM systems with rapid dispersive fading channels. IEEE Trans. Commun. **46**(7), 902–915 (1998)
13. O Edfors, M Sandell, Beek van de JJ, SK Wilson, PO Borjesson, OFDM channel estimation by singular value decomposition. IEEE Trans. Commun. **46**(7), 931–939 (1998)
14. O Edfors, M Sandell, Beek van de JJ, SK Wilson, PO Borjesson, Analysis of DFT-based channel estimators for OFDM. Wirel. Pers. Commun. **12**(1), 55–70 (2000)
15. M Morelli, U Mengali, A comparison of pilot-aided channel estimation methods for OFDM systems. IEEE Trans. Sig. Process. **49**(2), 3065–3073 (2001)
16. J Oliver, R Aravind, KMM Prabhu, Sparse channel estimation in OFDM systems by threshold-based pruning. IEEE Electron. Lett. **44**(13), 830–832 (2008)
17. S Rosati, GE Corazza, A Vanelli-Coralli, OFDM channel estimation based on impulse response decimation: analysis and novel algorithms. IEEE Trans. Commun. **60**(7), 1996–2008 (2012)
18. O Shalvi, E Weinstein, New criteria for blind deconvolution of non-minimum phase systems (channels). IEEE Trans. Inf. Theory. **36**, 312–321 (1990)
19. E Moulines, P Duhamel, JF Cardoso, S Mayrargue, Subspace methods for the blind identification of multichannel FIR filters. IEEE Trans. Sig. Process. **43**(2), 516–525 (1995)
20. S Zhou, GB Giannakis, Finite-alphabet based channel estimation for OFDM and related multicarrier systems. IEEE Trans. Commun. **49**(8), 1402–1414 (2001)
21. AP Petropulu, R Zhang, R Lin, Blind OFDM channel estimation through simple linear precoding. IEEE Trans. Wirel. Commun. **3**(2), 647–655 (2004)
22. C Li, S Roy, Subspaced-based blind channel estimation for OFDM by exploiting virtual carriers. IEEE Trans Wirel. Commun. **2**(1), 141–150 (2003)
23. B Su, PP Vaidyanathan, Subspace-based blind channel identification for cyclic prefix systems using few received blocks. IEEE Trans. Signal Process. **55**(10), 4979–4993 (2007)
24. F Gao, Y Zeng, A Nallanathan, T-S Ng, Robust subspace blind channel estimation for cyclic prefixed MIMO OFDM systems: algorithm, identifiability and performance analysis. IEEE J. Sel. Areas Commun. **26**(2), 378–388 (2008)
25. Y-C Pan, S-M Phoong, An improved subspace-based algorithm for blind channel identification using few received blocks. IEEE Trans. Commun. **61**(9), 3710–3720 (2013)
26. B Su, Subspace-based blind and semiblind channel estimation in OFDM systems with virtual carriers using few received symbols. International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 100–104 (2014)
27. F Gao, R Zhang, Y-C Liang, Optimal channel estimation and training design for two-way relay networks. IEEE Trans. Commun. **57**(10), 3024–3033 (2009)
28. F Gao, R Zhang, Y-C Liang, Channel estimation for OFDM modulated two-way relay networks. IEEE Trans. Signal Process. **57**(11), 4443–4455 (2009)
29. L Sanguinetti, AA D'Amico, Y Rong, A tutorial on the optimization of amplify-and-forward MIMO relay systems. IEEE J. Sel. Areas Commun. **30**(8), 1331–1346 (2012)
30. CWR Chiong, Y Rong, Y Xiang, Channel training algorithms for two-way MIMO relay systems. IEEE Trans. Signal Process. **61**(16), 3988–3998 (2013)
31. S Abdallah, IN Psaromiligkos, Blind channel estimation for amplify-and-forward two-way relay networks employing M-PSK modulation. IEEE Trans. Signal Process. **60**(7), 3604–3615 (2012)
32. Q Zhao, Z Zhou, J Li, B Vucetic, Joint semi-blind channel estimation and synchronization in two-way relay networks. IEEE Trans. Veh. Technol. **63**(7), 3276–3293 (2014)
33. X Xie, M Peng, B Zhao, W Wang, Y Hua, Maximum a posteriori based channel estimation strategy for two-way relaying channels. IEEE Trans. Wirel. Commun. **13**(1), 450–463 (2014)
34. X Liao, L Fan, F Gao, Blind channel estimation for OFDM modulated two-way relay network. Wireless Communications and Networking Conference (WCNC), 1–5 (2010)
35. T-C Lin, S-M Phoong, Blind channel estimation in OFDM-based amplify-and-forward two-way relay networks. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (2016)
36. M Morelli, U Mengali, A comparison of pilot-aided channel estimation methods for OFDM systems. IEEE Trans. Signal Process. **49**(12), 3065–3073 (2001)
37. B Picinbono, Second-order complex random vectors and normal distributions. IEEE Trans. Signal Process. **44**(10), 2637–2640 (1996)
38. HJ Larson, BO Shubert, *Probabilistic Models in Engineering Sciences, Vols. I and II, first edition*. (Wiley, New York, 1979)
39. RA Horn, CR Johnson, *Matrix Analysis*. (Cambridge University Press, Cambridge, 1985)