Iterative LMMSE Individual Channel Estimation over Relay Networks with Multiple Transmit and Receive Antennas

Jianpeng Ma, Student Member, IEEE, Shun Zhang, Member, IEEE, Hongyan Li, Member, IEEE, Nan Zhao, Senior Member, IEEE, and Arumugam Nallanathan, Fellow, IEEE

Abstract—In this paper, we investigate the individual channel estimation over the three-node one-way relay network (OWRN), where all nodes are equipped with multiple antennas. We first examine the scenario where the relay node is equipped with single antenna. An iterative linear minimum mean-square-error (LMMSE) method, which has fast convergence speed, is proposed to estimate the individual channels. The closed-form least square (LS) channel estimator is also derived through matrix unitary diagonalization to provide one good initialization point for the iterative LMMSE estimator. To evaluate the performance of the proposed algorithm, we present two performance lower bounds: Bayesian Cramér lower bound (BCRB) and linear estimation lower bound (LELB). Then, the training block and the relay amplification factor are optimized through minimizing the LELB. After that, our studies are extended to the general case where all the nodes are equipped with multiple nodes. Finally, numerical results are provided to corroborate our proposed studies.

Index Terms—Bayesian Cramér lower bound, individual channel estimation, linear estimation lower bound, one-way relay network.

I. INTRODUCTION

The cooperative relay network is efficient to combat the wireless channel fading and enhance the transmission capacity [2]–[4]. As one of the key technologies for the future wireless networks, the cooperative relay network has attracted more and more attention. Several efficient techniques, such as the relay beamforming [5], [6], the best relay selection [7], [8], the multi-antenna relay [9], [10], the distributed space time coding (DSTC) [11], [12], have been proposed to take the advantages of the wireless relay networks.

The accurate channel state information (CSI) at some or all the nodes in the relay networks is essential for achieving the promised performance of the above mentioned techniques. Several channel estimation algorithms at the destination nodes of the amplify-and-forward (AF) one-way relay networks (OWRN) have been proposed. In [15], the expectation conditional maximization (ECM) based iterative algorithm was applied to jointly estimate the channel and detect the data. In [16], the maximum a posteriori (MAP) based estimation schemes are developed to estimate the composite channel. In [17], Wang et al. investigated the effect of antenna correlations on the design of training sequences and channel estimation performance. In [18], one weighted least-square (WLS) channel estimator has been developed for the multiple-input multiple-output (MIMO) relay communication systems. The algorithms in [13]–[18] only estimate the cascaded channels from the source node to the relay node and then to the destination node. However, the individual channels (in-channels) along the link from the source node to the relay node and that from the relay node to the destination node are necessary for some optimal system designs in certain scenarios, such as the relay beamforming [19] and the subcarrier pairing [20].

To estimate the in-channels, Gao et al. proposed one superimposed training framework [22], where the relay superimposes its own training over the received one. Due to its advantages in efficiency and low-overhead, most of the recent in-channel estimation algorithms were proposed under the superimposed training framework [23]–[25]. In [23], Xie et al. designed two maximum likelihood (ML) estimators to recover the in-channels. In [24], Zhang et al. examined the performances of the two in-channel estimators under the superimposed training framework, i.e., the fully data-aided (FDA) estimator and the partially data-aided (PDA) estimator. The FDA estimator has knowledge about the training symbols from both the source and the relay, while the PDA estimator only has the statistical information about the data from the source and the full information of the training symbols superimposed by the relay. Moreover, Zhang et al. also studied the in-channel estimation under the time selective flat fading scenario [25]. Specifically, the unscented Kalman filter (UKF) was used to track the time-varying in-channels, while the unscented Rauch-Tung-Striebel smoother (URTSS) was adopted to smooth the UKF’s results.

The works [22]–[25] only considered the classical relay scenario, where each node in the network is equipped with single antenna. Yet, some attentions have been paid to the in-channel estimation in the multiple-input and multiple-output (MIMO) based OWRN, where some or all nodes are equipped with multiple antennas. In [26], Rong et al. compared the two-stage and the superimposed training schemes for the in-channel estimation in the MIMO-based relay networks, and optimized the training structures through minimizing the mean-squared error (MSE) of the in-channel estimation. In [27], Jing et al. designed one low-complex approximated ML in-channel estimator for the three-node AF relay network.

The authors are with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi’an 710071, P. R. China. (Email: jppmax-du@gmail.com; zhangshunsdu@gmail.com; hyli@mail.xidian.edu.cn).

N. Zhao is with the School of Inform. and Commun. Eng., Dalian University of Technology, Dalian, Liaoning, P. R. China (email: zhaonan@dlut.edu.cn).

A. Nallanathan is with the Department of Informatics, Kings College London, London, UK (Email: arumugam.nallanathan@kcl.ac.uk).

A preliminary version of this paper was published in part on proceedings of IEEE International Conference on Communications (ICC) [1].
model of the typical three-node OWRN with multiple-antennas at both the source and the destination is presented in Section II. In Section III, we develop the iterative LMMSE in-channel estimator under the superimposed training framework, complete its convergence analysis and derive the closed-formed LS estimator to provide the initialization point for the iterative estimator. Two performance lower bounds are given in Section IV. Then, we optimize the training structure and amplification factor in Section V. Section VI provides numerical simulation results to show the performance of the proposed algorithm. Moreover, in Section VII, our studies are extended to the scenario, where all the nodes in our system mode are equipped with multiple antennas. Finally, we draw our conclusions in Section VIII.

Notation: The vectors and the matrices are separately represented by the boldface small and capital letters. The matrix transpose, complex conjugate and Hermitian transpose are denoted by $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$, respectively. $E(\cdot)$ denotes the expectation operation. $[x]_i$ is $i$-th entry of the vector $x$, while $[X]_{i,j}$ is the $(i,j)$-th entry of the matrix $X$. The $N \times N$ identity matrix is denoted by $I_N$. $\text{tr}(X)$ represents the trace of $X$, and $\text{vec}(X)$ denotes the column vector formed by stacking the columns of $X$. $\otimes$ is the Kronecker product, and $\|x\|^2$ denotes the two-norm of the vector $x$. $|X|$ is the determinant of the matrix $X$.

II. SYSTEM MODEL

As shown in Fig. 1, we consider a typical three-node OWRN with one source $S$, one relay $R$, and one destination $D$ under the flat fading channel scenario. There is only one antenna at $R$, while the nodes $S$ and $D$ are equipped with $N_s$ and $N_d$ antennas, respectively. Without loss of generality, we do not consider the direct link between $S$ and $D$. The in-channel vectors along the links $S \to R$ and $R \to D$ can be separately written as $h = [h_0, h_1, \ldots, h_{N_r-1}]^T \in \mathbb{C}^{N_r \times 1}$ and $g^T = [g_0, g_1, \ldots, g_{N_d-1}]^T \in \mathbb{C}^{1 \times N_d}$, where $h_i$ represents the channel between the $i$-th antenna of $S$ and $R$, and $g_j$ denotes the channel between $R$ and the $j$-th antenna of $D$. All the in-channels are assumed to be zero-mean circularly symmetric complex Gaussian (CSCG) random variables, and the covariance matrices for $h$ and $g$ can be separately written as $\mathbf{R}_h = \sigma_h^2 \mathbf{I}_{N_s}$ and $\mathbf{R}_g = \sigma_g^2 \mathbf{I}_{N_d}$. Furthermore, we assume that $h$ is independent of $g$, and both $h$ and $g$ are block fading, which means that they do not change within one phase but may change from phase to phase.

Under the superimposed training framework, one round of the training symbols transmission from $S$ to $D$ can be partitioned into two phases. During phase I, $S$ sends one source-training block $T_s \in \mathbb{C}^{C_T \times N_s}$ to $R$, where $\tau$ is the number of the time slots for training block transmission. The received training at $R$ during phase I can be expressed as

$$r_r = T_s h + n_r,$$

where $n_r \in \mathbb{C}^{C_r \times 1}$ is the additive white Gaussian noise (AWGN) vector with zero mean and the covariance matrix $\sigma_n^2 I_T$. [Fig. 1. The system model and transmission protocol for three-node OWRN with multiple antennas at both source and destination.]

with multiple antennas at the source node and the destination node. In [28], the optimal training sequences in the sense of the minimal MSE is derived for the superimposed training based in-channel estimation under the MIMO-based relay networks. However, we still face some challenges. Since the number of parameters to be estimated in the multiple-antennas case is much larger than that in the single-antenna case, the ML, minimum MSE (MMSE) and maximum a posteriori probability (MAP) estimators for the in-channel estimation in multiple-antennas case require high dimensional searches or multiple integrals and cannot be implemented in an easy way. Theoretically, the linear MMSE (LMMSE) estimator is an alternative when the optimal ML, MMSE and MAP estimators are difficult to implement. Unfortunately, due to the product between the in-channels from the source to the relay and that from the relay to the destination, the observation data model at the destination is not a linear model with respect to each in-channel. Therefore, the LMMSE estimator cannot be directly utilized.

In this paper, we apply the superimposed training based in-channel estimation framework to the classical three-node OWRN, where both the source and the relay have multiple antennas. When the flat channel fading scenario is considered, we can achieve the following observation: for a specific realization of the in-channels along the link from the source to the relay (or from the relay to the destination), the observation data model at the destination is one Bayesian linear model with respect to the in-channels along the link from the source to the relay. Based on the above observation, we propose an iterative LMMSE in-channel estimator to address the challenges mentioned in the previous paragraph. Then, we formulate the closed-formed LS estimator through the matrix unitary diagonalization to provide one good initialization point for the iterative algorithm. In order to evaluate the proposed algorithm, we derive two performance lower bounds: the classical Bayesian Cramér lower bound (CRB) [29], [30], and the linear estimation lower bound (LELB). Interestingly, the LELB is proved to be more tighter than the Bayesian CRB. Moreover, we optimize the optimal training blocks and the relay’s amplification factor through minimizing the LELB. Finally, our studies are extended to one general case, where all the three nodes (the source, the relay, and the destination) are equipped with multiple antennas.

The rest of the paper is organized as follows. The system
During phase II, $\mathbb{R}$ superimposes its training sequence $t_r \in C^{T \times 1}$ over its received training $r_r$, and the resultant training from $\mathbb{R}$ can be denoted as [22]–[25]

$$r_t = \alpha r_r + t_r. \quad (2)$$

Then, $\mathbb{R}$ forwards $r_t$ to $\mathbb{D}$, and the received training at $\mathbb{D}$ can be written as

$$Y = \alpha T_s h g^T + t_r g^T + \alpha n_r g^T + N_d,$$

where $N_d \in C^{T \times N_d}$ denotes the AWGN matrix at $\mathbb{D}$, whose entries are i.i.d CSG random variables with zero mean and variance $\sigma_n^2$.

We set the average transmitting power of $\mathbb{R}$ as $P_r$, i.e.,

$$P_r = E_h\{\|\alpha (T_s h + n_r) + t_r\|^2\}$$

$$= \alpha^2 \sigma_h^2 P_s + \sigma_n^2 \alpha^2 P_s + P_t,$$

where $P_s = tr\{T_s T_s^H\}$ is the power of the source training block and $P_t = \|t_r\|^2$ is the power of the relay training sequence. Then the amplification factor $\alpha$ can be expressed as

$$\alpha \triangleq \sqrt{\frac{P_r - P_t}{\sigma_h^2 P_s + \sigma_n^2}}. \quad (5)$$

### III. IN-CHANNEL ESTIMATION

#### A. Iterative LMMSE In-Channel Estimator

Before preceding, let us define $y = \text{vec}(Y)$ and $n_d = \text{vec}(N_d)$. Resorting to the Kronecker product property vec$(ABC) = (C^T \otimes A) \text{vec}(B)$ [31], we can obtain

$$y = (\alpha (I_{N_d} \otimes T_s h) + (I_{N_d} \otimes t_r)) g + n$$

$$= (\alpha (I_{N_d} \otimes T_s h) + (I_{N_d} \otimes t_r)) g + n,$$

where the equivalent noise vector $n$ defined as the corresponding term is related to the specific realization of $g$.

When the in-channel statistics is known, the optimal in-channel estimation method is the MAP estimator, which can be explicitly expressed as

$$\hat{h} \mid g = \arg \max_{h \mid g} p(y \mid h, g)p(h \mid g),$$

where $p(y \mid h, g)$ denotes the probability density function (PDF) of $y$ conditioned on $h$, $g$, and $p(h \mid g)$ is the joint PDF of $h$, $g$. With the equation (6) and the statistical characteristics of $h$ and $g$, we can derive the following equations as

$$\ln p(y \mid h, g) = \text{Const.} - \ln |C_{h \mid g}| - (y - \mu)^T C_{h \mid g}^{-1}(y - \mu),$$

$$\ln p(h \mid g) = \text{Const.} - \sigma_h^{-2} \|h\|^2 - \sigma_g^{-2} \|g\|^2,$$

where

$$\mu = \alpha (g \otimes T_s h) + (g \otimes t_r) t_r = (\alpha (I_{N_d} \otimes T_s h) + (I_{N_d} \otimes t_r)) g,$$

and

$$C_{h \mid g} = \text{E}\{nn^H \mid h, g\} = \sigma_n^2 (\alpha^2 gg^H + I_{N_d}) \otimes I_r.$$

It is noticed that the Kronecker product property $(A \otimes B)(C \otimes D) = AC \otimes BD$ is utilized in the above derivation. After straightforward calculation, the MAP in-channel estimator can be represented as

$$\hat{h} \mid g = \arg \min_{h \mid g} (y - \mu)^T C_{h \mid g}^{-1}(y - \mu)$$

$$+ \ln |C_{h \mid g}| + \sigma_h^{-2} \|h\|^2 + \sigma_g^{-2} \|g\|^2. \quad (10)$$

Unfortunately, due to its complicated structure, especially the presence of $C_{h \mid g}$, the MAP estimator requires a high dimensional search and is difficult to implement. Theoretically, the LMMSE estimator, which minimizes the MSE for the unknown parameters’ estimation under the constraint that the estimator must be linear, can be adopted as one suboptimal estimator under the Bayesian framework. However, due to the presence of nonlinear term $\|g \otimes T_s h\|$, the observation data model in (6) is not a Bayesian linear model with respect to $h$ and $g$. Thus, the Bayesian Gaussian-Markov theorem does not hold here, and the LMMSE estimator cannot be directly used to estimate the in-channels $h$ and $g$. Instead, we would like to conceive an iterative LMMSE in-channel estimator.

The proposed iterative LMMSE in-channel estimator is based on the following observation: for specific realization of $g$ (or $h$), the observation data model of $y$ in (6) is a Bayesian linear model with respect to $h$ (or $g$). Hence, with the Bayesian Gaussian-Markov theorem [32], the LMMSE estimation of $h$ conditioned on given $g$ can be formulated as

$$\hat{h} \mid g = \alpha \left( R_{h_1}^{-1} + \alpha^2 (g \otimes T_s h)^T C_{n \mid g}^{-1}(g \otimes T_s h) \right)^{-1}$$

$$\left( g \otimes T_s h \right)^T C_{n \mid g}^{-1} (y - (g \otimes t_r) t_r), \quad (11)$$

where the covariance matrix

$$C_{n \mid g} = \text{E}\{nn^H \mid g\} = \sigma_n^2 (\alpha^2 gg^H + I_{N_d}) \otimes I_r. \quad (12)$$

Utilizing the matrix property equations $(I + AB)^{-1} = I - A(I + BA)^{-1} B$ and $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$, we can derive the inverse of $C_{n \mid g}$ as

$$C_{n \mid g}^{-1} = \sigma_n^{-2} \left( I_{N_d} - \alpha^2 (1 + \alpha^2 \|g\|^2) g^H g \right) \otimes I_r.$$

Substituting (13) into (11), we can reexpress (11) as

$$\hat{h} \mid g = \alpha \left( R_{h_1}^{-1} + \frac{\alpha^2 \|g\|^2}{\sigma_n^2 (1 + \alpha^2 \|g\|^2)} T_s^H T_s \right)^{-1}$$

$$\left( g^H (I_{N_d} - \frac{\alpha^2 \|g\|^2}{\sigma_n^2 (1 + \alpha^2 \|g\|^2)} T_s^H T_s ) \right) (y - (g \otimes t_r) t_r)$$

$$= \frac{\alpha}{\sigma_n^2 (1 + \alpha^2 \|g\|^2)} \left( R_{h_1}^{-1} + \frac{\alpha^2 \|g\|^2}{\sigma_n^2 (1 + \alpha^2 \|g\|^2)} T_s^H T_s \right)^{-1}$$

$$T_s^H (y - t_r g^T) g^*, \quad (14)$$

where the matrix equation vec$(ABC) = (C^T \otimes A) \text{vec}(B)$ is used in the derivation.

Let us further investigate the estimation of $g$ for a specific $h$. Similarly, the covariance matrix of $n$ conditioned on $h$ can be denoted as

$$C_{n \mid h} = \text{E}\{nn^H \mid h\} = \sigma_n^2 (\alpha^2 R_g + I_{N_d}) \otimes I_r. \quad (15)$$
Algorithm 1 The Iterative LMMSE In-Channel Estimator for In-Channels h and g

Initialization: Initial LS in-channel estimation
1: Unitarily diagonalize the Hermitian matrix $Y^H \Xi Y$ to obtain the eigenvector that corresponds to the maximal eigenvalue of $Y^H \Xi Y$, i.e., $v_0$, and derive $\hat{g}_{LS}$ according to (29).
2: Calculate $\|g\|_{LS}$ according to (23), and obtain $\hat{g}_{LS}$ as

$$\hat{g}_{LS} = \|g\|_{LS} \hat{g}_{LS}.$$ 

3: Calculate $h_{LS}$ according to (19) by substituting $\hat{g}_{LS}$ into it.
4: $k = 0$, $\hat{h}_0 = \hat{h}_{LS}$, $\hat{g}_0 = \hat{g}_{LS}$.
5: repeat
6: $k = k + 1$.
7: Calculate the current $\hat{h}_k$ with the linear estimator (14) by substituting the previous $\hat{g}_{k-1}$ into it.
8: Calculate the current $\hat{g}_k$ with the linear estimator (17) by substituting the current $\hat{h}_k$ into it.
9: until some termination criterion is satisfied.
10: return $\hat{h}_k, \hat{g}_k$.

Resorting to the fact $R_g = \sigma_g^2 I_{N_d}$, we can derive the inverse of $C_{n|h}$ as

$$C_{n|h}^{-1} = \frac{1}{\sigma_n^2(\sigma^2 \sigma_g^2 + 1)} I_{F N_d}.$$

Then, for given h, the standard LMMSE estimation of g is given by

$$\hat{g}|h = \left(R_g^{-1} + (I_{N_d} \otimes (\alpha T_s h + t_r))^H (\alpha T_s h + t_r) \right)^{-1} (I_{N_d} \otimes (\alpha T_s h + t_r))^H y = \frac{1}{\sigma_n^2(\sigma^2 \sigma_g^2 + 1)} \left(R_g^{-1} + \frac{\|\alpha T_s h + t_r\|^2}{\sigma_g^2(\sigma^2 \sigma_g^2 + 1)} I_{N_d}\right)^{-1} \left(I_{N_d} \otimes (\alpha T_s h + t_r)^H\right) y = \left(\sigma_n^2(\alpha^2 + \sigma_g^2) + \|\alpha T_s h + t_r\|^2\right)^{-1} Y^T(\alpha T_s h + t_r)^*.$$

(17)

B. Initial LS In-Channel Estimator

As is well known, a good initial point is essential for an iterative estimator since it affects the convergence point and speed of convergence. Obviously, the LS estimator

$$\{\hat{h}_{LS}, \hat{g}_{LS}\} = \arg \min_{h, g} \|y - \alpha (g \otimes T_s) h - (g \otimes I_r) t_r\|^2$$

$$= \arg \min_{h, g} \|y - \alpha T_s h g^T - t_r g^T\|_F^2$$

(18)

could serve as a good initial point for the iterative LMMSE in-channel estimator.

With given g, the LS estimation of h can be expressed as

$$\hat{h}_{LS} = \alpha^{-1}\|g\|^{-2} \left(g^H (T_s^H T_s) g^H\right) (y - (g \otimes I_r) t_r) = \alpha^{-1}\|g\|^{-2} (T_s^H T_s)^{-1} T_s^H (Y - t_r g^T) g^*.$$ 

(19)

Incorporating (19) into (18) and recalling (6), we can rewrite (18) as

$$\hat{g}_{LS} = \arg \min_g \|Y - t_r g^T - g\|_F^{-2} W(Y - t_r g^T) g^T g^T.$$ 

(20)

where $W = T_s (T_s^H T_s)^{-1} T_s^H$. For simplicity, instead of minimizing the right hand side (RHS) of (20) through optimizing g directly, we can minimize it first with respect to the norm of g, i.e., $\|g\|$, and then with respect to the normalized version of g, i.e., $g = \frac{g}{\|g\|}$. Hence, we should reformulate (20) as

$$\arg \min_{g, \|g\|} \|Y - \|g\| t_r g^T - W(Y - \|g\| t_r g^T) g^T g^T\|_F^2 = \arg \min_{g, \|g\|} \|Y - \|g\| t_r g^T\|_F^2 - \|W(Y - \|g\| t_r g^T) g^T g^T\|_F^2 = \arg \min_{g, \|g\|} \|g\|^2 (I_r - W) t_r^2 - \|g\|^2 \|W^g\|^2.$$ 

(22)

Fixing g and taking the first order derivative of (22)’s RHS term with respect to $\|g\|$, we can derive the LS estimation of g as

$$\hat{g}_{LS} = \arg \max_g L(g).$$ 

(24)

where $L(g) = \Re^2 (t_r^H (I_r - W) Y g^*) + \|I_r - W\| t_r^2 \|W^g\|^2$.

From the basic inequalities, we can prove that $L(g)$ satisfy

$$L(g) \leq g^T Y^H \Xi Y g^*,$$

(25)

where the Hermitian matrix $\Xi = \|I_r - W\| t_r^2 \|W^g\|^2$. From the inequality holds when the term $t_r^H (I_r - W) Y g^*$ is real. Note that the quadratic polynomial $g^T Y^H \Xi Y g^*$ can serve as an upper bound of the objective function $L(g)$. Instead of directly solving (24), we can first look into the following optimization:

$$\nu_{opt} = \arg \max_{\nu \geq 1} \nu^H Y^H \Xi Y \nu,$$

(26)

where $\nu$ denotes the $N_d \times 1$ column vector. It can be readily checked that $Y^H \Xi Y$ is Hermitian and can be decomposed as

$$Y^H \Xi Y = V \text{diag}(\lambda_0, \lambda_1, \ldots, \lambda_{N_d-1}) V^H,$$

(27)

where $V = [v_0, v_1, \ldots, v_{N_d-1}]$ is an unitary matrix, i.e., $V V^H = V^H V = I$ and the real variables $\lambda_0, \lambda_1, \ldots, \lambda_{N_d-1}$ are the eigenvalues of $Y^H \Xi Y$. Without loss of generality, we can assume $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{N_d-1}$. 


From the Rayleigh-Ritz theorem [31], it can be obtained that
\[
\lambda_0 = \max_{\|\nu\|^2=1} \nu^H Y^H \Xi Y \nu,
\]
where the maximal value is achieved, if \( \nu_{opt} \) equals the eigenvector of \( Y^H \Xi Y \) corresponding to the eigenvalue \( \lambda_0 \), i.e., \( \nu_{opt} = \nu_0 \).

With the above analysis, we can conclude that \( L(g) \leq \nu_0^H Y^H \Xi Y \nu_0 \) and the upper bound can be achieved only if \( \hat{g}_{LS} = \nu_0^H \Xi \omega \nu_0 \).

Let us use \( \hat{h}_k \) and \( \hat{g}_k \) to denote the estimations of \( h \) and \( g \) during the \( k \)-th iteration, respectively. Then, the proposed iterative LMMSE in-channel estimation algorithm is outlined in Algorithm 1.

### C. Convergence Analysis

To analyse the convergence property of proposed iterative LMMSE in-channel estimator, we derive the following Lemma.

**Lemma 1:** Define the variable \( \alpha > 0 \) and the \( N \times N \) positive semi-definite Hermitian matrix \( X \). For the following matrix function
\[
f(X) = (\alpha I + A^H X^{-1} A)^{-1},
\]
we have \( \text{tr}(f(X_2)) - \text{tr}(f(X_1)) < 0 \) if \( \text{tr}(X_2) < \text{tr}(X_1) \).

**Proof:** See Appendix A.

It is clear that the current estimation of \( h \) (or \( g \)) is subjected to the error propagation from the previous imperfect estimation of \( g \) (or \( h \)). During the \( k \)-th iteration, the average MSEs (AMSEs) for both \( \hat{h}_k \) and \( \hat{g}_k \) can be separatedly expressed as
\[
\begin{align*}
\text{tr}(C_{\Delta h_k}) &= \text{tr}\left(E\{\Delta h_k \Delta h_k^H\}\right), \\
\text{tr}(C_{\Delta g_k}) &= \text{tr}\left(E\{\Delta g_k \Delta g_k^H\}\right).
\end{align*}
\]

where \( \Delta h_k = h - \hat{h}_k \) and \( \Delta g_k = g - \hat{g}_k \). Plugging \( g = \hat{g}_k + \Delta g_k \) into (6), we can obtain that
\[
\begin{align*}
y &= \alpha (\hat{g}_k \otimes T_s) h + (\hat{g}_k \otimes t_r) \text{tr} \\
&+ \alpha (\Delta g_k \otimes T_s) h + (\Delta g_k \otimes I_r) t_r + n,
\end{align*}
\]

where \( \omega | \Delta g_k \) represents the equivalent noise during the estimation of \( h \) conditioned on given \( \Delta g_k \). Correspondingly, its covariance matrix can be listed as
\[
C_{\omega | \Delta g_k} = \alpha^2 \sigma_h^2 C_{\Delta g_k} \otimes (T_s T_s^H) + (I_{N_d} \otimes t_r) C_{\Delta g_k} (I_{N_d} \otimes T_s^H) + C_{n|g}.
\]

Furthermore, we can derive the covariance matrix for the LMMSE estimation \( \hat{h}_{k+1} \) conditioned on given \( \Delta g_k \), as [32]
\[
C_{\Delta h_{k+1} | \Delta g_k} = \left(R_h^{-1} + \alpha^2 (\hat{g}_k \otimes T_s^H)(C_{\omega | \Delta g_k})^{-1} (\hat{g}_k \otimes T_s)\right)^{-1}.
\]

Since all the matrices terms in (34) and (35) are positive semi-definite, it can be concluded from Lemma 1 that the decrease of the AMSE for \( \hat{g}_k \), i.e., \( \text{tr}(C_{\Delta g_k}) \), will reduce the AMSE of \( h_{k+1} \), i.e., \( \text{tr}(C_{\Delta h_{k+1}}) \). Similarly, substituting \( h = h_{k+1} + \Delta h_{k+1} \) into (6), we get
\[
y = \alpha (I_{N_d} \otimes T_s \hat{h}_{k+1}) + (I_{N_d} \otimes t_r) g + \alpha (I_{N_d} \otimes T_s \Delta h_{k+1}) g + n,
\]

and
\[
C_{\omega | \Delta h_{k+1}} = \alpha^2 \sigma_g^2 (I_{N_d} \otimes T_s) C_{\Delta h_{k+1}} (I_{N_d} \otimes T_s^H) + C_{n|h}.
\]

Then, the covariance matrix for \( g \)’s estimation error conditioned on given \( \Delta g_{k+1} \) in \( (k + 1) \)-th iteration step can be written as
\[
C_{\Delta g_{k+1} | \Delta h_{k+1}} = \left(R_g^{-1} + (I_{N_d} \otimes (\alpha T_s \hat{h}_{k+1} + t_r))^H (C_{\omega | \Delta h_{k+1}})^{-1} (I_{N_d} \otimes (\alpha T_s \hat{h}_{k+1} + t_r))^H\right)^{-1}.
\]

Basing on (37) (38) and Lemma1, we can also find that decrease of the AMSE for \( \hat{h}_{k+1} \), i.e., \( \text{tr}(C_{\Delta h_{k+1}}) \), will reduce the AMSE of \( \hat{g}_{k+1} \), i.e., \( \text{tr}(C_{\Delta g_{k+1}}) \).

Obviously, the estimation error of \( h \) (or \( g \)) is lower bounded by estimation with exact given \( g \) (or \( h \)), which is given by
\[
\begin{align*}
\text{tr}(C_{\Delta h_{|g}}) &= \text{tr}\left((R_h^{-1} + \alpha^2 (g \otimes T_s) C_{n|h})^{-1} (I_{N_d} \otimes (\alpha T_s g + t_r))^H\right) \\
&= \text{tr}\left((\sigma_h^{-2} I_{N_d} + \frac{\alpha^2 \|g\|^2}{\sigma_h^2 (1 + \alpha^2 \|g\|^2) T_s^H T_s})^{-1}\right).
\end{align*}
\]

With a good initial point provided by LS estimator, the estimation errors of \( h \) and \( g \) decrease alternately. Moreover, as the estimation error has a lower bound, it is easy to draw the conclusion that the proposed iterative LMMSE in-channel estimator converges to a minimum estimation error.

### IV. PERFORMANCE LOWER BOUNDS

In this section, we first derive the LELB to evaluate the performance of the proposed iterative LMMSE in-channel estimator. Then, to solidify the derived LELB, we will also derive the Bayesian CRB, which is widely used to evaluate the performance of the estimators under the Bayesian framework. Interestingly, it will be proved that the LELB is tighter than the Bayesian CRB.
A. Linear Estimation Lower Bound (LELB)

To derive a closed-form LELB, we first introduce the following lemma.

Lemma 2: With the variable $a > 0, x > 0$ and the $N \times N$ matrix $A > 0$, the matrix function

$$f(x) = \text{tr} \left( (aI + xA)^{-1} \right) \tag{41}$$

is convex for $x > 0$.

Proof: See Appendix C.

From the convex optimization theory, we have the following inequality [33]

$$E[f(x)] \geq f(E[x]), \tag{42}$$

where $f(x)$ is one convex function, and $x$ is a random variable.

As derived in section III.C, the lower bound of the estimation error $\text{tr} \{C_{\Delta h_n} \} \{C_{\Delta h_n} \}$ with a specific $g$ ($h$) is given by (39) and (40). By taking the expectation of $\text{tr} \{C_{\Delta h_n} \}$ and $\text{tr} \{C_{\Delta g_n} \}$ over $g$ and $h$, respectively, we can derive the closed-form LELBs for $h$ and $g$ with the help of Lemma 2 and inequality (42) as follows:

$$E_g \left\{ \text{tr} \left( C_{\Delta h_n} \right) \right\} = E_g \left\{ \text{tr} \left[ \left( \sigma_h^2 T_s + \frac{\alpha^2 \|g\|^2}{\sigma_n^2 (1 + \alpha^2 \|g\|^2)} T_s T_s \right)^{-1} \right] \right\}, \tag{43}$$

$$\geq \text{tr} \left[ \left( \sigma_h^2 T_s + \frac{\alpha^2 \|g\|^2}{\sigma_n^2 (1 + \alpha^2 \|g\|^2)} T_s T_s \right)^{-1} \right] = E_h \left\{ \text{tr} \left( C_{\Delta g_n} \right) \right\} = E_h \left\{ \text{tr} \left[ \left( \sigma_g^2 T_s + \frac{\alpha^2 \|g\|^2}{\sigma_n^2 (1 + \alpha^2 \|g\|^2)} T_s T_s \right)^{-1} \right] \right\}, \tag{44}$$

Following the similar methods in [23]–[25], we can calculate $F$ as

$$F = E_{h,g} \left\{ \begin{bmatrix} \frac{\partial^2 p(y|b,g)}{\partial g^T \partial g} & \frac{\partial^2 p(y|b,g)}{\partial g \partial h} \\ \frac{\partial^2 p(y|b,g)}{\partial g \partial h} & \frac{\partial^2 p(y|b,g)}{\partial h^2} \end{bmatrix} \right\} + E_{h,g} \left\{ \begin{bmatrix} \frac{\partial p(y|b,g)}{\partial g} & \frac{\partial p(y|b,g)}{\partial g \partial h} \\ \frac{\partial p(y|b,g)}{\partial g \partial h} & \frac{\partial p(y|b,g)}{\partial h^2} \end{bmatrix} \right\}, \tag{46}$$

where $F_{11} \in \mathbb{C}^{N_s \times N_s}, F_{12} \in \mathbb{C}^{N_s \times N_d}, F_{21} \in \mathbb{C}^{N_d \times N_s}$, and $F_{22} \in \mathbb{C}^{N_d \times N_d}$ are the corresponding sub-matrices of $F$, $C_{n|h,g}$ equals $C_{n|g}$ in (12), the partial derivatives are defined as

$$\frac{\partial \mu}{\partial g} = \left( \frac{\partial \mu}{\partial g_0}, \frac{\partial \mu}{\partial g_1}, \ldots, \frac{\partial \mu}{\partial g_{N_d-1}} \right), \tag{47}$$

and $\Sigma$ is a $N_d \times N_d$ matrix with the $(i,j)$-th element as

$$\Sigma_{ij} = \text{tr} \left( C_{n|h,g}^{-1} \frac{\partial C_{n|h,g}}{\partial g_i} C_{n|h,g}^{-1} \frac{\partial C_{n|h,g}}{\partial g_j} \right). \tag{48}$$

Notice that the facts that $n$ and $h$ are separately independent on $h$ and $g$ are utilized in derivation. From (12), it can be derived

$$\frac{\partial C_{n|h,g}}{\partial g_i} = \sigma_n^2 \alpha^2 e_{i} g_H^T \otimes I_{r}, \tag{49}$$

where $e_i$ denotes the $N_d \times 1$ basis vector with 1 for the $i$-th element and 0 otherwise. Plugging (13), (48) into (47) and taking some tedious mathematical manipulations, we can derive $E_{h,g} \{ \Sigma_{ij} \}$ as

$$E_{h,g} \{ \Sigma_{ij} \} = E_g \left\{ \frac{\tau^4}{1 + \alpha^2 \|g\|^2} \text{tr} \left( g e_i^T g e_j g H \right) \right\} - E_g \left\{ \frac{\tau^6 \|g\|^2}{(1 + \alpha^2 \|g\|^2)^2} \text{tr} \left( g e_i^T g g H e_j g H \right) \right\} = E_g \left\{ \frac{\tau^4 \|g\|^2}{1 + \alpha^2 \|g\|^2} \left( 1 - \frac{\alpha^2 \|g\|^2}{1 + \alpha^2 \|g\|^2} \right) \sigma(i-j) \right\} = E_g \left\{ \frac{\tau^4 \|g\|^2}{1 + \alpha^2 \|g\|^2} \left( 1 - \frac{\alpha^2 \|g\|^2}{1 + \alpha^2 \|g\|^2} \right) \sigma(i-j) \right\}, \tag{49}$$

where $i, j = 0, 1, \ldots, N_d - 1$. The equations $\text{tr}(AB) = \text{tr}(BA), \text{tr}(A \otimes B) = \text{tr}(A) \text{tr}(B)$ and the fact that $g_0$ and $g_j$ have the same distribution are used in the above derivation.

With the definition of $\mu$ in (7), it can be achieved that

$$\frac{\partial \mu}{\partial g} = \alpha (I_{N_d} \otimes T_h) + (I_{N_d} \otimes t_r), \tag{50}$$

$$\frac{\partial \mu}{\partial h} = \alpha (g \otimes T_s).$$
Combining (13) and (50), we have the following results:

\[
E_{h,g}\left\{ \frac{\partial \mu^H}{\partial h} C_{n|h,g}^{-1} \frac{\partial \mu}{\partial h^T} \right\} = \sigma_n^{-2}E_{g}\left\{ \frac{\alpha^2||g||^2}{1 + \alpha^2||g||^2} \right\} T_s^H T_s \tag{51}
\]

\[
E_{h,g}\left\{ \frac{\partial \mu^H}{\partial g} C_{n|h,g}^{-1} \frac{\partial \mu}{\partial g^T} \right\} = \left( E_{h,g}\left\{ \frac{\partial \mu^H}{\partial h^T} C_{n|h,g}^{-1} \frac{\partial \mu}{\partial g^T} \right\} \right)^H = 0_{N_d \times N_s} \tag{52}
\]

\[
E_{h,g}\left\{ \frac{\partial \mu^H}{\partial g} C_{n|h,g}^{-1} \frac{\partial \mu}{\partial g^T} \right\} = \sigma_n^{-2}E_{h,g}\left\{ \alpha^2 ||T_s h + t_r ||^2 \right\} E_{g}\left\{ 1 - \frac{\alpha^2||g_0||^2}{1 + \alpha^2||g||^2} \right\} I_{N_d} \tag{53}
\]

\[
= (\sigma_n^{-2}P_r - \tau)E_{g}\left\{ 1 - \frac{\alpha^2||g_0||^2}{1 + \alpha^2||g||^2} \right\} I_{N_d} \tag{54}
\]

With and (46), (49) and (53), the sub-matrices \( F_{ij} \) (\( i,j = 1,2 \)) of \( F \) can be derived as

\[
F_{12} = F_{12}^H = 0_{N_s \times N_d} \tag{55}
\]

\[
F_{22} = \sigma_n^{-2}E_{g}\left\{ \alpha^2 ||g||^2 \right\} T_s^H T_s + R_h^{-1} \tag{56}
\]

\[
= \omega_1 T_s^H T_s + \sigma_h^{-2} I_{N_s}, \tag{57}
\]

\[
F_{11} = E_{g}\left\{ \tau \alpha^2 ||g||^2 \right\} \left( 1 - \frac{\alpha^2||g_0||^2}{1 + \alpha^2||g||^2} \right) I_{N_d} \tag{58}
\]

\[
= \omega_2 I_{N_d} + \omega_3 I_{N_d} + \sigma_g^{-2} I_{N_d}. \tag{59}
\]

The proof of Theorem 1 is completed.

V. OPTIMAL TRAINING AND AMPLIFICATION FACTOR DESIGN

Since LELB is tighter than BCRB, we will adopt the LELB as the criterion to design the training block and amplification factor. Fortunately, the LELB has simple analytical solutions, which can be seen from (43) and (44). Hence, it can provide sufficient insight on the training parameter optimization.

A. Training Design

It can be concluded from (43) and (44) that the source-training block \( T_s \) only affects the LELB, and that the structure of relay-training block \( t_r \) has no impact either on LELB or on LELB_g. Therefore, the optimal training design can be done through minimizing the LELB, with respect to \( T_s \), which can be formulated as follows

\[
\text{min}_{T_s} \text{LELB}_{T_s} \quad \text{s.t.} \quad \text{tr}\{T_s^H T_s\} = P_s
\]

For any positive-definite \( N_s \times N_s \) matrix \( A \), we have [35]

\[
\text{tr}(A^{-1}) \geq \sum_{i=0}^{N_s-1} \frac{1}{|A_{ii}|}, \tag{61}
\]

where the equality holds when \( A \) is diagonal. With this inequality, it can be derived that

\[
\text{LELB}_{T_s} \geq \sum_{i=0}^{N_s-1} \frac{1}{\sigma_h^{-2} + \omega_1 |T_s^H T_s|_{ii}} \tag{62}
\]

where the equality in (a) holds when \( T_s^H T_s \) is diagonal, and the equality in (b) holds when each diagonal entry of \( T_s^H T_s \) equals \( P_s/N_s \) [36]. Thus, the optimal \( T_s \) should satisfy the following constraint.

\[
T_s^H T_s = \frac{P_s}{N_s} I_{N_s}. \tag{63}
\]

Notice that the number \( \tau \) of the time slots for training block transmission should be no less than \( N_s \) to satisfy the above constraint. Since the system throughput decreases with the increase of \( \tau \), the optimal choice of \( \tau \) is \( \tau = N_s + 1 \) to ensure that \( t_r \) is also orthogonal with \( T_s \). With the above analysis, an example of the optimal training blocks \( T_s \) and \( t_r \) can be constructed from the \((N_s+1) \times (N_s+1)\) discrete Fourier transform (DFT) matrix as

\[
[T_s]_{i,j} = \sqrt{\frac{P_s}{\tau N_s}} e^{-2\pi j i/N_s}, \quad i = 0, \ldots, \tau-1, \quad j = 0, \ldots, N_s-1. \tag{64}
\]

\[
[t_r]_i = \sqrt{\frac{P_t}{\tau}} e^{-2\pi i N_s}, \quad i = 0, \ldots, \tau-1. \tag{65}
\]
presented on the top of the next page, where \( \xi = \frac{P_{r}\sigma_{n}^{-2}+\tau\sigma_{a}^{2}}{N_{s}N_{a}\sigma_{a}^{2}+\tau\sigma_{a}^{2}} \)
and the derivative formula \( \frac{\delta}{\delta x} F_{1}(a; b; x) = \frac{\tau}{x} F_{1}(a+1; b+1; x) \)
of the confluent hypergeometric function is utilized in the derivation. Due to its complicated structure, it is difficult to derive the analytic expression of the optimal \( \alpha \) through solving the equation \( f'(\alpha) = 0 \). However, its numerical solution can be conveniently obtained with the gradient descent algorithm [39].

VI. SIMULATIONS

In this section, the numerical results are provided to examine the proposed iterative LMMSE estimator. All the in-channels and noise are assumed to have unit variance, i.e., \( \sigma_{h}^{2} = \sigma_{g}^{2} = \sigma_{n}^{2} = 1 \). Without loss of generality, we set \( N_{s} = N_{d} = N \). We fix \( P_{r} = 2P_{s} \) and the signal to noise ratio is defined as \( \text{SNR} = P_{s}/\sigma_{n} = P_{s} \). The normalized MSEs, BCRBs, and LELBs (NMSEs, NBCRBs and NLELBs) of \( h, g \) are used as the figures of the merit, which are defined as

\[
\text{NMSE}_{\theta} = \frac{E\{||\theta - \hat{\theta}||^{2}\}}{N},
\]
\[
\text{NBCRB}_{\theta} = \frac{\text{BCRB}_{\theta}}{N},
\]
\[
\text{NLELB}_{\theta} = \frac{\text{LELB}_{\theta}}{N},
\]

where \( \theta = h, g \). In total, \( 10^{4} \) Monte-Carlo runs are adopted for numerical average.

Firstly, we examine the performance and convergence property of the proposed iterative LMMSE estimator. Here, we set \( \alpha = 0.8 \) and utilize the optimal orthogonal training block in (64) and (65).

Fig. 2 presents the NMSE curves of the initial LS estimation and the proposed iterative LMMSE in-channel estimator. It can be seen that the iterative LMMSE in-channel estimator can effectively improve the accuracy of the initial LS in-channel estimation in terms of NMSE of both \( h \) and \( g \). Compared with the cases of iterations 3 and 5, the iterative LMMSE in-channel estimator with iteration=1 can achieve larger performance gains. The iterative LMMSE algorithm only takes five iterations to arrive at a steady state, which shows that the algorithm has fast convergence speed. Moreover, the NMSE curves versus number of iteration is show in Fig. 3 to explicitly exhibit the convergence of proposed estimator.

Fig. 4 and Fig. 5 compare NLELBs and NBCRBs with \( N = 2 \) and \( N = 6 \), respectively. Five iterations are used for the iterative LMMSE algorithm. We obtain the following observations: the in-channels’ \( h \) and \( g \) NMSE curves of the iterative LMMSE estimator cannot approximate the NLELB curves, which can be explained by the in-channel estimation error propagation phenomena; the NLELB is more tighter than NBCRB, which matches the theoretical analysis in Theorem 1; When \( N \) increases from \( N = 2 \) to \( N = 6 \), NLELB of both \( h \) and \( g \) become quite close to the NMSE, but the NBCRB is still far away from NMSE.
\[ f(\alpha) = \lambda N_s \gamma h_s \left( \sigma_h^{-2} + \frac{P_s}{N_s \sigma_h^2} e^{-\sigma_h^{-2} \alpha^{-2}} + \frac{P_s}{N_s \sigma_h^2} e^{-\sigma_h^{-2} \alpha^{-2}} \right) \]

\[ f'(\alpha) = 2 \xi \lambda N_s \gamma h_s \left( \sigma_h^{-2} + \frac{P_s}{N_s \sigma_h^2} e^{-\sigma_h^{-2} \alpha^{-2}} + \frac{P_s}{N_s \sigma_h^2} e^{-\sigma_h^{-2} \alpha^{-2}} \right) \]

\[ 1 F_1(N_d - 1; N_d; \sigma_g^{-2} \alpha^{-2}) - \frac{N_d - 1}{N_d \alpha \sigma_g^2} 1 F_1(N_d; N_d + 1; \sigma_g^{-2} \alpha^{-2}) + 2 \alpha \sigma_g^2 \gamma N_d \left( \sigma_g^{-2} + \frac{\sigma_n^{-2} P_r - \tau}{\alpha^2 \sigma_g^2 + 1} \right) \]

Fig. 6. Comparison of the in-channels \( h, g \) estimation NMSEs versus SNR under different \( \tau \). Fig. 4. Comparison between the LELB and BCRB versus SNR with \( N = 2 \). Fig. 7. The weighted sum NMSEs versus \( \alpha \) with SNR = 30dB. Fig. 5. Comparison between the LELB and BCRB versus SNR with \( N = 6 \). Fig. 8. The weighted sum NMSEs versus \( \alpha \) with SNR = 20dB.

are presented in Fig. 6. Comparing the NMSE with different \( \tau \), we can find that the NMSE\( h \) decreases significantly when \( \tau \) increases from \( N - 1 \) to \( N \) as the training sequences for different source antennas becomes orthogonal, and that NMSE\( g \) decreases when \( \tau \) increases from \( N \) to \( N + 1 \) since the relay’s training sequences becomes orthogonal with source’s training block. However, both the NMSE\( h \) and NMSE\( g \) can’t
be lowered any further by continually increasing $\tau$, which validates our analytical study. It can be also seen that $\text{NMSE}_g$ increases with increasing $\tau$ at low SNR regime. This phenomenon is not strange and can be explained as follow: as the relay’s total power is constrained, more noise power merges into received signal power with increase of $\tau$, which leads to decrease in SNR at destination.

Lastly, we would like to check our designing of optimal $\alpha$. Fig. 7 and Fig. 8 present the curves of weighted sum of $\text{NMSE}_g$ and $\text{NMSE}_h$ versus $\alpha$ with SNR = 30dB and SNR = 20dB, respectively. Three different relay power, i.e., $P_r = 2P_s$, $P_r = 1.5P_s$, and $P_r = P_s$, are adopted for both cases. We set the weight $\gamma_h = 0.2$ and $\gamma_g = 0.8$ as analyzed in main text that the weight of $g$ should be greater than $h$ in general. As observed from both figures, the optimal $\alpha$ obtained by our designed approach (marked as red circle) are well-matched with the lowest points of weighted sum $\text{NMSE}_s$ of simulation, which demonstrated our analytical study. It is clearly that $\alpha$ increases with the increase of relay power $P_r$, which matches the fact that more power should be allocated for forwarding the received source training blocks if relay have more power. Comparing the optimal $\alpha$ with different SNR, we can also conclude that optimal $\alpha$ decreases with decrement of SNR. This result comes from the fact that more power should be allocated for relay’s training sequence to guarantee the estimation of $g$ in low SNR as $g$ has a greater weight.

VII. THE GENERAL CASE WHERE RELAY HAS MULTIPLE ANTENNAS

In previous sections, we only consider the specific case where the relay has single antenna. In this section, following the similar approach in [27], we extend our studies to the general scenario, where the relay has $N_r > 1$ antennas.

As shown in Fig. 9, let us denote the in-channel vector along $S$ to $R$’s $i$-th ($i = 0, 1, \cdots, N_r-1$) antenna and that along $R$’s $i$-th antenna to $D$ as $h_i$ and $g_i$, respectively. One round of the training symbol transmission from $S$ to $D$ is still partitioned into two phases.

During phase I, $S$ sends one source-training block $T_s$ to $R$. Then, the received training at $R$’s $i$-th antenna during phase I can be expressed as

$$r_{r,i} = T_s h_i + n_{r,i},$$

(71)

Different from the single antenna case, we divide the phase II into $N_r$ stages. In the $i$-th stage, $R$ superimposes the relay-training sequence $t_{r,i}$ in $r_{r,i}$, i.e.,

$$r_{r,i} = \alpha_i r_{r,i} + t_{r,i},$$

(72)

and forwards $r_{r,i}$ to $D$ through its $i$-th antenna. Then, the received training at $D$ in stage $i$ can be written as

$$Y_i = \alpha_i T_s h_i g_i T + t_{r,i} g_i T + \alpha_i n_{r,i} g_i T + N_{d,i}.$$  

(73)

Since each relay’s antenna forwards its own received training in each stage independently, with respect to each relay antenna, the problem becomes the same as that for the case where relay has single antenna. Thus, $h_i$ and $g_i$ can be estimated by the proposed iterative LMMSE in-channel estimator from observation (73) in stage $i$. The performance analysis, optimal training design, and optimal amplification factor design for the single antenna case also apply to the case where relay has multiple antennas. It is worth to point out that the the dimension of the source-training block $T_s$ does not increase with relay’s antenna number $N_r$.

VIII. CONCLUSION

In this paper, we investigated the in-channel estimation in the classical three-node AF OWRN with multiple antennas at all the nodes. The typical scenario, where the relay node is equipped with single antenna, was first examined. Under the superimposed training framework, we developed an iterative LMMSE in-channel estimator, whose initial point was provided by the LS in-channel estimator formed by the matrix unitary diagonalization. We further derived LELB and BCRB to evaluate the proposed algorithms. The optimal training and amplification factor was designed by minimizing the LELB. Furthermore, simulation results were provided to corroborate our studies. Finally, we extended our studies to the general case, where all the nodes are equipped with multiple antennas.

APPENDIX A

PROOF OF LEMMA 1

The term $\text{tr}(f(X))$ can be expressed as

$$\text{tr}(f(X)) = \text{tr}((aI + A^H X^{-1} A)^{-1})$$

$$= aN + \text{tr}(A^H X^{-1} A)$$

$$= aN + \frac{1}{\text{tr}(X(AA^H)^{-1})}. \hspace{1cm} (74)$$

From the fact that $(AA^H)^{-1} \succ 0$, it can be derived that $\text{tr}(X_2(AA^H)^{-1}) < \text{tr}(X_1(AA^H)^{-1})$ if $\text{tr}(X_2) < \text{tr}(X_1)$, where $X_1 \succeq 0$ and $X_2 \succeq 0$. Therefore, we can obtain $\text{tr}(f(X_2)) < \text{tr}(f(X_1))$. 

Fig. 9. The system model and training blocks transmission protocol for OWRM with multiple antennas at all nodes.
APPENDIX B

DEFINITIONS AND CALCULATIONS OF $\omega_1$, $\omega_2$ AND $\omega_3$

We first calculate the factor $\omega_1$ as

$$\omega_1 = \frac{\sigma_n^{-2}}{\alpha_n^2} \mathbb{E}_g \left\{ \alpha_n^2 \|g\|^2 \right\}$$

$$= \frac{\sigma_n^{-2}}{\alpha_n^2} \mathbb{E}_g \left\{ \frac{2}{\sigma_n^2 \alpha_n^2} \|g\|^2 \cdot \frac{1}{2 \sigma_n^2 \alpha_n^2 + 2 \sigma_n^2 \|g\|^2} \right\}. \quad (75)$$

With the statistics of $g$, it can be checked that the term $2\sigma_n^{-2} \alpha_n^{-2} + 2\sigma_n^{-2} \|g\|^2$ is the non-central chi-squared distributed with the degrees of freedom $2N_d$ and the non-centrality parameter $2\sigma_n^{-2} \alpha_n^{-2}$. Thus, we can derive that

$$\mathbb{E}_g \left\{ \frac{1}{2 \sigma_n^{-2} \alpha_n^{-2} + 2 \sigma_n^{-2} \|g\|^2} \right\}
= \frac{1}{2N_d} e^{-\frac{\sigma_n^{-2} \alpha_n^{-2}}{2}} F_1(N_d - 1; N_d; \sigma_n^{-2} \alpha_n^{-2}), \quad (76)$$

where

$$F_1(a; b; z) = \frac{\Gamma(b)}{\Gamma(a) \Gamma(b-a)} \int_0^1 t^{a-1} (1-t)^{b-a-1} e^{zt} dt \quad (77)$$

is the confluent hypergeometric function. Plugging (76) into (75), we get the closed-form expression of $\omega_1$.

Following the similar methods, the variables $\omega_2$ and $\omega_3$ in (56) can be separately written as

$$\omega_2 = \mathbb{E}_g \left\{ \tau \alpha_n^4 \|g\|^2 \frac{1 - \frac{\alpha_n^2 \|g\|^2}{1 + \alpha_n^2 \|g\|^2}}{1 + \alpha_n^2 \|g\|^2} \right\}, \quad (78)$$

$$\omega_3 = (\sigma_n^{-2} P_r - \tau) \mathbb{E}_g \left\{ \frac{1 - \frac{\alpha_n^2 \|g\|^2}{1 + \alpha_n^2 \|g\|^2}}{1 + \alpha_n^2 \|g\|^2} \right\}. \quad (79)$$

Due to the complicated structures of the expectations in (78) and (79), it is difficult for us to achieve simple analytical solutions for both $\omega_2$ and $\omega_3$. Nonetheless, we can use the Gibbs sampling technique and generate samples $g_1, g_2, \cdots, g_{N_c}$ from its PDF $CN(0, R_G)$. Then, the numerical scheme to calculate $\omega_2$ and $\omega_3$ can be listed as

$$\omega_2 \approx \frac{1}{N_c} \sum_{k=1}^{N_c} \{ \tau \alpha_n^4 \|g_k\|^2 \frac{1 - \frac{\alpha_n^2 \|g_k\|^2}{1 + \alpha_n^2 \|g_k\|^2}}{1 + \alpha_n^2 \|g_k\|^2} \}, \quad (80)$$

$$\omega_3 \approx \frac{1}{N_c} \sum_{k=1}^{N_c} (\sigma_n^{-2} P_r - \tau) \frac{1 - \frac{\alpha_n^2 \|g_k\|^2}{1 + \alpha_n^2 \|g_k\|^2}}{1 + \alpha_n^2 \|g_k\|^2} \}. \quad (81)$$

APPENDIX C

PROOF OF LEMMA 2

With the matrix property $\text{tr}(X^{-1}) = 1/\text{tr}(X)$, $f(x)$ can be rewritten as

$$f(x) = \frac{1}{\alpha N + x \text{tr}(A)}. \quad (82)$$

Moreover, the second derivative of $f(x)$ with respect to $x$ can be denoted as

$$f''(x) = \frac{2 \text{tr}^2(A)}{(\alpha N + x \text{tr}(A))^3} > 0, \quad (x > 0) \quad (83)$$

where $\text{tr}(A) > 0$ for $A > 0$. From the fact that $f''(x) > 0$, we can conclude that $f(x)$ is convex for $x > 0$.

REFERENCES


