ESTIMATION OF CHANNELS IN SYSTEMS WITH INTELLIGENT REFLECTING SURFACES

Michael Joham, Hangze Gao, and Wolfgang Utschick

Department of Electrical and Computer Engineering
Technical University of Munich, Arcisstraße 21, 80333 Munich, Germany
email: joham@tum.de, hangzegao@gmail.com, utschick@tum.de

ABSTRACT

We consider channel estimation for systems equipped with an intelligent reflecting surface (IRS). We develop least squares (LS) and minimum mean square error (MMSE) estimation for such systems. The appropriate system models are developed and we also discuss the parameters which can be estimated in such a setup because there exists a difficulty due to the ambiguity for the two channels connecting with the IRS. The MMSE estimator is based on a Kronecker product approximation of the channel covariance matrix. The simulations results illustrate the advantage of the optimized pilots and the optimized phase allocations for the channel estimation.

Index Terms— Intelligent reflecting surfaces, channel estimation, pilot optimization, least squares, minimum mean square error.

1. INTRODUCTION

The usage of reconfigurable intelligent systems (RIS) or IRS is considered to be a key technology for B5G/6G systems [1], e.g., because an IRS has the potential to enable the communication to a user being in a blocked position. As mentioned in [1], using semi-passive IRS is not cost-efficient. Hence, we concentrate on fully passive IRS which only can reflect the impinging waves with every element.

A system including an IRS with a direct channel per user and a cascaded channel including the IRS is assumed. It is crucial to have available reliable estimates of the different channels to be able to design appropriate strategies for data transmission. The main difficulty can be found in the passiveness of the IRS, i.e., it is unable to process the impinging waves, cannot transmit training data, and, thus, cannot estimate the channels. To circumvent this problem, an on/off strategy was discussed in [2, 3, 4, 5] where the IRS is completely switched off to estimate the direct channel and the elements are separately activated sequentially to estimate the different cascaded channels. However, it was shown in [4] that this strategy is suboptimal since a lot of potentially reflected power is lost. An alternative to cope with the passiveness of the IRS is to neglect the direct path as in [6, 7]. Since usually a large number of elements at the IRS are assumed, a grouping of the IRS elements was considered in [8, 9] having an impact on the channel estimation. Discrete phase shifts for the elements of the IRS were considered, e.g., in [8].

The majority of previous papers on channel estimation for systems with IRS employed LS, e.g., [2, 10, 11, 9]. Minimum variance unbiased estimation was investigated in [4] and MMSE estimation in [12]. The DFT matrix was identified to be the optimum phase allocation during the estimation phase. In [13], a tensor modeling approach for channel estimation in IRS systems was considered.

Contributions: We formulate the system models for different single-user systems and highlight the similarity of multi-user channel estimation to the single-user case also when the system comprises an IRS in Section 2. Additionally, we discuss the inherent ambiguity naturally included in the cascaded channel in Section 3. In Section 4, we consider the LS and MMSE estimation of channels including an IRS and derive the corresponding jointly optimal pilot sequences and phase allocations during the training period. The simulation results show that the optimized pilots and phase allocation clearly outperform other strategies.

2. SYSTEM MODEL

The IRS comprises L passive elements which can only influence the phase shifts when reflecting impinging waves but do not change the amplitudes. Accordingly, the operation of a single IRS element can be represented by the factor

$$v_\ell = \mathrm{e}^{\mathrm{j}\,\theta_\ell} \tag{1}$$

with the corresponding phase shift $\theta_\ell \in [0, 2\pi)$ such that $|v_\ell| = 1$. The factors of the whole IRS are collected in

$$\boldsymbol{v} = [v_1, \ldots, v_L] \in \mathbb{C}^L$$

whose elements must obey $|v_i| = 1$ for all i due to (1).

For a single-user multiple-input multiple-output (MIMO) system, let the direct channel be denoted by $\boldsymbol{H}_0 \in \mathbb{C}^{M\times N}$, that is, the transmitter is equipped with N antennas and the receiver has M antennas. The channel connecting the transmitter and the IRS is $\boldsymbol{H}_1 \in \mathbb{C}^{N\times L}$ and from the IRS to the receiver, the channel is $\boldsymbol{H}_2 \in \mathbb{C}^{M\times L}$. With the transmitted signal $\boldsymbol{x} \in \mathbb{C}^N$, the additive stationary noise $\boldsymbol{n}' \sim \mathcal{N}_\mathbb{C}(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{n}})$, $\boldsymbol{\Phi} = \mathrm{diag}(\boldsymbol{v}) \in \mathbb{C}^{L\times L}$, and the cascaded channel $\boldsymbol{H}_2\boldsymbol{\Phi}\boldsymbol{H}_1^{\mathrm{T}}$, the received signal can be expressed as

$$\boldsymbol{y}' = \boldsymbol{H}_0\boldsymbol{x} + \boldsymbol{H}_2\boldsymbol{\Phi}\boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{n}'. \tag{2}$$

For the particular case with $M = N = 1$, we obtain the single-input single-output SISO system [cf. (2)]

$$y' = h_0 x + \boldsymbol{h}_2^{\mathrm{T}}\boldsymbol{\Phi}\boldsymbol{h}_1 x + n' = h_0 x + \boldsymbol{h}_1^{\mathrm{T}}\boldsymbol{H}_2\boldsymbol{v}x + n' \tag{3}$$

with $n' \sim \mathcal{N}_\mathbb{C}(0, \sigma^2)$, $h_0 \in \mathbb{C}$, $\boldsymbol{H}_2 = \mathrm{diag}(\boldsymbol{h}_2)$, and $\boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathbb{C}^L$.

When $N = 1$ but $M > 1$, i.e., resulting in a single-input multiple-output (SIMO) system, the received signal reads as

$$\boldsymbol{y}' = \boldsymbol{h}_0 x + \boldsymbol{H}_2\boldsymbol{\Phi}\boldsymbol{h}_1 x + \boldsymbol{n}' = \boldsymbol{h}_0 x + \boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{v}x + \boldsymbol{n}' \tag{4}$$

with the direct channel $\boldsymbol{h}_0 \in \mathbb{C}^M$, $\boldsymbol{h}_1 \in \mathbb{C}^L$, $\boldsymbol{H}_1 = \mathrm{diag}(\boldsymbol{h}_1)$, $\boldsymbol{H}_2 \in \mathbb{C}^{M\times L}$, and $\boldsymbol{n}' \sim \mathcal{N}_\mathbb{C}(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{n}})$.

Finally, in the multiple-input single-output (MISO) scenario where $M = 1$ and $N > 1$, the received signal is [cf. (2)]

$$y' = \boldsymbol{h}_0^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{h}_2^{\mathrm{T}}\boldsymbol{\Phi}\boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{x} + n' = \boldsymbol{h}_0^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{H}_2\boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{x} + n' \tag{5}$$

with $\boldsymbol{h}_0 \in \mathbb{C}^N$, $\boldsymbol{H}_1 \in \mathbb{C}^{N\times L}$, $\boldsymbol{h}_2 \in \mathbb{C}^L$, $\boldsymbol{H}_2 = \mathrm{diag}(\boldsymbol{h}_2)$, and $n' \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$.

2.1. Multi-User Downlink System

In the downlink, the channel estimation problem is basically a single-user case because the received signal of user k can be written as [cf. (2)]

$$\boldsymbol{y}'_k = \boldsymbol{H}_{0,k}\boldsymbol{x} + \boldsymbol{H}_{2,k}\boldsymbol{\Phi}\boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{n}'_k$$

with $\boldsymbol{H}_{0,k} \in \mathbb{C}^{M_k\times N}$, $\boldsymbol{H}_{2,k} \in \mathbb{C}^{M_k\times L}$, $\boldsymbol{H}_1^{\mathrm{T}} \in \mathbb{C}^{N\times L}$, and $\boldsymbol{n}' \sim \mathcal{N}_{\mathbb{C}}(0, \boldsymbol{C}_{\boldsymbol{n}})$. Here, M_k denotes user k's number of antennas. Thus, for the estimation of the downlink channels, the discussion on single-user MIMO systems can be employed.

2.2. Multi-User Uplink System

The received signal in the K users uplink reads as

$$\boldsymbol{y}' = \sum_{k=1}^{K}(\boldsymbol{H}_{0,k} + \boldsymbol{H}_2\boldsymbol{\Phi}\boldsymbol{H}_{1,k}^{\mathrm{T}})\boldsymbol{x}_k + \boldsymbol{n}'$$

where $\boldsymbol{x}_k \in \mathbb{C}^{N_k}$ and N_k denotes the number of antennas at user k. Combining the direct channels of the users in $\boldsymbol{H}_0 = [\boldsymbol{H}_{0,1}, \ldots, \boldsymbol{H}_{0,K}] \in \mathbb{C}^{M\times N}$, the user to IRS channels in $\boldsymbol{H}_1 = [\boldsymbol{H}_{1,1}^{\mathrm{T}}, \ldots, \boldsymbol{H}_{1,K}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{N\times L}$, and $\boldsymbol{x} = [\boldsymbol{x}_1^{\mathrm{T}}, \ldots, \boldsymbol{x}_K^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^N$ with $N = \sum_{k=1}^{K} N_k$, the channel model in (2) can be obtained. Therefore, the channel estimation in the uplink can again be treated like a single-user MIMO problem.

As we discuss in the next section, it is not possible to estimate the two components $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ of the cascaded channel explicitly in the single-user cases. Unlike the direct channel $\boldsymbol{H}_0$, only the corresponding combination of the cascaded channels can be estimated. In the uplink, however, all the cascaded channels of the different users include the same component $\boldsymbol{H}_2$. Therefore, it is possible to find $\boldsymbol{H}_2$ and the components $\boldsymbol{H}_{1,1}, \ldots, \boldsymbol{H}_{1,K}$. For the later application of the estimates, however, the explicit knowledge of the components is not necessary.

3. PARAMETERS TO BE ESTIMATED

Consider the MIMO model in (2). The transmitted signal $\boldsymbol{x}$ is transformed by the composite channel

$$\boldsymbol{H}_{\text{comp}} = \boldsymbol{H}_0 + \boldsymbol{H}_2\boldsymbol{\Phi}\boldsymbol{H}_1^{\mathrm{T}} \tag{6}$$

i.e., $\boldsymbol{y}' = \boldsymbol{H}_{\text{comp}}\boldsymbol{x} + \boldsymbol{n}'$ [cf. (2)]. We do not consider any particular structure for the channels $\boldsymbol{H}_0$, $\boldsymbol{H}_1$, and $\boldsymbol{H}_2$, e.g., like sparsity, to keep the discussion as general as possible although such a structure might reduce the effort for channel estimation.

Note that $\boldsymbol{H}_{\text{comp}}$ depends on the factors $\boldsymbol{v}$ of the IRS. Due to the diagonal structure of the IRS representing matrix, i.e., $\boldsymbol{\Phi} = \mathrm{diag}(\boldsymbol{v})$, it is possible to include any arbitrary diagonal matrix $\boldsymbol{D} \in \mathbb{C}^{L\times L}$ comprising non-zero (not necessarily with unit magnitude) diagonal elements without changing the composite channel, i.e.,

$$\boldsymbol{H}_{\text{comp}} = \boldsymbol{H}_0 + \boldsymbol{H}_2\boldsymbol{D}^{-1}\boldsymbol{\Phi}\boldsymbol{D}\boldsymbol{H}_1^{\mathrm{T}}$$

showing that by introducing $\boldsymbol{H}'_2 = \boldsymbol{H}_2\boldsymbol{D}^{-1}$ and $\boldsymbol{H}'_1 = \boldsymbol{H}_1\boldsymbol{D}$, we can write $\boldsymbol{H}_{\text{comp}} = \boldsymbol{H}_0 + \boldsymbol{H}'_2\boldsymbol{\Phi}\boldsymbol{H}_1'^{,\mathrm{T}}$. This discussion highlights that it is impossible to estimate the true channels $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ separately due to the ambiguity by $\boldsymbol{D}$.

Let $\boldsymbol{h}_{1,i}$ and $\boldsymbol{h}_{2,i}$ denote the i-th column of $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$, respectively. Therefore, we can write

$$\boldsymbol{H}_{\text{comp}} = \boldsymbol{H}_0 + \sum_{i=1}^{L}\boldsymbol{h}_{2,i}\boldsymbol{h}_{1,i}^{\mathrm{T}}v_i.$$

We can see that only the knowledge of $\boldsymbol{h}_{2,i}\boldsymbol{h}_{1,i}^{\mathrm{T}}$ or its vectorization $\boldsymbol{h}_{1,i}\otimes\boldsymbol{h}_{2,i}$ (see [14, Theorem 3.4]) suffices to express the composite channel depending on the IRS factors $\boldsymbol{v}$ where '$\otimes$' denotes the Kronecker product. As all the relevant figures of merit, e.g., spectral efficiency or mean square error, depend on $\boldsymbol{H}_{\text{comp}}$ and not explicitly on $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$, it suffices to find the terms $\boldsymbol{h}_{1,i}\otimes\boldsymbol{h}_{2,i}\ \forall i = 1, \ldots, L$. Likewise, $h_{\text{comp}} = h_0 + \boldsymbol{h}_1^{\mathrm{T}}\boldsymbol{H}_2\boldsymbol{v}$ [see (3)], $\boldsymbol{h}_{\text{comp}} = \boldsymbol{h}_0 + \boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{v}$ [see (4)], and $\boldsymbol{h}_{\text{comp}}^{\mathrm{T}} = \boldsymbol{h}_0^{\mathrm{T}} + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{H}_2\boldsymbol{H}_1^{\mathrm{T}}$ [see (5)] for the SISO, SIMO, and MISO cases, respectively, exhibit the same properties, i.e., the components of the cascaded channels cannot be estimated separately due to the ambiguity by a diagonal $\boldsymbol{D}$.

We'd like to stress that the components $\boldsymbol{H}_0$, $\boldsymbol{H}_1$, and $\boldsymbol{H}_2$ must be estimated and not only $\boldsymbol{H}_{\text{comp}}$ [see (6)] such that the properties of the system can be influenced by the IRS factors $v_1, \ldots, v_L$.

4. CHANNEL ESTIMATION

For estimating the channel, it is necessary that the transmitter sends N_{p} pilots which we collect in $\boldsymbol{X} \in \mathbb{C}^{N_{\mathrm{p}}\times N}$. Therefore, the received signals for the N_{p} pilots read as

$$\boldsymbol{Y}' = \boldsymbol{H}_0\boldsymbol{X}^{\mathrm{T}} + \boldsymbol{H}_2\boldsymbol{\Phi}\boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}} + \boldsymbol{N}' \in \mathbb{C}^{M\times N_{\mathrm{p}}} \tag{7}$$

with the noise matrix $\boldsymbol{N}'$ whose zero-mean columns are assumed to have the same covariance matrix $\boldsymbol{C}_{\boldsymbol{n}}$ but are mutually independent. Vectorizing $\boldsymbol{Y}'$, i.e., $\boldsymbol{y} = \mathrm{vec}(\boldsymbol{Y}')$ gives

$$\begin{aligned}\boldsymbol{y} &= (\boldsymbol{X}\otimes\mathbf{I})\left(\mathrm{vec}(\boldsymbol{H}_0) + \mathrm{vec}\left(\boldsymbol{H}_2\boldsymbol{\Phi}\boldsymbol{H}_1^{\mathrm{T}}\right)\right) + \mathrm{vec}(\boldsymbol{N}')\\ &= (\boldsymbol{X}\otimes\mathbf{I})\boldsymbol{h}_0 + (\boldsymbol{X}\otimes\mathbf{I})(\boldsymbol{H}_1\circledast\boldsymbol{H}_2)\boldsymbol{v} + \boldsymbol{n} \in \mathbb{C}^{MN_{\mathrm{p}}}\end{aligned} \tag{8}$$

with $\boldsymbol{h}_0 = \mathrm{vec}(\boldsymbol{H}_0)$ and the column-wise Khatri-Rao product '$\circledast$'. For the second line, [14, Theorem 3.13] has been employed. Note that $\boldsymbol{n} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \mathbf{I}\otimes\boldsymbol{C}_{\boldsymbol{n}})$ due to the assumed stationarity of $\boldsymbol{n}'$. With

$$\boldsymbol{H} = [\boldsymbol{h}_0, \boldsymbol{H}_1\circledast\boldsymbol{H}_2] \in \mathbb{C}^{MN\times L+1} \tag{9}$$

(8) can be rewritten as

$$\boldsymbol{y} = (\boldsymbol{X}\otimes\mathbf{I})\boldsymbol{H}\boldsymbol{v}' + \boldsymbol{n}$$

where $\boldsymbol{v}' = [1, \boldsymbol{v}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{L+1}$.

For a fixed IRS phase allocation, suppose that the composite channel $\boldsymbol{H}_{\text{comp}}$ [see (6)] is estimated with the filter $\boldsymbol{G}_x \in \mathbb{C}^{N\times N_{\mathrm{p}}}$, i.e., $\hat{\boldsymbol{H}}_{\text{comp}} = \boldsymbol{Y}'\boldsymbol{G}_x^{\mathrm{T}}$ or $\hat{\boldsymbol{h}}_{\text{comp}} = (\boldsymbol{G}_x\boldsymbol{X}\otimes\mathbf{I})\boldsymbol{H}\boldsymbol{v}' + (\boldsymbol{G}_x\otimes\mathbf{I})\boldsymbol{n}$ after vectorization, that is, $\hat{\boldsymbol{h}}_{\text{comp}} = \mathrm{vec}(\hat{\boldsymbol{H}}_{\text{comp}})$.

To illuminate the whole channel, N_v different phase allocations are considered which we collect in

$$\boldsymbol{V} = [\boldsymbol{v}'_1, \ldots, \boldsymbol{v}'_{N_v}] \in \mathbb{C}^{L+1\times N_v}.$$

Note that $\boldsymbol{V}$ has unit-magnitude entries [cf. (1)] and its first row is the all-ones vector. Assuming the same pilots for all phase allocations, we get

$$\boldsymbol{Y} = (\boldsymbol{X}\otimes\mathbf{I})\boldsymbol{H}\boldsymbol{V} + \boldsymbol{N} \in \mathbb{C}^{MN_{\mathrm{p}}\times N_v}. \tag{10}$$

The noise matrix $\boldsymbol{N}$ comprises mutually independent zero-mean columns with covariance matrix $\mathbf{I} \otimes \boldsymbol{C}_{\boldsymbol{n}}$.

To estimate the complete channel matrix $\boldsymbol{H}$, the filter $\boldsymbol{G}_x$ must be applied as discussed above combined with the filter $\boldsymbol{G}_v \in \mathbb{C}^{N_v \times L+1}$, i.e.,

$$\begin{aligned}\hat{\boldsymbol{H}} &= (\boldsymbol{G}_x \otimes \mathbf{I})\boldsymbol{Y}\boldsymbol{G}_v \\ &= (\boldsymbol{G}_x\boldsymbol{X} \otimes \mathbf{I})\boldsymbol{H}\boldsymbol{V}\boldsymbol{G}_v + (\boldsymbol{G}_x \otimes \mathbf{I})\boldsymbol{N}\boldsymbol{G}_v.\end{aligned} \tag{11}$$

Note that

$$\mathrm{E}[\boldsymbol{N}\boldsymbol{G}_v\boldsymbol{G}_v^{\mathrm{H}}\boldsymbol{N}^{\mathrm{H}}] = (\mathbf{I} \otimes \boldsymbol{C}_{\boldsymbol{n}}) \operatorname{tr}(\boldsymbol{G}_v\boldsymbol{G}_v^{\mathrm{H}})$$

due to the independence of the columns of $\boldsymbol{N}$.

With similar steps, we can find for SISO systems,

$$\hat{h}^{\mathrm{T}} = \boldsymbol{g}_x^{\mathrm{T}}\boldsymbol{x}\boldsymbol{h}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{G}_v + \boldsymbol{g}_x^{\mathrm{T}}\boldsymbol{N}\boldsymbol{G}_v \tag{12}$$

where $\boldsymbol{g}_x \in \mathbb{C}^{N_{\mathrm{p}}}$, $\boldsymbol{h} = [h_0, \boldsymbol{h}_1^{\mathrm{T}}\boldsymbol{H}_2]$, and $\boldsymbol{H}_2 = \operatorname{diag}(\boldsymbol{h}_2)$. For a SIMO system, we get

$$\hat{\boldsymbol{H}} = (\boldsymbol{G}_x\boldsymbol{x} \otimes \mathbf{I})\boldsymbol{H}\boldsymbol{V}\boldsymbol{G}_v + (\boldsymbol{G}_x\boldsymbol{x} \otimes \mathbf{I})\boldsymbol{N}\boldsymbol{G}_v \tag{13}$$

with $\boldsymbol{H} = [\boldsymbol{h}_0, \boldsymbol{h}_1^{\mathrm{T}} \circledast \boldsymbol{H}_2]$ and for a MISO system,

$$\hat{\boldsymbol{H}} = \boldsymbol{G}_x\boldsymbol{X}\boldsymbol{H}\boldsymbol{V}\boldsymbol{G}_v + \boldsymbol{G}_x\boldsymbol{N}\boldsymbol{G}_v \tag{14}$$

where $\boldsymbol{H} = [\boldsymbol{h}_0, \boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{H}_2]$ and $\boldsymbol{H}_2 = \operatorname{diag}(\boldsymbol{h}_2)$. We observe that the expressions for the different channels types are similar. Thus, we concentrate on the MIMO case in the following.

4.1. Least Squares Channel Estimation

The LS estimate tries to match to the model in (10), that is,

$$\hat{\boldsymbol{H}}_{\mathrm{LS}} = \underset{\boldsymbol{H}}{\operatorname{argmin}} \left\|\boldsymbol{Y} - (\boldsymbol{X} \otimes \mathbf{I})\boldsymbol{H}\boldsymbol{V}\right\|_{\mathrm{F}}^2. \tag{15}$$

The solution reads as

$$\hat{\boldsymbol{H}}_{\mathrm{LS}} = \left(\boldsymbol{X}^{+} \otimes \mathbf{I}\right)\boldsymbol{Y}\boldsymbol{V}^{+}$$

with the Moore-Penrose pseudoinverse $\boldsymbol{A}^{+}$ of $\boldsymbol{A}$. For full-rank and tall $\boldsymbol{X}$, $\boldsymbol{X}^{+} = (\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{H}}$, and for full-rank and wide $\boldsymbol{V}$, $\boldsymbol{V}^{+} = \boldsymbol{V}^{\mathrm{H}}(\boldsymbol{V}\boldsymbol{V}^{\mathrm{H}})^{-1}$. Thus,

$$\boldsymbol{G}_{x,\mathrm{LS}} = (\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{H}} \qquad \boldsymbol{G}_{v,\mathrm{LS}} = \boldsymbol{V}^{\mathrm{H}}(\boldsymbol{V}\boldsymbol{V}^{\mathrm{H}})^{-1}. \tag{16}$$

The mean square error (MSE) of the unbiased LS estimate is

$$\begin{aligned}\mathrm{E}[\|\boldsymbol{H} - \hat{\boldsymbol{H}}_{\mathrm{LS}}\|_{\mathrm{F}}^2] &= \mathrm{E}[\|(\boldsymbol{G}_{x,\mathrm{LS}} \otimes \mathbf{I})\boldsymbol{N}\boldsymbol{G}_{v,\mathrm{LS}}\|_{\mathrm{F}}^2] \\ &= \operatorname{tr}(\boldsymbol{G}_x\boldsymbol{G}_x^{\mathrm{H}} \otimes \boldsymbol{C}_{\boldsymbol{n}}) \operatorname{tr}(\boldsymbol{G}_v\boldsymbol{G}_v^{\mathrm{H}}) \\ &= \operatorname{tr}((\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X})^{-1}) \operatorname{tr}(\boldsymbol{C}_{\boldsymbol{n}}) \operatorname{tr}((\boldsymbol{V}\boldsymbol{V}^{\mathrm{H}})^{-1})\end{aligned}$$

which, due to the unbiasedness of $\hat{\boldsymbol{H}}_{\mathrm{LS}}$, is independent of the statistical properties of $\boldsymbol{H}$. As shown in Appendix A, the minimum of the MSE under the constraints $\operatorname{tr}(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}) = P_{\mathrm{tx}}$ and $\operatorname{tr}(\boldsymbol{V}\boldsymbol{V}^{\mathrm{H}}) = (L+1)^2$ can be reached by choosing $\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X} = \frac{P_{\mathrm{tx}}}{N}\mathbf{I}$ and $\boldsymbol{V}\boldsymbol{V}^{\mathrm{H}} = (L+1)\mathbf{I}$. To minimize also the time necessary for training, both, the pilot matrix $\boldsymbol{X}$ and the phase allocation matrix $\boldsymbol{V}$, should be square, i.e., $\boldsymbol{X} \in \mathbb{C}^{N \times N}$ and $\boldsymbol{V} \in \mathbb{C}^{L+1 \times L+1}$. Therefore, the MSE with LS channel estimation is minimized by scaled unitary $\boldsymbol{X}$ and $\boldsymbol{V}$. Any scaled unitary $N \times N$ matrix can be used for $\boldsymbol{X}$, e.g., the scaled DFT matrix $\sqrt{\frac{P_{\mathrm{tx}}}{N}}\boldsymbol{F}_N$. For $\boldsymbol{V}$, the scaled $L+1 \times L+1$ DFT matrix $\sqrt{L+1}\boldsymbol{F}_{L+1}$ is particularly suited as all of its elements have unit magnitudes and its first row is the all-ones vector.

4.2. Minimum Mean Square Error Channel Estimation

The combined channel $\boldsymbol{H}$ is assumed to be zero-mean and has the covariance matrix $\mathrm{E}[\boldsymbol{H}\boldsymbol{A}\boldsymbol{H}^{\mathrm{H}}] = \operatorname{tr}(\boldsymbol{A})\boldsymbol{R}_{\boldsymbol{H}}$ for any $\boldsymbol{A} \in \mathbb{C}^{L+1 \times L+1}$ due to the independence of the columns of $\boldsymbol{H}$, it is possible to formulate the MMSE optimization

$$(\boldsymbol{G}_{x,\mathrm{MMSE}}, \boldsymbol{G}_{v,\mathrm{MMSE}}) = \underset{(\boldsymbol{G}_x,\boldsymbol{G}_v)}{\operatorname{argmin}} \mathrm{E}[\|\boldsymbol{H} - \hat{\boldsymbol{H}}\|_{\mathrm{F}}^2] \tag{17}$$

with the MSE $= \mathrm{E}[\|\boldsymbol{H} - \hat{\boldsymbol{H}}\|_{\mathrm{F}}^2]$. Unfortunately, the MMSE optimization in (17) is non-convex and difficult to solve. Consequently, we use the LS solution for $\boldsymbol{G}_{v,\mathrm{LS}} = \boldsymbol{V}(\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V})^{-1}$ [see (16)]. The simplified MSE reads as

$$\begin{aligned}\mathrm{MSE} = \operatorname{tr}(\boldsymbol{R}_{\boldsymbol{H}}) &- 2(L+1)\operatorname{Re}\{\operatorname{tr}((\boldsymbol{G}_x\boldsymbol{X} \otimes \mathbf{I})\boldsymbol{R}_{\boldsymbol{H}})\} \\ &+ (L+1)\operatorname{tr}((\boldsymbol{X}^{\mathrm{H}}\boldsymbol{G}_x^{\mathrm{H}}\boldsymbol{G}_x\boldsymbol{X} \otimes \mathbf{I})\boldsymbol{R}_{\boldsymbol{H}}) \\ &+ \operatorname{tr}(\boldsymbol{G}_x\boldsymbol{G}_x^{\mathrm{H}} \otimes \boldsymbol{C}_{\boldsymbol{n}}) \operatorname{tr}((\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V})^{-1})\end{aligned}$$

Based on Appendix A, the MSE is minimized w.r.t. the phase allocations under the constraint $\operatorname{tr}(\boldsymbol{V}\boldsymbol{V}^{\mathrm{H}}) = (L+1)^2$ by $\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V} = (L+1)\mathbf{I}$ which suggests that $\boldsymbol{V} = \sqrt{L+1}\boldsymbol{F}_{L+1}$ is again well suited with the DFT matrix $\boldsymbol{F}_{L+1}$. Note that $\operatorname{tr}((\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V})^{-1}) = 1$ since $(\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V})^{-1} = \frac{1}{L+1}\mathbf{I}$.

For finding $\boldsymbol{G}_x$, we apply an approximation as discussed in [15], i.e., we approximate $\boldsymbol{R}_{\boldsymbol{H}}$ by $\boldsymbol{R}_1 \otimes \boldsymbol{R}_2$ with positive definite $\boldsymbol{R}_1 \in \mathbb{C}^{N \times N}$ and $\boldsymbol{R}_2 \in \mathbb{C}^{M \times M}$. Thus, the MSE can be approximated as

$$\begin{aligned}\mathrm{MSE} \approx \operatorname{tr}(\boldsymbol{R}_{\boldsymbol{H}}) &- 2(L+1)\operatorname{Re}\{\operatorname{tr}(\boldsymbol{G}_x\boldsymbol{X}\boldsymbol{R}_1)\operatorname{tr}(\boldsymbol{R}_2)\} \\ &+ (L+1)\operatorname{tr}(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{G}_x^{\mathrm{H}}\boldsymbol{G}_x\boldsymbol{X}\boldsymbol{R}_1)\operatorname{tr}(\boldsymbol{R}_2) + \|\boldsymbol{G}_x\|_{\mathrm{F}}^2 \operatorname{tr}(\boldsymbol{C}_{\boldsymbol{n}}).\end{aligned}$$

The optimal choice for $\boldsymbol{G}_x$ minimizing the MSE reads as

$$\boldsymbol{G}_{x,\mathrm{MMSE}} = \boldsymbol{R}_1\boldsymbol{X}^{\mathrm{H}}\left(\boldsymbol{X}\boldsymbol{R}_1\boldsymbol{X}^{\mathrm{H}} + \gamma^{-1}\mathbf{I}\right)^{-1}$$

with $\gamma = (L+1)\operatorname{tr}(\boldsymbol{R}_2)/\operatorname{tr}(\boldsymbol{C}_{\boldsymbol{n}})$, leading to

$$\mathrm{MMSE} = \operatorname{tr}\left(\left(\boldsymbol{R}_1^{-1} + \gamma\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\right)^{-1}\right). \tag{18}$$

With the eigenvalue decomposition (EVD) $\boldsymbol{R}_1 = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{H}}$, we find

$$\mathrm{MMSE} = \frac{1}{\gamma}\operatorname{tr}(((\gamma\boldsymbol{D})^{-1} + \boldsymbol{U}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\boldsymbol{U})^{-1}).$$

It is optimum to set $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\boldsymbol{U} = \max(\mu'\mathbf{I} - (\gamma\boldsymbol{D})^{-1}, \mathbf{0})$ to further minimize the MMSE (see Appendix B). Hence, $\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\varXi}\boldsymbol{U}^{\mathrm{H}}$ with $\boldsymbol{\varXi} = \max(\mu'\mathbf{I} - (\gamma\boldsymbol{D})^{-1}, \mathbf{0})$ where μ' must be chosen such that $\operatorname{tr}(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}) = \operatorname{tr}(\boldsymbol{\varXi}) = P_{\mathrm{tx}}$. Thus, a possible choice for the MMSE minimizing pilots is

$$\boldsymbol{X} = \boldsymbol{\varXi}^{\frac{1}{2}}\boldsymbol{U}^{\mathrm{H}} \in \mathbb{C}^{N \times N} \tag{19}$$

which highlights that the length of the training sequences N_{p} doesn't need to be larger than the number of transmitting antennas N. Similar results for MMSE pilots in MIMO systems are given in [16, 17].

5. SIMULATION RESULTS

For the numerical analysis, we consider a base station (BS) equipped with a uniform linear array comprising $N = 8$ antennas and the elements at the IRS are arranged as a square ($8 \times 8 = 64$ elements). The path loss is modelled by $\mathrm{PL} = \mathrm{PL}_0 + \alpha_{\mathrm{PL}}10\log_{10}(d/d_0) + X_{\mathrm{g}}$,

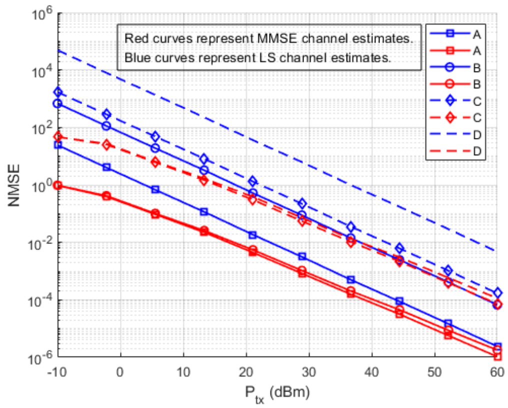

Fig. 1. Comparison of Optimized and Non-Optimized Training

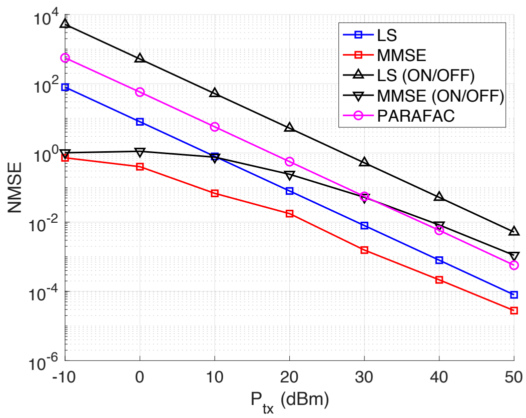

Fig. 2. Comparison of IRS channel estimation algorithms

where $\alpha = 3.7$ for the link between the BS and the user, $\alpha = 2.0$ between the BS and the IRS, and $\alpha = 2.8$ for the channel between the IRS and the user. The reference distance is 1 m and the corresponding reference path loss is $\mathrm{PL}_0 = -10\,\mathrm{dB}$. The attenuation caused by shadow fading is $X_\mathrm{g} = -30\,\mathrm{dB}$ and the noise variance is $\sigma^2 = -80\,\mathrm{dBm}$. The BS is located at $(0, 0, 10)$ m, the IRS at $(50, 0, 10)$ m, and the UE at $(40, 0, 0)$ m. Additionally, the direct path suffers from a shadowing effect and has a 30 dB attenuation. The noise covariance matrix is $\boldsymbol{C_n} = \mathbf{I}$. The covariance matrix $\boldsymbol{R_H}$ results from the assumed angular spread of 2°.

Different combinations of optimized (OPT) and non-optimized (NOPT) training allocations are presented in Fig. 1 where the normalized mean square errors (NMSE) of LS and MMSE estimators are presented, defined as $\mathrm{NMSE} = \mathrm{E}[\|\boldsymbol{H} - \hat{\boldsymbol{H}}\|_\mathrm{F}^2 / \|\boldsymbol{H}\|_\mathrm{F}^2]$. Case A corresponds to OPT pilots and OPT phase allocations that lead to the best results for LS and MMSE estimation. Using random (NOPT) pilots with OPT phase allocations (Case B) has little effect for MMSE estimation but reduces the performance of LS channel estimation considerably. Employing OPT pilots and random phase shifts (Case C) has a detrimental effect for both estimation schemes. Case D with NOPT pilots and NOPT phase shifts gives the worst NMSE. These results illustrate the importance of optimizing the pilots and the phase allocations during training.

In Fig. 2, the NMSEs resulting from the different channel estimation methods are shown. Firstly, it can be observed that the NMSE for a method with ON/OFF (e.g., [4, 5]). operation is clearly higher than that for the corresponding channel estimation without ON/OFF procedure. Secondly, even optimized LS outperforms PARAFAC which is based on a tensor decomposition and partially employs LS [6]. Additionally, the MMSE channel estimators outperform the corresponding LS estimators due to the application of the covariance information.

A. MINIMUM OF TRACE OF INVERSE

Consider the minimization of the trace of the inverse under a trace constraint, i.e.,

$$\boldsymbol{A}_\mathrm{opt} = \operatorname*{argmin}_{\boldsymbol{A}} \mathrm{tr}(\boldsymbol{A}^{-1}) \quad \text{s.t. } \mathrm{tr}(\boldsymbol{A}) = n^2 \text{ and } \boldsymbol{A} \succ \mathbf{0} \tag{20}$$

with $\boldsymbol{A} \in \mathbb{C}^{n \times n}$. This optimization is convex since the constraints are linear and the cost function is convex since $\frac{\partial^2}{\partial t^2}\mathrm{tr}((\boldsymbol{A} + t\boldsymbol{B})^{-1})|_{t=0} > 0$ for $\boldsymbol{A} \succ \mathbf{0}$ and $\boldsymbol{B} = \boldsymbol{B}^\mathrm{H}$.

From the corresponding Lagrangian function

$$L(\boldsymbol{A}, \mu, \boldsymbol{\Lambda}) = \mathrm{tr}(\boldsymbol{A}^{-1}) + \mu(\mathrm{tr}(\boldsymbol{A}) - n^2) - \mathrm{tr}(\boldsymbol{\Lambda}\boldsymbol{A})$$

where $\mu \in \mathbb{R}$ and $\boldsymbol{\Lambda} \succeq \mathbf{0}$, we obtain by differentiating w.r.t. $\boldsymbol{A}^\mathrm{T}$ that $-\boldsymbol{A}^{-2} + \mu\mathbf{I} - \boldsymbol{\Lambda} = \mathbf{0}$. Multiplying by $\boldsymbol{A}^2$ yields $\boldsymbol{A}^2 = \frac{1}{\mu}\mathbf{I}$. With the constraint $\mathrm{tr}(\boldsymbol{A}) = n^2$, we finally obtain

$$\boldsymbol{A}_\mathrm{opt} = n\mathbf{I}. \tag{21}$$

B. MINIMUM OF TRACE OF INVERSE WITH ADDITIONAL DIAGONAL TERM

By the following optimization, the trace of the inverse of the sum of the non-negative definite matrix $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ and a positive definite diagonal matrix $\boldsymbol{\Psi} \in \mathbb{R}^{n \times n}$ is minimized

$$\boldsymbol{A}_\mathrm{opt} = \operatorname*{argmin}_{\boldsymbol{A}} \mathrm{tr}((\boldsymbol{\Psi} + \boldsymbol{A})^{-1}) \text{ s.t. } \mathrm{tr}(\boldsymbol{A}) = \alpha \text{ and } \boldsymbol{A} \succeq \mathbf{0}. \tag{22}$$

This optimization is convex due to the convexity of the cost function combined with linear constraints.

Differentiating the corresponding Lagrangian function w.r.t. $\boldsymbol{A}^\mathrm{T}$, where $\mu \in \mathbb{R}$ and $\boldsymbol{\Lambda} \succeq \mathbf{0}$ are the Lagrangian multipliers, gives $-(\boldsymbol{\Psi} + \boldsymbol{A})^{-2} + \mu\,\mathbf{I} - \boldsymbol{\Lambda} = \mathbf{0}$. The multiplication by $\boldsymbol{\Psi} + \boldsymbol{A}$ leads to $-(\boldsymbol{\Psi} + \boldsymbol{A})^{-1} + \mu(\boldsymbol{\Psi} + \boldsymbol{A}) - \boldsymbol{\Psi}\boldsymbol{\Lambda} = \mathbf{0}$ where we have exploited $\boldsymbol{A}\boldsymbol{\Lambda} = \mathbf{0}$. Note that $\boldsymbol{\Psi}\boldsymbol{\Lambda}$ must be Hermitian since all other terms are Hermitian. It can be inferred that $\boldsymbol{\Lambda}$ must be diagonal if the diagonal elements of $\boldsymbol{\Psi}$ are different. Multiplying by $\boldsymbol{\Psi} + \boldsymbol{A}$ another time gives the expression $-\mathbf{I} + \mu(\boldsymbol{\Psi} + \boldsymbol{A})^2 - \boldsymbol{\Psi}^2\boldsymbol{\Lambda} = \mathbf{0}$ which shows that $\boldsymbol{A}$ must be diagonal as all other terms are diagonal. Consequently, we have obtained following scalar equations

$$-1 + \mu(\psi_i + a_i)^2 - \psi_i\lambda_i = 0 \quad \forall i \in \{1, \ldots, n\}$$

where $a_i \geqslant 0$ is the i-th diagonal element of $\boldsymbol{A}$, $\psi_i > 0$ denotes the i-th diagonal element of $\boldsymbol{\Psi}$, and $\lambda_i \geqslant 0$ is the i-th diagonal element of $\boldsymbol{\Lambda}$. Since $\lambda_i = 0$ if $a_i \neq 0$, $a_i = \max(\mu' - \psi_i, 0)$ with $\mu' = 1/\sqrt{\mu}$. Assuming the optimal number of non-zero diagonal elements of $\boldsymbol{A}$ is n_opt, the Lagrangian multiplier is given by $\mu' = \frac{\alpha}{n_\mathrm{opt}} + \frac{1}{n_\mathrm{opt}}\sum_{j=1}^{n_\mathrm{opt}} \psi_j$ resulting from $\mathrm{tr}(\boldsymbol{A}) = \alpha$. Therefore, the optimal solution to (22) reads as

$$\boldsymbol{A} = \max(\mu'\mathbf{I} - \boldsymbol{\Psi}, \mathbf{0}) \tag{23}$$

where the element-wise maximum is taken by $\max(\bullet, \bullet)$.

C. REFERENCES

[1] Hexa-X, "Deliverable D4.1 AI-driven communication & computation co-design: Gap analysis and blueprint," Tech. Rep. 101015956, 5G PPP, 2021.

[2] D. Mishra and H. Johansson, "Channel Estimation and Low-Complexity Beamforming Design for Passive Intelligent Surface Assisted MISO Wireless Energy Transfer," in *Proc. 44th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2019)*, May 2019, pp. 4659–4663.

[3] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent Reflecting Surface Meets OFDM: Protocol Design and Rate Maximization," *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4522–4535, July 2020.

[4] T. L. Jensen and E. De Carvalho, "An Optimal Channel Estimation Scheme for Intelligent Reflecting Surfaces Based on a Minimum Variance Unbiased Estimator," in *Proc. 45th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2020)*, May 2020, pp. 5000–5004.

[5] Z. Wang, L. Liu, and S. Cui, "Channel Estimation for Intelligent Reflecting Surface Assisted Multiuser Communications: Framework, Algorithms, and Analysis," *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6607–6620, October 2020.

[6] L. Wei, C. Huang, G. C. Alexandropoulos, and C. Yuen, "Parallel Factor Decomposition Channel Estimation in RIS-Assisted Multi-User MISO Communication," in *Proc. 2020 IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, June 2020.

[7] M. Guo and M. Cenk Gursoy, "Channel Estimation for Intelligent Reflecting Surface Assisted Wireless Communications," in *Proc. 2021 IEEE Wireless Communications and Networking Conference (WCNC)*, March 2021.

[8] C. You, B. Zheng, and R. Zhang, "Channel Estimation and Passive Beamforming for Intelligent Reflecting Surface: Discrete Phase Shift and Progressive Refinement," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2604–2620, November 2020.

[9] Y. Pan and Z. Deng, "Channel Estimation for Wireless Communication Systems Aided by Large Intelligent Reflecting Surface," in *Proc. 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, March 2021.

[10] C. You, B. Zheng, and R. Zhang, "Intelligent Reflecting Surface Assisted Multi-User OFDMA: Channel Estimation and Training Design," *IEEE Transactions on Wireless Communications*, vol. 19, no. 12, pp. 8315–8329, December 2020.

[11] Z. Zhou, N. Ge, Z. Wang, and L. Hanzo, "Joint Transmit Precoding and Reconfigurable Intelligent Surface Phase Adjustment: A Decomposition-Aided Channel Estimation Approach," *IEEE Transactions on Communications*, vol. 69, no. 2, pp. 1228–1243, February 2021.

[12] Q. Nadeem, H. Alwazani, A. Kammoun, A. Chaaban, M. Debbah, and M. Alouini, "Intelligent Reflecting Surface-Assisted Multi-User MISO Communication: Channel Estimation and Beamforming Design," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 661–680, May 2020.

[13] G. T. de Araújo, A. L. F. de Almeida, and R. Boyer, "Channel Estimation for Intelligent Reflecting Surface Assisted MIMO Systems: A Tensor Modeling Approach," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 789–802, April 2021.

[14] J. W. Brewer, "Kronecker Products and Matrix Calculus in System Theory," *IEEE Transactions on Circuits and Systems*, vol. CAS-25, no. 9, pp. 772–781, September 1978.

[15] C.F. Van Loan and N. Pitsianis, "Approximation with Kronecker Products," in *Linear Algebra for Large Scale and Real-Time Applications*, M. S. Moonen, G. H. Golub, and B. L. R. De Moor, Eds., vol. 232 of *NATO ASI Series (Series E: Applied Sciences)*, pp. 293–314. Springer, Dordrecht, 1993.

[16] J. Choi and Y. Lee, "Optimum Pilot Pattern for MMSE Channel Estimation in Single-Carrier MIMO Systems," in *Proc. IEEE 60th Vehicular Technology Conference, 2004 (VTC2004-Fall)*, September 2004, pp. 1382–1385.

[17] E. Björnson and B. Ottersten, "A Framework for Training-Based Estimation in Arbitrarily Correlated Rician MIMO Channels With Rician Disturbance," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1807–1820, March 2010.