THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Adaptive LS and MMSE based Beamformer Design for Multiuser MIMO Interference Channels

Citation for published version:
Ratnarajah, T 2016, 'Adaptive LS and MMSE based Beamformer Design for Multiuser MIMO Interference Channels', *IEEE Transactions on Vehicular Technology*.

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
IEEE Transactions on Vehicular Technology

OPEN ACCESS

Adaptive LS and MMSE based Beamformer Design for Multiuser MIMO Interference Channels

S. Morteza Razavi, Tharmalingam Ratnarajah, *Senior Member, IEEE*

Abstract—In the presence of perfect channel state information (CSI), the achievable degrees of freedom (DoF) in wireless interference networks can be linearly scaled up with the number of users. Achievability is based on the idea of interference alignment (IA). However, in the presence of imperfect CSI, the sum rate becomes degraded and full DoF may not be achievable anymore. In this paper, we propose novel least squares (LS) and minimum mean square error (MMSE) based IA schemes which adaptively design beamformers by relying on the availability of imperfect CSI and the knowledge of the channel estimation error variance in advance. Interestingly and unlike the other robust algorithms, the proposed adaptive schemes do not impose extra computational complexity compared to the nonadaptive ones. It is shown that the LS based IA is able to outperform interference leakage minimization algorithms under both perfect and imperfect CSI. Furthermore, we compare the performance of the proposed MMSE based IA with maximum signal-to-interference-plus-noise ratio (Max-SINR) algorithm. We show that while under perfect CSI the MMSE based IA achieves the same performance as that of Max-SINR, the former outperforms the latter under CSI mismatch. Meanwhile, it is shown that the proposed MMSE based IA needs less CSI to be available and has less computational complexity compared to Max-SINR.

Keywords—Adaptive beamformer design, interference channels, least squares, minimum mean square error, multiuser MIMO.

I. Introduction

INTERFERENCE alignment (IA) is an appealing interference management technique since it enables us to achieve significant throughput in wireless interference networks such that the total number of degrees of freedom (DoF) can be linearly scaled up with the number of users. This is in contrast to the orthogonal medium access techniques like time division multiple access (TDMA) and frequency division multiple access (FDMA) where the total DoF is one. In [1], it has been shown that in a K-user interference channel (IC) with a single antenna at each node, and with time-varying or frequency-selective channel coefficients, it is possible to achieve $\frac{K}{2}$ DoF by coding across sufficiently large symbol extension of the

channel. However, instead of aligning interfering signals in time and by deploying multiple antennas at transmit/receive nodes, it is possible to achieve IA without the need of symbol extension.

Although there are various IA algorithms based on the concept of signal space alignment offered by multiple antennas (see e.g., [2]–[13]), in this paper, we place our focus on designing two distinguishable IA algorithms based on least squares (LS) and minimum mean square error (MMSE) criteria which are hereafter referred to as LS and MMSE based designs for multiuser MIMO IC. Unlike standard IA methods were primarily designed based on the availability of perfect CSI, the optimization criteria of the proposed adaptive algorithms are set up based on the availability of imperfect channel estimations without imposing extra computational complexity compared to the nonadaptive ones. This makes the proposed algorithms be *adaptive* in a sense that the beamformers can be designed with the knowledge of the channel estimation error variance in advance. In this case, the beamformer design under perfect CSI becomes a special scenario when the error variance is set to zero. This makes the proposed schemes more general than standard IA techniques. It is worthwhile to note that although there are several improved IA methods under the assumption of imperfect CSI in the literature (see e.g., [13]–[18]), almost all of them just considered the digital feedback which is a very special case of the imperfect CSI scenario discussed in this work. More importantly and unlike the other methods, the proposed algorithms with imperfect CSI do not introduce any extra computational complexity compared to the case of perfect CSI.

Most of the previously proposed IA algorithms are closely related to either minimum weighted leakage interference (Min-WLI) or maximum signal-to-interference-plus-noise ratio (Max-SINR) algorithms proposed in [2]. For example, alternating minimization (Alt-Min) algorithm defined in [9] is algorithmically identical to Min-WLI such that both are categorized as *interference leakage minimization* algorithms since their goal is to minimize the leakage interference within the desired signal subspaces without considering the effect of noise. Also, Max-SINR is a special case of the MMSE IA in [11] and the weighted MMSE IA defined in [3]. Therefore without loss of generality and in the remainder of the paper, we compare the performance of the proposed schemes with these two representative IA algorithms, i.e., Min-WLI and Max-SINR.

It is shown that the proposed LS based IA, which similarly does not consider the effect of noise for beamformer design, is able to outperform interference leakage minimization algorithms under both perfect and imperfect CSI. This is due to

the fact that unlike Min-WLI and Alt-Min algorithms which do not consider the direct links to design the beamformers, the proposed LS based scheme does so.

The idea of MMSE IA was first introduced in [11] for single-stream-per-user transmission, and then generalized to the case of multi-stream-per-user communication (see e.g., [3]). In this paper, however, we propose a novel MMSE based IA which yields unitary precoders and combiners. Since the proposed approach is implicitly built on MMSE criterion to design beamformers, it might be slightly suboptimal compared to the weighted MMSE IA techniques as discussed in [3], [11]. However, note that the precoders obtained by the weighted MMSE IA are not unitary, and they further need an extra optimization step to meet the power constraint. Moreover, this power-constraint optimization step has no closed-form solution and has to be done numerically within *each iteration*. Nevertheless, the beamformers obtained by the proposed MMSE based IA are unitary, and consequently they do not require such an extra power-constraint optimization step within each iteration, which eventually results in simpler implementation. Furthermore, the employment of the unitary beamformers facilitates the usage of the proposed algorithms in realistic scenarios including digital feedback due to the ease in codebook design.

Compared to Max-SINR, while the proposed MMSE based IA is a user-by-user approach, the former is a stream-by-stream approach as explained in [19]. Consequently, it is shown that the proposed MMSE based IA needs less CSI to be available compared to Max-SINR in order to calculate the beamformers. Plus, the former possesses less computational complexity compared to the latter. We also prove that the proposed MMSE based design achieves the same performance as Max-SINR under perfect CSI. Subject to imperfect CSI, however, the former outperforms the latter.

Moreover, it is shown that under perfect CSI, the proposed LS based IA results in diagonalized subchannels for all SNR ranges. In other words, after premultiplying the received signal by the corresponding combiner, the interferences are first suppressed and the decoupled subchannels are then diagonalized. This is in contrast to the previously proposed IA schemes like interference leakage minimization algorithms wherein the resulted subchannels are full matrices. In this case and in order to employ waterfilling, while standard IA schemes like Min-WLI and Alt-Min need to use singular value decomposition (SVD), the proposed LS based IA dissolves the need of such decomposition since the decoupled subchannels have been already diagonalized. It is also worthwhile to point out that since at sufficiently high SNRs, MMSE based IA boils down to LS based IA, the resulted subchannels of the proposed MMSE based IA are also diagonalized at high enough SNRs.

This paper is organized as follows: Section II presents the basic concepts regarding the perfect and imperfect CSI models as well as the signal postprocessing at receive nodes. In Sections III and IV, respectively, we propose LS and MMSE based IA algorithms by setting up the optimization problem based on the knowledge of imperfect CSI and channel estimation error variance. Section V contains some discussions towards the behavior of the proposed algorithms. In Section VI, we use numerical results to show that the proposed schemes outperform standard IA techniques under both perfect and imperfect CSI. Finally the paper ends up with conclusions in Section VII.

A. Notations

Throughout the paper, a is a scalar, $\mathbf{a}$ is a vector, and $\mathbf{A}$ is a matrix. The superscript $(\cdot)^{\mathsf{H}}$ represents the Hermitian transpose. $\mathbb{E}\{\cdot\}$ and $\mathrm{Tr}\,[\cdot]$ are the expectation and trace operators, respectively. While $\|\cdot\|_2$ denotes the vector 2-norm, "$\Longleftarrow$" designates assignment through an "in-place" manner. $\mathrm{orth}\,(\mathbf{A})$ represents the unitary part of the orthogonal-triangular (QR) decomposition of $\mathbf{A}$. Furthermore, we define the vectorization and determinant operators as $\mathrm{vec}(\cdot)$ and $\det(\cdot)$, respectively.

II. PRELIMINARIES

A. System Model and Standard IA

We consider a symmetric K-user MIMO interference channel with $2K$ nodes, K of which are transmitters while the other K are receivers. Each transmitter intends to communicate with a single receiver in a one-to-one mapping as illustrated in Fig. 1. More specifically, each transmitter with N antennas communicates with its corresponding M-antenna receiver by sending d independent data streams. The channel output at the k-th receiver can be shown as

$$\mathbf{y}_k = \mathbf{H}_{k,k}\mathbf{x}_k + \sum_{\substack{j=1 \\ j\neq k}}^{K} \mathbf{H}_{k,j}\mathbf{x}_j + \mathbf{z}_k \ , \qquad (1)$$

where $\mathbf{y}_k \in \mathbb{C}^{M\times 1}$ is the received signal at node k, $\mathbf{x}_k \in \mathbb{C}^{N\times 1}$ is the transmitted signal from the k-th transmitter and $\mathbf{x}_j \in \mathbb{C}^{N\times 1}$ is the interference caused by transmitter j. $\mathbf{H}_{k,j} \in \mathbb{C}^{M\times N}$ denotes the channel from the j-th transmitter to the k-th receiver. The magnitudes of the fading coefficients are considered to be nonzero finite constants. We also assume a block fading model where all links are static for the duration of a transmission but may change between successive transmission, i.e., constant MIMO scenario. More specifically, the entries of channel matrices between transmitter j and receiver k can be modeled by i.i.d. Gaussian random variables with zero mean and unit variance, i.e., $\mathrm{vec}\,(\mathbf{H}_{k,j}) \sim \mathcal{N}_{\mathbb{C}}\,(\mathbf{0}, \mathbf{I})$. $\mathbf{z}_k \in \mathbb{C}^{M\times 1}$ represents the circularly symmetric additive white Gaussian noise with zero mean and variance σ^2 per entry, i.e., $\mathbf{z}_k \sim \mathcal{N}_{\mathbb{C}}\,(\mathbf{0}, \sigma^2\mathbf{I})$. Each transmitted signal $\mathbf{x}_k$ is equal to $\mathbf{V}_k\mathbf{c}_k$ where $\{\mathbf{V}_k\}_{k=1}^{K} \in \mathbb{C}^{N\times d}$ are the truncated unitary transmit beamforming matrices (precoders), and $\mathbf{c}_k \in \mathbb{C}^{d\times 1}$ is the data stream meant for the k-th receiver such that $\mathbb{E}\,\{\mathbf{c}_k\mathbf{c}_k^{\mathsf{H}}\} = P\mathbf{I}$. Without losing the generality, we assume uniform power allocation across all data streams of all users which is asymptotically optimal. We also define $\gamma = P/\sigma^2$ as the nominal SNR throughout the paper.

By considering a ZF based receiver, the conditions for perfect IA are [2], [20]

$$\mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,j}\mathbf{V}_j = \mathbf{0} \ , \quad \forall j \neq k \ , \qquad (2)$$

$$\mathrm{rank}\left(\mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{V}_k\right) = d \ , \qquad (3)$$
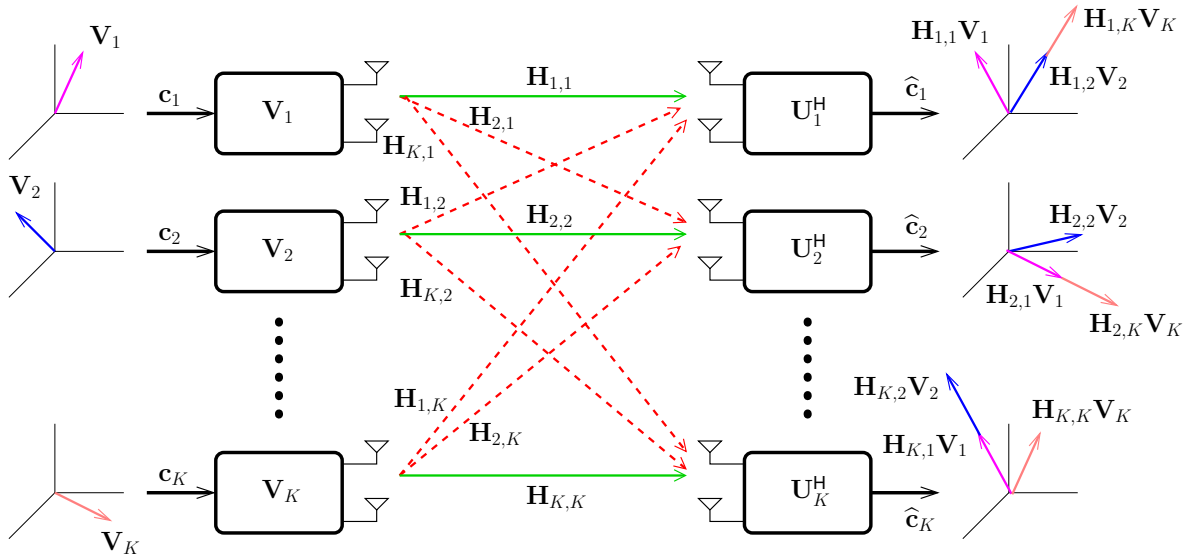
Fig. 1. K-user interference channel where solid green arrows denote direct links and dash red arrows represent interfering (cross) links. $\mathbf{c}_k$ designates the input data of transmit node k whereas $\widehat{\mathbf{c}}_k$ indicates the recovered data at receive node k.

where $\{\mathbf{U}_j\}_{j=1}^{K}$ designate the truncated unitary interference suppression matrices (combiners).

In other words, IA aims to design precoders $\mathbf{V}_k$ such that at each receive node, the unwanted interferences are aligned within a reduced subspace of the received signal space which is supposed to be independent of the desired signal subspace. This is shown in Fig. 1 wherein, for example, at the first receive node the interference subspaces $\mathbf{H}_{1,2}\mathbf{V}_2$ and $\mathbf{H}_{1,K}\mathbf{V}_K$ are aligned and are independent of the desired signal subspace $\mathbf{H}_{1,1}\mathbf{V}_1$. Therefore, with respect to (2), the first receive node premultiplies its received signal with $\mathbf{U}_1^{\mathsf{H}}$ which nulls out the aligned interference without suppressing the desired signal.

B. Imperfect CSI Model

Similar to the same assumption as in [21]–[23] and regardless of distributed or centralized processing, we assume that all precoders and combiners are constructed with the knowledge of *unified* CSI mismatch. We further model the CSI mismatch as

$$\widehat{\mathbf{H}}_{k,j} = \mathbf{H}_{k,j} + \mathbf{E}_{k,j} , \qquad (4)$$

where the channel measurement error $\mathbf{E}_{k,j}$ is thought to be independent of actual channel matrix $\mathbf{H}_{k,j}$. Analogous to [22], we consider $\mathbf{E}_{k,j}$ as a Gaussian matrix consisting of i.i.d. elements with mean zero and variance τ, i.e.,

$$\text{vec}\left(\mathbf{E}_{k,j}\right) \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \tau \mathbf{I}\right) \text{ with } \tau \triangleq \beta\gamma^{-\alpha}, \quad \beta > 0, \quad \alpha \geq 0 , \qquad (5)$$

where α is the SNR exponent and β is the SNR scaling factor, and both are considered to be independent of the SNR. In this case, the error variance can depend on SNR ($\alpha \neq 0$) or be independent of that ($\alpha = 0$). Notice the variance model in (5) is versatile since it is potentially able to accommodate a variety of distinct scenarios, e.g., reciprocal channels ($\alpha = 1$) and CSI

feedback ($\alpha = 0$) [22], [24]. Also, $0 < \alpha < 1$ may reflect the scenario in which feedback power is much smaller than feedforward power. This might be the case where the mobile and BS powers are not of the same order. Therefore, the BS can reciprocally learn the forward link, but instead of full DoF, i.e., Kd, only an α fraction of that, i.e, αKd, are achievable [22].

τ can be further interpreted as a parameter that captures the quality of the channel estimation which is possible to be known a priori, depending on the channel dynamics and channel estimation schemes (see e.g., [25] and references therein).

To facilitate the performance analysis of IA under CSI mismatch model in (4), it is more appropriate to have the statistical properties of $\mathbf{H}_{k,j}$ conditioned on $\widehat{\mathbf{H}}_{k,j}$ by using following lemma [26]:

Lemma 1: Conditioned on $\widehat{\mathbf{H}}_{k,j}$, $\mathbf{H}_{k,j}$ has a Gaussian distribution with mean $\widehat{\mathbf{H}}_{k,j}/(1+\tau)$ and statistically independent elements of variance $\tau/(1+\tau)$, i.e.,

$$\mathbf{H}_{k,j} = \frac{1}{1+\tau}\widehat{\mathbf{H}}_{k,j} + \breve{\mathbf{H}}_{k,j} , \qquad (6)$$

where the *auxiliary* random matrix $\text{vec}\left(\breve{\mathbf{H}}_{k,j}\right) \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \frac{\tau}{1+\tau}\mathbf{I}\right)$ is statistically independent of $\widehat{\mathbf{H}}_{k,j}$.

C. Signal Postprocessing at Receive Nodes

In this subsection, we briefly address the data recovery at receive nodes. Without loss of generality, we assume that each receive node uses a linear ZF equalizer. It is also worth mentioning that the results of this subsection can be readily generalized to the case of channel inversion, i.e., preprocessing

the signals at transmit nodes instead of postprocessing the signals at receive nodes.

We first assume that perfect CSI is available. In this case, if we define

$$
\mathbf{V} = \begin{pmatrix} \mathbf{V}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{V}_K \end{pmatrix} , \tag{7}
$$

$$
\mathbf{U} = \begin{pmatrix} \mathbf{U}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{U}_K \end{pmatrix} , \tag{8}
$$

$$
\mathbf{H} = \begin{pmatrix} \mathbf{H}_{1,1} & \mathbf{H}_{1,2} & \cdots & \mathbf{H}_{1,K} \\ \mathbf{H}_{2,1} & \mathbf{H}_{2,2} & \cdots & \mathbf{H}_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{K,1} & \mathbf{H}_{K,2} & \cdots & \mathbf{H}_{K,K} \end{pmatrix} , \tag{9}
$$

with respect to (2) and at high enough SNRs, we have

$$
\mathbf{U}^{\mathsf{H}}\mathbf{H}\mathbf{V} = \begin{pmatrix} \overline{\mathbf{H}}_{1,1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \overline{\mathbf{H}}_{K,K} \end{pmatrix} , \tag{10}
$$

where $\overline{\mathbf{H}}_{k,k} = \mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{V}_k$. More specifically, by premultiplying the received signal at receiver k by $\mathbf{U}_k^{\mathsf{H}}$ we have

$$
\mathbf{U}_k^{\mathsf{H}}\mathbf{y}_k = \mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{x}_k + \underbrace{\mathbf{U}_k^{\mathsf{H}}\sum_{\substack{j=1 \\ j \neq k}}^{K}\mathbf{H}_{k,j}\mathbf{x}_j + \mathbf{U}_k^{\mathsf{H}}\mathbf{z}_k}_{\overline{\mathbf{z}}_k}
$$

$$
\Rightarrow \overline{\mathbf{y}}_k = \overline{\mathbf{H}}_{k,k}\mathbf{c}_k + \overline{\mathbf{z}}_k , \tag{11}
$$

where $\overline{\mathbf{y}}_k = \mathbf{U}_k^{\mathsf{H}}\mathbf{y}_k$, and following (2), at high enough SNRs we have $\overline{\mathbf{z}}_k = \mathbf{U}_k^{\mathsf{H}}\mathbf{z}_k$. Therefore the transmitted symbol vector $\mathbf{c}_k$ can be easily recovered through premultiplying $\overline{\mathbf{y}}_k$ by (pseudo-) inverse of $\overline{\mathbf{H}}_{k,k}$, i.e., $\mathrm{pinv}\left(\overline{\mathbf{H}}_{k,k}\right)$.

Now we assume that all precoders and combiners are constructed based on imperfect CSI in (4). Consequently, (2) can be written as

$$
\widehat{\mathbf{U}}_k^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}\widehat{\mathbf{V}}_j = \mathbf{0} \quad , \quad \forall j \neq k , \tag{12}
$$

where all $\widehat{\mathbf{U}}_k$ and $\widehat{\mathbf{V}}_j$ are calculated based on the fact that only imperfect channel estimations $\widehat{\mathbf{H}}_{k,j}$ are available. In this case the received signal at node k in (1) can be written as

$$
\widehat{\mathbf{y}}_k = \mathbf{H}_{k,k}\widehat{\mathbf{V}}_k\mathbf{c}_k + \sum_{\substack{j=1 \\ j \neq k}}^{K}\mathbf{H}_{k,j}\widehat{\mathbf{V}}_j\mathbf{c}_j + \mathbf{z}_k . \tag{13}
$$

If the perfect direct link, i.e., $\mathbf{H}_{k,k}$, is available at receive node k, recovering $\mathbf{c}_k$ at the corresponding node is rather straightforward; however, if instead of perfect CSI, we assume that the receive node k has only access to imperfect direct link, i.e., $\widehat{\mathbf{H}}_{k,k}$, data recovery becomes a bit tricky. Therefore in the remainder of this subsection, we address this issue by

considering the fact that only imperfect direct link $\widehat{\mathbf{H}}_{k,k}$ is available at receive node k. Now at the kth receive node, the received signal is premultiplied by $\widehat{\mathbf{U}}_k^{\mathsf{H}}$, and therefore we have

$$
\begin{aligned}
\widehat{\mathbf{U}}_k^{\mathsf{H}}\widehat{\mathbf{y}}_k &= \widehat{\mathbf{U}}_k^{\mathsf{H}}\mathbf{H}_{k,k}\widehat{\mathbf{V}}_k\mathbf{c}_k + \sum_{\substack{j=1 \\ j \neq k}}^{K}\widehat{\mathbf{U}}_k^{\mathsf{H}}\mathbf{H}_{k,j}\widehat{\mathbf{V}}_j\mathbf{c}_j + \widehat{\mathbf{U}}_k^{\mathsf{H}}\mathbf{z}_k \\
&\overset{①}{=} \widehat{\mathbf{U}}_k^{\mathsf{H}}\left(\frac{1}{1+\tau}\widehat{\mathbf{H}}_{k,k}+\breve{\mathbf{H}}_{k,k}\right)\widehat{\mathbf{V}}_k\mathbf{c}_k \\
&\quad + \sum_{\substack{j=1 \\ j \neq k}}^{K}\widehat{\mathbf{U}}_k^{\mathsf{H}}\left(\frac{1}{1+\tau}\widehat{\mathbf{H}}_{k,j}+\breve{\mathbf{H}}_{k,j}\right)\widehat{\mathbf{V}}_j\mathbf{c}_j+\widehat{\mathbf{U}}_k^{\mathsf{H}}\mathbf{z}_k \\
&\overset{②}{=} \underbrace{\frac{1}{1+\tau}\widehat{\mathbf{U}}_k^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}\widehat{\mathbf{V}}_k\mathbf{c}_k}_{\text{desired term}} + \underbrace{\sum_{j=1}^{K}\widehat{\mathbf{U}}_k^{\mathsf{H}}\breve{\mathbf{H}}_{k,j}\widehat{\mathbf{V}}_j\mathbf{c}_j + \widehat{\mathbf{U}}_k^{\mathsf{H}}\mathbf{z}_k}_{\text{interference plus noise term}} ,
\end{aligned} \tag{14}
$$

where ① follows from (6), and ② is due to (12). Thus to recover $\mathbf{c}_k$, $\widehat{\mathbf{U}}_k^{\mathsf{H}}\widehat{\mathbf{y}}_k$ should be premultiplied by $(1+\tau) \times \mathrm{pinv}\left(\widehat{\mathbf{U}}_k^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}\widehat{\mathbf{V}}_k\right)$. In other words, when the receive node is in possession of imperfect direct link $\widehat{\mathbf{H}}_{k,k}$, to have an unbiased detection, the received signal should be scaled up by $(1+\tau)$.

III. LS BASED IA

In this section, we propose an optimized beamformer design for MIMO IC based on LS criterion and the knowledge of imperfect CSI.

Note that due to the coupled nature of the beamformer design for multiuser MIMO IC, there are no closed form solutions for IA, except for a few particular cases (see e.g., [1], [27]). Consequently, finding precoders and combiners requests an iterative procedure in general. Therefore, first the precoders get fixed and the combiners are sought, and then the combiners get fixed and the precoders are sought, through an iterative manner.

For the proposed LS based IA and without loss of generality, we consider the adaptive design under imperfect CSI as the major optimization problem, and as it will be shown later, the standard design under perfect CSI is a special case of the adaptive design by setting error variance equal to zero. In order to calculate the beamformers adaptively, we assume that the variance of channel estimation error is known in advance. Given randomly initialized precoders and with respect to (1), the optimization problem to find the appropriate combiners based on LS criterion can be considered as

$$
\min_{\mathbf{U}_k} \mathbb{E}\left\{\left\|\mathbf{U}_k^{\mathsf{H}}\sum_{j=1}^{K}\mathbf{H}_{k,j}\mathbf{V}_j\mathbf{c}_j - \mathbf{c}_k\right\|_2^2\right\} \quad \forall k = 1, \ldots, K . \tag{15}
$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2391195, IEEE Transactions on Vehicular Technology

5

Now the objective function of the optimization problem in (15) can be defined as

$$
F_{\mathrm{LS}}^{\mathrm{U}} = \mathbb{E} \left\{ \mathrm{Tr} \left[\left(\mathbf{U}_k^{\mathsf{H}} \sum_{j=1}^{K} \mathbf{H}_{k,j} \mathbf{V}_j \mathbf{c}_j - \mathbf{c}_k \right) \right. \right.
$$
$$
\left. \left. \times \left(\mathbf{U}_k^{\mathsf{H}} \sum_{j=1}^{K} \mathbf{H}_{k,j} \mathbf{V}_j \mathbf{c}_j - \mathbf{c}_k \right)^{\mathsf{H}} \right] \right\} . \tag{16}
$$

To further continue, we consider the following assumptions:

Remark 1: We assume that the transmitted data vector $\mathbf{c}_k$ consists of i.i.d. symbols, i.e., $\mathbb{E} \left\{ \mathbf{c}_k \mathbf{c}_j^{\mathsf{H}} \right\} = \mathbf{0}$, $j \neq k$, and also the noise vector $\mathbf{z}_k$ is independent of the data vector $\mathbf{c}_k$ as well as the channel matrices $\mathbf{H}_{k,j}$.

With respect to Remark 1, $F_{\mathrm{LS}}^{\mathrm{U}}$ in (16) can be rewritten as

$$
F_{\mathrm{LS}}^{\mathrm{U}} = \mathrm{Tr} \left[P\, \mathbf{U}_k^{\mathsf{H}} \sum_{j=1}^{K} \mathbf{H}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \mathbf{H}_{k,j}^{\mathsf{H}} \mathbf{U}_k - P\, \mathbf{V}_k^{\mathsf{H}} \mathbf{H}_{k,k}^{\mathsf{H}} \mathbf{U}_k \right.
$$
$$
\left. - P\, \mathbf{U}_k^{\mathsf{H}} \mathbf{H}_{k,k} \mathbf{V}_k \right] + Pd . \tag{17}
$$

To obtain the sought combiners based on minimizing the objective function in (17), we differentiate $F_{\mathrm{LS}}^{\mathrm{U}}$ with respect to $\mathbf{U}_k$ by first considering the following assumptions [24], [28]:

1) $\mathbf{U}_k$ and $\mathbf{U}_k^{\mathsf{H}}$ are treated as independent variables.
2) $\dfrac{\partial\,\mathrm{Tr}\,[\mathbf{A}\mathbf{U}_k]}{\partial \mathbf{U}_k} = \dfrac{\partial\,\mathrm{Tr}\,[\mathbf{U}_k \mathbf{A}]}{\partial \mathbf{U}_k} = \mathbf{A}.$

Therefore, with respect to the two preceding assumptions, we have

$$
\frac{\partial F_{\mathrm{LS}}^{\mathrm{U}}}{\partial \mathbf{U}_k} = P\, \mathbf{U}_k^{\mathsf{H}} \sum_{j=1}^{K} \mathbf{H}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \mathbf{H}_{k,j}^{\mathsf{H}} - P\, \mathbf{V}_k^{\mathsf{H}} \mathbf{H}_{k,k}^{\mathsf{H}} . \tag{18}
$$

Thus, by considering (6), equation (18) can be rewritten as

$$
\frac{\partial F_{\mathrm{LS}}^{\mathrm{U}}}{\partial \mathbf{U}_k} = P\, \mathbf{U}_k^{\mathsf{H}} \sum_{j=1}^{K} \left(\frac{\widehat{\mathbf{H}}_{k,j}}{1+\tau} + \breve{\mathbf{H}}_{k,j} \right) \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \left(\frac{\widehat{\mathbf{H}}_{k,j}}{1+\tau} + \breve{\mathbf{H}}_{k,j} \right)^{\mathsf{H}}
$$
$$
- P\, \mathbf{V}_k^{\mathsf{H}} \left(\frac{\widehat{\mathbf{H}}_{k,k}}{1+\tau} + \breve{\mathbf{H}}_{k,k} \right)^{\mathsf{H}} . \tag{19}
$$

Note that $\partial F_{\mathrm{LS}}^{\mathrm{U}} / \partial \mathbf{U}_k$ is now dependent on both $\widehat{\mathbf{H}}_{k,j}$ and $\breve{\mathbf{H}}_{k,j}$. To make $\partial F_{\mathrm{LS}}^{\mathrm{U}} / \partial \mathbf{U}_k$ dependent only on $\widehat{\mathbf{H}}_{k,j}$, we can take the expectation over the auxiliary random matrix $\breve{\mathbf{H}}_{k,j}$ by noticing the following two lemmas:

Lemma 2:
$$
\mathbb{E}_{\breve{\mathbf{H}}|\widehat{\mathbf{H}}} \left\{ \widehat{\mathbf{H}}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \breve{\mathbf{H}}_{k,j}^{\mathsf{H}} \right\} = \mathbb{E}_{\breve{\mathbf{H}}|\widehat{\mathbf{H}}} \left\{ \breve{\mathbf{H}}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \widehat{\mathbf{H}}_{k,j}^{\mathsf{H}} \right\} = \mathbf{0} . \tag{20}
$$

Proof: All precoders and combiners are constructed upon channel estimates $\widehat{\mathbf{H}}_{k,j}$ which based on Lemma 1 are independent of $\breve{\mathbf{H}}_{k,j}$. ∎

Lemma 3: If $\mathbf{A} \in \mathbb{C}^{M \times N}$ represents a Gaussian matrix with i.i.d. elements of mean zero and variance a, and $\mathbf{B} \in \mathbb{C}^{N \times d}$ refers to a truncated unitary matrix independent of $\mathbf{A}$, then $\mathbb{E}_{\mathbf{A}} \left\{ \mathbf{A} \mathbf{B} \mathbf{B}^{\mathsf{H}} \mathbf{A}^{\mathsf{H}} \right\} = ad\, \mathbf{I}$.

Proof: Since $\mathbf{A}$ is a Gaussian matrix, it is bi-unitarily invariant, and consequently the joint distribution of its entries equals that of $\mathbf{A}\mathbf{B}$ for any truncated unitary matrix $\mathbf{B}$ independent of $\mathbf{A}$ [29]. Therefore $\mathbf{A}\mathbf{B}$ is equivalent to a Gaussian matrix with i.i.d. elements of mean zero and variance a. Since $\mathbf{A}\mathbf{B}$ has d independent columns, the claim follows. ∎

Following Lemma 2 and Lemma 3 and by taking the expectation of $\partial F_{\mathrm{LS}}^{\mathrm{U}} / \partial \mathbf{U}_k$ over the auxiliary random matrix $\breve{\mathbf{H}}_{k,j}$, we have

$$
\mathbb{E}_{\breve{\mathbf{H}}_{k,j}} \left\{ \frac{\partial F_{\mathrm{LS}}^{\mathrm{U}}}{\partial \mathbf{U}_k} \right\} = P\, \mathbf{U}_k^{\mathsf{H}} \left(\frac{1}{(1+\tau)^2} \sum_{j=1}^{K} \widehat{\mathbf{H}}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \widehat{\mathbf{H}}_{k,j}^{\mathsf{H}} \right.
$$
$$
\left. + \frac{Kd\tau}{1+\tau} \mathbf{I} \right) - \frac{P}{1+\tau} \mathbf{V}_k^{\mathsf{H}} \widehat{\mathbf{H}}_{k,k}^{\mathsf{H}} . \tag{21}
$$

Now the sought combiner $\mathbf{U}_k$ can be obtained by setting $\mathbb{E}_{\breve{\mathbf{H}}_{k,j}} \left\{ \dfrac{\partial F_{\mathrm{LS}}^{\mathrm{U}}}{\partial \mathbf{U}_k} \right\}$ equal to zero, which yields

$$
P\, \mathbf{U}_k^{\mathsf{H}} \left(\frac{1}{(1+\tau)^2} \sum_{j=1}^{K} \widehat{\mathbf{H}}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \widehat{\mathbf{H}}_{k,j}^{\mathsf{H}} + \frac{Kd\tau}{1+\tau} \mathbf{I} \right)
$$
$$
- \frac{P}{1+\tau} \mathbf{V}_k^{\mathsf{H}} \widehat{\mathbf{H}}_{k,k}^{\mathsf{H}} = \mathbf{0}
$$
$$
\Rightarrow \mathbf{U}_k = (1+\tau) \left(\sum_{j=1}^{K} \widehat{\mathbf{H}}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \widehat{\mathbf{H}}_{k,j}^{\mathsf{H}} + \varepsilon_{\mathrm{LS}}^{\mathrm{U}} \mathbf{I} \right)^{-1} \widehat{\mathbf{H}}_{k,k} \mathbf{V}_k , \tag{22}
$$

such that $\varepsilon_{\mathrm{LS}}^{\mathrm{U}} = \tau\,(1+\tau)\,Kd$.

Now we turn our focus to obtain the precoders by first considering the following lemma:

Lemma 4: (Reciprocity of Alignment) [2]—If Kd DoF are achievable on the original interference network then the same Kd DoF are also achievable on the reciprocal interference network and vice versa. IA based on this reciprocity can be readily attained by choosing the precoders and combiners on the reciprocal channels as the combiners and precoders of the original channel, respectively. In this case, we have $\overleftarrow{\mathbf{H}}_{k,j} = \overrightarrow{\mathbf{H}}_{j,k}^{\mathsf{H}}$ where $\overleftarrow{\mathbf{H}}_{k,j}$ is the channel between transmitter j and receiver k in reciprocal network and $\overrightarrow{\mathbf{H}}_{j,k}$ is the channel between transmitter k and receiver j in original network.

With respect to Lemma 4 and given randomly initialized combiners, the optimization problem based on LS criterion can be considered as

$$
\min_{\mathbf{V}_k} \mathbb{E} \left\{ \left\| \mathbf{V}_k^{\mathsf{H}} \sum_{j=1}^{K} \mathbf{H}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{c}_j - \mathbf{c}_k \right\|_2^2 \right\} \quad \forall k = 1, \ldots, K . \tag{23}
$$

Now the objective function of the optimization problem in (23) can be defined as

$$F_{\mathrm{LS}}^{\mathrm{V}} = \mathbb{E}\left\{ \mathrm{Tr}\left[\left(\mathbf{V}_k^{\mathsf{H}} \sum_{j=1}^{K} \mathbf{H}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{c}_j - \mathbf{c}_k \right) \right.\right.$$
$$\left.\left. \times \left(\mathbf{V}_k^{\mathsf{H}} \sum_{j=1}^{K} \mathbf{H}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{c}_j - \mathbf{c}_k \right)^{\mathsf{H}} \right] \right\}. \quad (24)$$

With respect to Remark 1, $F_{\mathrm{LS}}^{\mathrm{V}}$ in (24) can be rewritten as

$$F_{\mathrm{LS}}^{\mathrm{V}} = \mathrm{Tr}\left[P\,\mathbf{V}_k^{\mathsf{H}} \sum_{j=1}^{K} \mathbf{H}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \mathbf{H}_{j,k} \mathbf{V}_k - P\,\mathbf{U}_k^{\mathsf{H}} \mathbf{H}_{k,k} \mathbf{V}_k \right.$$
$$\left. - P\,\mathbf{V}_k^{\mathsf{H}} \mathbf{H}_{k,k}^{\mathsf{H}} \mathbf{U}_k \right] + Pd. \quad (25)$$

Similarly, it is straightforward to show that

$$\mathbb{E}_{\breve{\mathbf{H}}_{j,k}}\left\{ \frac{\partial F_{\mathrm{LS}}^{\mathrm{V}}}{\partial \mathbf{V}_k} \right\} = P\,\mathbf{V}_k^{\mathsf{H}}\left(\frac{1}{(1+\tau)^2} \sum_{j=1}^{K} \widehat{\mathbf{H}}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \widehat{\mathbf{H}}_{j,k} \right.$$
$$\left. + \frac{Kd\tau}{1+\tau}\mathbf{I} \right) - \frac{P}{1+\tau}\mathbf{U}_k^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}. \quad (26)$$

The sought combiner $\mathbf{V}_k$ can then be obtained by setting $\mathbb{E}_{\breve{\mathbf{H}}_{j,k}}\left\{ \frac{\partial F_{\mathrm{LS}}^{\mathrm{V}}}{\partial \mathbf{V}_k} \right\}$ equal to zero, which yields

$$P\,\mathbf{V}_k^{\mathsf{H}}\left(\frac{1}{(1+\tau)^2} \sum_{j=1}^{K} \widehat{\mathbf{H}}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \widehat{\mathbf{H}}_{j,k} + \frac{Kd\tau}{1+\tau}\mathbf{I} \right)$$
$$- \frac{P}{1+\tau}\mathbf{U}_k^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k} = \mathbf{0}$$
$$\Rightarrow \mathbf{V}_k = (1+\tau)\left(\sum_{j=1}^{K} \widehat{\mathbf{H}}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \widehat{\mathbf{H}}_{j,k} + \varepsilon_{\mathrm{LS}}^{\mathrm{V}} \mathbf{I} \right)^{-1} \widehat{\mathbf{H}}_{k,k}^{\mathsf{H}} \mathbf{U}_k, \quad (27)$$

where $\varepsilon_{\mathrm{LS}}^{\mathrm{V}} = \tau(1+\tau)Kd$. As mentioned earlier and analogous to the standard IA techniques, due to the coupled nature of the beamformer design, finding precoders and combiners requests an iterative algorithm in general. Therefore, with respect to the fact that *unitary precoders and combiners* are more desirable, the proposed algorithm, which iteratively optimizes the precoders and combiners, can be concisely presented as follows:

LS Based IA

1: Set $\varepsilon_{\mathrm{LS}} := \tau(1+\tau)Kd$
2: Initialize random unitary matrices $\mathbf{V}_k, \forall\, k$

3: $\mathbf{U}_k = (1+\tau)\left(\sum_{j=1}^{K} \widehat{\mathbf{H}}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \widehat{\mathbf{H}}_{k,j}^{\mathsf{H}} + \varepsilon_{\mathrm{LS}}\mathbf{I} \right)^{-1} \widehat{\mathbf{H}}_{k,k}\mathbf{V}_k$
4: $\mathbf{U}_k \Longleftarrow \mathrm{orth}\left(\mathbf{U}_k \right)$
5: $\mathbf{V}_k = (1+\tau)\left(\sum_{j=1}^{K} \widehat{\mathbf{H}}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \widehat{\mathbf{H}}_{j,k} + \varepsilon_{\mathrm{LS}}\mathbf{I} \right)^{-1} \widehat{\mathbf{H}}_{k,k}^{\mathsf{H}} \mathbf{U}_k$
6: $\mathbf{V}_k \Longleftarrow \mathrm{orth}\left(\mathbf{V}_k \right)$
7: Go to Step 3 and repeat

Remark 2: Although the aforementioned algorithm has been designed based on the channel estimations $\widehat{\mathbf{H}}_{k,j}$, it can be readily used for the perfect CSI. In this case, we need to set $\tau = 0$, and replace $\widehat{\mathbf{H}}_{k,j}$ with $\mathbf{H}_{k,j}$.

Remark 3: As revealed in LS based IA, to design beamformers for user k, the direct link, i.e., $\mathbf{H}_{k,k}$, is taken into account. This is in contrast to interference leakage minimization algorithms where $\mathbf{H}_{k,k}$ is not considered to design beamformers for user k, as denoted in Min-WLI [2] and Alt-Min [9] algorithms. Therefore, it is expected that the proposed LS based IA should outperform interference leakage minimization algorithms under both perfect and imperfect CSI.

IV. MMSE BASED IA

In Section III, we proposed an LS based beamformer design for multiuser MIMO IC. Nonetheless, as it is well-known, MMSE based designs outperform LS based schemes since they further consider the effect of noise to design beamformers. Therefore, in this section, we propose an MMSE based IA, which is similar to Max-SINR algorithm, considers the effect of noise to calculate beamformers. Since we will largely compare the performance of the proposed MMSE based IA with Max-SINR, it is appropriate to recall this algorithm. However, as stated in [12], [19], [22], Max-SINR without orthogonalization can not achieve full DoF, since at least one of its precoders and/or combiners may become rank-deficient, and this can further result in reduced multiplexing gain. Therefore, when using Max-SINR, if the goal is to preserve full DoF, the inclusion of orthogonalization steps is mandatory. Consequently, in the sequel, we use Max-SINR with orthogonalization which can be represented as follows [22], [23]:

Max-SINR (with orthogonalization)

1: Set $\varepsilon' := \gamma^{-1}$
2: Initialize random unit-norm vectors $\mathbf{v}_{k\ell}, \forall\, k, \ell$
3: $\mathbf{T}_k^{\ell} = \sum_{j=1}^{K} \mathbf{H}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \mathbf{H}_{k,j}^{\mathsf{H}} - \mathbf{H}_{k,k}\mathbf{v}_{k\ell}\mathbf{v}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}^{\mathsf{H}} + \varepsilon'\,\mathbf{I}$
4: $\mathbf{u}_{k\ell} = \dfrac{\left(\mathbf{T}_k^{\ell}\right)^{-1}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}}{\left\| \left(\mathbf{T}_k^{\ell}\right)^{-1}\mathbf{H}_{k,k}\mathbf{v}_{k\ell} \right\|_2} \qquad \ell = 1, \dots, d$
5: $\mathbf{U}_k \Longleftarrow \mathrm{orth}\left(\mathbf{U}_k \right)$
6: $\mathbf{R}_k^{\ell} = \sum_{j=1}^{K} \mathbf{H}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \mathbf{H}_{j,k} - \mathbf{H}_{k,k}^{\mathsf{H}}\mathbf{u}_{k\ell}\mathbf{u}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k} + \varepsilon'\,\mathbf{I}$

7: $\quad \mathbf{v}_{k\ell} = \dfrac{\left(\mathbf{R}_k^\ell\right)^{-1}\mathbf{H}_{k,k}^{\mathsf{H}}\mathbf{u}_{k\ell}}{\left\|\left(\mathbf{R}_k^\ell\right)^{-1}\mathbf{H}_{k,k}^{\mathsf{H}}\mathbf{u}_{k\ell}\right\|_2} \qquad \ell = 1,\dots,d$

8: $\quad \mathbf{V}_k \Longleftarrow \mathrm{orth}\left(\mathbf{V}_k\right)$

9: Go to Step 3 and repeat

Hereafter and for the sake of simplicity, we call Max-SINR in lieu of Max-SINR with orthogonalization.

To derive the desired beamformers based on the MMSE criterion and similar to the same approach as in Section III, we consider an adaptive design based on the knowledge of imperfect CSI as the major optimization problem, and as it will be shown later, the standard design under perfect CSI is a special case of the adaptive design under imperfect CSI by setting error variance equal to zero. Given randomly initialized precoders and with respect to (1), the optimization problem based on MMSE criterion can be considered as

$$\min_{\mathbf{U}_k}\ \mathbb{E}\left\{\left\|\mathbf{U}_k^{\mathsf{H}}\mathbf{y}_k - \mathbf{c}_k\right\|_2^2\right\} \quad \forall k=1,\dots,K\ . \qquad (28)$$

In this case, the objective function in (28) can de shown as

$$
F_{\mathrm{MMSE}}^{\mathrm{U}} = \mathbb{E}\left\{ \mathrm{Tr}\left[\left(\mathbf{U}_k^{\mathsf{H}}\sum_{j=1}^{K}\mathbf{H}_{k,j}\mathbf{V}_j\mathbf{c}_j + \mathbf{U}_k^{\mathsf{H}}\mathbf{z}_k - \mathbf{c}_k\right)\right.\right.
$$
$$
\left.\left. \times \left(\mathbf{U}_k^{\mathsf{H}}\sum_{j=1}^{K}\mathbf{H}_{k,j}\mathbf{V}_j\mathbf{c}_j + \mathbf{U}_k^{\mathsf{H}}\mathbf{z}_k - \mathbf{c}_k\right)^{\mathsf{H}}\right]\right\}\ . \qquad (29)
$$

With respect to Remark 1, (29) is reduced to

$$
F_{\mathrm{MMSE}}^{\mathrm{U}} = \mathrm{Tr}\left[P\,\mathbf{U}_k^{\mathsf{H}}\sum_{j=1}^{K}\mathbf{H}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}}\mathbf{U}_k + \sigma^2\mathbf{U}_k^{\mathsf{H}}\mathbf{U}_k \right.
$$
$$
\left. -\,P\,\mathbf{V}_k^{\mathsf{H}}\mathbf{H}_{k,k}^{\mathsf{H}}\mathbf{U}_k - P\,\mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{V}_k \right] + Pd\ . \qquad (30)
$$

Since minimizing the optimization problem in (30) requires differentiation, we have

$$
\frac{\partial F_{\mathrm{MMSE}}^{\mathrm{U}}}{\partial \mathbf{U}_k} = P\,\mathbf{U}_k^{\mathsf{H}}\sum_{j=1}^{K}\mathbf{H}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}} + \sigma^2\mathbf{U}_k^{\mathsf{H}} - P\,\mathbf{V}_k^{\mathsf{H}}\mathbf{H}_{k,k}^{\mathsf{H}}\ . \qquad (31)
$$

By considering (6), equation (31) can be further rewritten as

$$
\frac{\partial F_{\mathrm{MMSE}}^{\mathrm{U}}}{\partial \mathbf{U}_k} = P\mathbf{U}_k^{\mathsf{H}}\sum_{j=1}^{K}\left(\frac{\widehat{\mathbf{H}}_{k,j}}{1+\tau}+\breve{\mathbf{H}}_{k,j}\right)\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\left(\frac{\widehat{\mathbf{H}}_{k,j}}{1+\tau}+\breve{\mathbf{H}}_{k,j}\right)^{\mathsf{H}}
$$
$$
+\,\sigma^2\mathbf{U}_k^{\mathsf{H}} - P\,\mathbf{V}_k^{\mathsf{H}}\left(\frac{\widehat{\mathbf{H}}_{k,k}}{1+\tau}+\breve{\mathbf{H}}_{k,k}\right)^{\mathsf{H}}\ . \qquad (32)
$$

Note that $\partial F_{\mathrm{MMSE}}^{\mathrm{U}}/\partial\mathbf{U}_k$ is now dependent on both $\widehat{\mathbf{H}}_{k,j}$ and $\breve{\mathbf{H}}_{k,j}$. To make $\partial F_{\mathrm{MMSE}}^{\mathrm{U}}/\partial\mathbf{U}_k$ dependent only on $\widehat{\mathbf{H}}_{k,j}$, and with respect to Lemma 2 and Lemma 3, we take the expectation

of $\partial F_{\mathrm{MMSE}}^{\mathrm{U}}/\partial\mathbf{U}_k$ over the auxiliary random matrix $\breve{\mathbf{H}}_{k,j}$ which yields

$$
\mathbb{E}_{\breve{\mathbf{H}}_{k,j}}\left\{\frac{\partial F_{\mathrm{MMSE}}^{\mathrm{U}}}{\partial\mathbf{U}_k}\right\} = P\,\mathbf{U}_k^{\mathsf{H}}\left(\frac{1}{(1+\tau)^2}\sum_{j=1}^{K}\widehat{\mathbf{H}}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}}\right.
$$
$$
\left.+\,\frac{Kd\tau}{1+\tau}\mathbf{I}\right) + \sigma^2\mathbf{U}_k^{\mathsf{H}} - \frac{P}{1+\tau}\,\mathbf{V}_k^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}}\ . \qquad (33)
$$

Now the sought combiner $\mathbf{U}_k$ can be obtained by setting $\mathbb{E}_{\breve{\mathbf{H}}_{k,j}}\left\{\dfrac{\partial F_{\mathrm{MMSE}}^{\mathrm{U}}}{\partial\mathbf{U}_k}\right\}$ equal to zero, which yields

$$
P\,\mathbf{U}_k^{\mathsf{H}}\left(\frac{1}{(1+\tau)^2}\sum_{j=1}^{K}\widehat{\mathbf{H}}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}} + \frac{Kd\tau}{1+\tau}\mathbf{I}\right) + \sigma^2\mathbf{U}_k^{\mathsf{H}}
$$
$$
-\,\frac{P}{1+\tau}\,\mathbf{V}_k^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}} = \mathbf{0}
$$
$$
\Rightarrow \mathbf{U}_k = (1+\tau)\left(\sum_{j=1}^{K}\widehat{\mathbf{H}}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}} + \varepsilon_{\mathrm{MMSE}}^{\mathrm{U}}\mathbf{I}\right)^{-1}\widehat{\mathbf{H}}_{k,k}\mathbf{V}_k\ , \qquad (34)
$$

where
$$
\varepsilon_{\mathrm{MMSE}}^{\mathrm{U}} = \gamma^{-1}\left(1+\tau\right)^2 + \tau\left(1+\tau\right)Kd\ . \qquad (35)
$$

Now we turn our focus to obtain the precoders. With respect to Lemma 4 and given randomly initialized combiners, the optimization criterion based on MMSE design can be considered as

$$
\min_{\mathbf{V}_k}\ \mathbb{E}\left\{\left\|\mathbf{V}_k^{\mathsf{H}}\mathbf{y}_k' - \mathbf{c}_k\right\|_2^2\right\} \quad \forall k=1,\dots,K\ , \qquad (36)
$$

where $\mathbf{y}_k'$ is defined as

$$
\mathbf{y}_k' = \sum_{j=1}^{K}\mathbf{H}_{j,k}^{\mathsf{H}}\mathbf{U}_j\mathbf{c}_j + \mathbf{z}_k'\ , \qquad (37)
$$

such that $\mathbf{z}_k' \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0},\sigma^2\mathbf{I}\right)$.

In this case, the objective function in (36) can be shown as

$$
F_{\mathrm{MMSE}}^{\mathrm{V}} = \mathbb{E}\left\{ \mathrm{Tr}\left[\left(\mathbf{V}_k^{\mathsf{H}}\sum_{j=1}^{K}\mathbf{H}_{j,k}^{\mathsf{H}}\mathbf{U}_j\mathbf{c}_j + \mathbf{V}_k^{\mathsf{H}}\mathbf{z}_k' - \mathbf{c}_k\right)\right.\right.
$$
$$
\left.\left. \times \left(\mathbf{V}_k^{\mathsf{H}}\sum_{j=1}^{K}\mathbf{H}_{j,k}^{\mathsf{H}}\mathbf{U}_j\mathbf{c}_j + \mathbf{V}_k^{\mathsf{H}}\mathbf{z}_k' - \mathbf{c}_k\right)^{\mathsf{H}}\right]\right\}\ . \qquad (38)
$$

With respect to Remark 1, (38) is reduced to

$$
F_{\mathrm{MMSE}}^{\mathrm{V}} = \mathrm{Tr}\left[P\,\mathbf{V}_k^{\mathsf{H}}\sum_{j=1}^{K}\mathbf{H}_{j,k}^{\mathsf{H}}\mathbf{U}_j\mathbf{U}_j^{\mathsf{H}}\mathbf{H}_{j,k}\mathbf{V}_k + \sigma^2\mathbf{V}_k^{\mathsf{H}}\mathbf{V}_k \right.
$$
$$
\left. -\,P\,\mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{V}_k - P\,\mathbf{V}_k^{\mathsf{H}}\mathbf{H}_{k,k}^{\mathsf{H}}\mathbf{U}_k \right] + Pd\ . \qquad (39)
$$

Similarly, it is straightforward to show that

$$
\mathbb{E}_{\breve{\mathbf{H}}_{j,k}} \left\{ \frac{\partial F_{\text{MMSE}}^{\text{V}}}{\partial \mathbf{V}_k} \right\} = P \, \mathbf{V}_k^{\mathsf{H}} \left(\frac{1}{(1+\tau)^2} \sum_{j=1}^{K} \widehat{\mathbf{H}}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \widehat{\mathbf{H}}_{j,k} \right.
$$
$$
\left. + \frac{Kd\tau}{1+\tau} \mathbf{I} \right) + \sigma^2 \mathbf{V}_k^{\mathsf{H}} - \frac{P}{1+\tau} \mathbf{U}_k^{\mathsf{H}} \widehat{\mathbf{H}}_{k,k} \ .
$$

$$(40)$$

Now the sought combiner $\mathbf{V}_k$ can be obtained by setting $\mathbb{E}_{\breve{\mathbf{H}}_{j,k}} \left\{ \frac{\partial F_{\text{MMSE}}^{\text{V}}}{\partial \mathbf{V}_k} \right\}$ equal to zero, which yields

$$
P \, \mathbf{V}_k^{\mathsf{H}} \left(\frac{1}{(1+\tau)^2} \sum_{j=1}^{K} \widehat{\mathbf{H}}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \widehat{\mathbf{H}}_{j,k}^{\mathsf{H}} + \frac{Kd\tau}{1+\tau} \mathbf{I} \right) + \sigma^2 \mathbf{V}_k^{\mathsf{H}}
$$
$$
- \frac{P}{1+\tau} \mathbf{U}_k^{\mathsf{H}} \widehat{\mathbf{H}}_{k,k} = \mathbf{0}
$$
$$
\Rightarrow \mathbf{V}_k = (1+\tau) \left(\sum_{j=1}^{K} \widehat{\mathbf{H}}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \widehat{\mathbf{H}}_{j,k} + \varepsilon_{\text{MMSE}}^{\text{V}} \mathbf{I} \right)^{-1} \widehat{\mathbf{H}}_{k,k}^{\mathsf{H}} \mathbf{U}_k \ ,
$$

$$(41)$$

where

$$
\varepsilon_{\text{MMSE}}^{\text{V}} = \gamma^{-1} (1+\tau)^2 + \tau (1+\tau) Kd \ . \quad (42)
$$

With respect to the fact that *unitary precoders and combiners* are more desirable, and since only imperfect channel estimates $\widehat{\mathbf{H}}_{k,j}$ are available, the proposed algorithm, which iteratively optimizes the precoders and combiners, can be concisely presented as follows:

MMSE Based IA

1: Set $\varepsilon_{\text{MMSE}} := \gamma^{-1} (1+\tau)^2 + \tau (1+\tau) Kd$
2: Initialize random unitary matrices $\mathbf{V}_k, \forall\, k$
3: $\mathbf{U}_k = (1+\tau) \left(\sum_{j=1}^{K} \widehat{\mathbf{H}}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \widehat{\mathbf{H}}_{k,j}^{\mathsf{H}} + \varepsilon_{\text{MMSE}} \mathbf{I} \right)^{-1} \widehat{\mathbf{H}}_{k,k} \mathbf{V}_k$
4: $\mathbf{U}_k \Longleftarrow \text{orth}\,(\mathbf{U}_k)$
5: $\mathbf{V}_k = (1+\tau) \left(\sum_{j=1}^{K} \widehat{\mathbf{H}}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \widehat{\mathbf{H}}_{j,k} + \varepsilon_{\text{MMSE}} \mathbf{I} \right)^{-1} \widehat{\mathbf{H}}_{k,k}^{\mathsf{H}} \mathbf{U}_k$
6: $\mathbf{V}_k \Longleftarrow \text{orth}\,(\mathbf{V}_k)$
7: Go to Step 3 and repeat

Remark 4: Similar to the same discussions as in Remark 2, the proposed MMSE based design can be applied to the case of perfect CSI by setting $\tau = 0$ and replacing $\widehat{\mathbf{H}}_{k,j}$ with $\mathbf{H}_{k,j}$.

V. DISCUSSIONS

A. CSI Overhead

To compare the amount of CSI overhead and without loss of generality, we consider the number of covariance matrices needed for IA algorithms. By comparing Min-WLI with LS based IA, one can conclude that to calculate one specific combiner, Min-WLI needs $K-1$ covariance matrices whereas LS based IA requires K covariance matrices. Therefore, the proposed LS based IA needs slightly more CSI to be available than Min-WLI, but it achieves better performance.

Both Max-SINR and the proposed MMSE based IA consider the effect of noise to derive precoders and combiners. However, while the former is based on a stream-by-stream approach, the latter is based on a user-by-user approach. As seen in Step 3 of Max-SINR and MMSE based IA algorithms, to calculate the kth combiner, Max-SINR needs $K+d-1$ covariance matrices whereas MMSE based IA requires K covariance matrices. Therefore, the proposed MMSE based IA needs less CSI to be available than Max-SINR. This can also be perceived with respect to the fact that the Max-SINR is a stream-by-stream approach such that each column of each precoder or combiner is derived separately of the other columns (as shown in Step 4 of the Max-SINR IA scheme) while MMSE based IA is a user-by-user approach such that all columns of each precoder or combiner are calculated together (as demonstrated in Step 3 of the MMSE-based IA); and as it is well known, the use-by-user-based IA schemes are more desirable than their stream-by-stream-based counterparts due to the reduced amount of required CSI [19].

B. Computational Complexity

In this subsection, we compare the computational complexity of the proposed IA algorithms with that of standard IA schemes. Without loss of generality, we consider beamformer design under perfect CSI.

Note that for both Min-WLI and LS based IA, the beamformers are designed independent of the nominal SNR, i.e., the precoders and combiners are once calculated and then can be used for any SNR, as long as all links are constant. Therefore, the computational complexity of these two schemes are comparable and none of them gives considerable advantage over the other in terms of decreasing the computational complexity. In other words, the computational complexity of these two schemes depends on the variation of the links rather than the SNR. However, this does not hold for the proposed MMSE based IA and Max-SINR, since the beamformers obtained by either of these algorithms are dependent on SNR. In other words, the beamformers are to be recalculated each time the nominal SNR changes. Therefore if any of these algorithms has a slight advantage over the other for one specific nominal SNR, this can lead to a huge reduction in computational complexity for a wide range of SNRs. To demonstrate this superior performance of the proposed MMSE based IA over Max-SINR, we consider the calculations involved in just one iteration of either of these two algorithms and for one specific combiner, i.e., the kth combiner. As demonstrated in Steps 3–4 of Max-SINR, to compute $\mathbf{U}_k$, we need to calculate d matrix inverses whereas for the proposed MMSE based IA, we have to calculate only one matrix inverse of the same size. The same is true to calculate the kth precoder $\mathbf{V}_k$. Therefore, within one iteration, the proposed MMSE based IA calculates 2 matrix inverses whereas Max-SINR calculates $2d$ matrix inverses.

By considering the fact that these algorithms need at least hundreds of iterations to efficiently calculate the precoders and combiners, the proposed MMSE based IA results in much less computational complexity compared to Max-SINR, and this automatically translates to the reduced running time for the proposed MMSE based IA.

C. The Equivalence of MMSE based IA and Max-SINR

In this subsection, it is shown that Max-SINR is equivalent to the proposed MMSE based IA under perfect CSI, i.e., these two algorithms achieve exactly the same beamformers under perfect CSI. To do so, we should demonstrate that in each iteration, the precoders and combiners obtained by one of these algorithms are the same as the other's. Therefore and without loss of generality, we just show that the kth combiner in Step 5 of Max-SINR is the same as the one in Step 4 of the proposed MMSE based IA. To do so, we assume that τ in the proposed MMSE based IA has been set to zero and $\widehat{\mathbf{H}}_{j,k}$ has been replaced by $\mathbf{H}_{j,k}$. To further proceed, we consider the following lemma [30]:

Lemma 5: If $\mathbf{a} \in \mathbb{C}^{M \times 1}$ and $\mathbf{A} \in \mathbb{C}^{M \times M}$ then

$$\left(\mathbf{A} - \mathbf{a}\mathbf{a}^{\mathsf{H}}\right)^{-1} \mathbf{a} = \frac{\mathbf{A}^{-1}\mathbf{a}}{1 - \mathbf{a}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{a}} \ . \qquad (43)$$

With respect to Lemma 5, $\left(\mathbf{T}_k^\ell\right)^{-1} \mathbf{H}_{k,k}\mathbf{v}_{k\ell}$ in Step 4 of Max-SINR algorithm can be rewritten as

$$\left(\mathbf{T}_k^\ell\right)^{-1} \mathbf{H}_{k,k}\mathbf{v}_{k\ell} =$$
$$\frac{\left(\sum_{j=1}^{K}\mathbf{H}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}} + \gamma^{-1}\mathbf{I}\right)^{-1}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}}{1 - \mathbf{v}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}^{\mathsf{H}}\left(\sum_{j=1}^{K}\mathbf{H}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}} + \gamma^{-1}\mathbf{I}\right)^{-1}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}} \ . \qquad (44)$$

By considering the fact that the ℓth column of $\mathbf{U}_k$ is equal to

$$\mathbf{u}_{k\ell} = \frac{\left(\mathbf{T}_k^\ell\right)^{-1}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}}{\left\|\left(\mathbf{T}_k^\ell\right)^{-1}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}\right\|_2} \ , \qquad (45)$$

and by horizontally concatenating the d columns of $\mathbf{U}_k$, we have

$$\mathbf{U}_k = \left(\sum_{j=1}^{K}\mathbf{H}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}} + \gamma^{-1}\mathbf{I}\right)^{-1}$$
$$\times \mathbf{H}_{k,k}\mathbf{V}_k \begin{pmatrix} t_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & t_d \end{pmatrix} \ , \qquad (46)$$

such that t_ℓ, $\ell = 1, \ldots, d$, is equal to

$$t_\ell = \frac{1}{\left\|\left(\sum_{j=1}^{K}\mathbf{H}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}} + \gamma^{-1}\mathbf{I}\right)^{-1}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}\right\|_2} \ . \qquad (47)$$

Note that although $\mathbf{U}_k$ in (46) has unit-norm columns, it is not unitary, i.e., $\mathbf{U}_k^{\mathsf{H}}\mathbf{U}_k \neq \mathbf{I}$. However, the orthogonalization in Step 5 of Max-SINR substitutes $\mathbf{U}_k$ with the unitary part of its QR decomposition which is exactly the same as the one obtained in Step 4 of the proposed MMSE based IA. Similarly, it can be readily proved that in each iteration, the kth precoder of Max-SINR is the same as the one obtained through MMSE based IA. Consequently, the two algorithms achieve the same precoders and combiners and thus achieve the same performance under perfect CSI.

D. Convergence of the Algorithms

By considering (15) and (28), it can be seen that the proposed LS and MMSE based IA schemes, respectively, try to minimize the objective functions F_{LS}^{U} and $F_{\text{MMSE}}^{\text{U}}$ through an iterative manner in order to derive the kth combiner $\mathbf{U}_k$ when the matrices $\mathbf{V}_j$, $j = 1, \ldots, K$ are fixed. In this case and upon meeting the feasibility conditions [20], the objective functions in (16) and (29) become quadratic and therefore they have a solution. However, their solutions depend on the initialization such that by changing the fixed values of $\mathbf{V}_j$, $j = 1, \ldots, K$, different solutions can be obtained. Besides, since these objective functions represents the amount of leakage interference, the minimization of the objective functions implies on the minimization of the leakage interference which gradually leads to the convergence of the algorithms to local minima so that these local minima are dependent on the initial values of $\mathbf{V}_j$, $j = 1, \ldots, K$.

Moreover, it is worthy to note that as discussed in [19], [31], in all practical scenarios, the Max-SINR algorithm seems to converge to local minima. Consequently, since in Section V-C, it was mathematically proved that under perfect CSI the MMSE based IA is equivalent to the Max-SINR, one can conclude that the MMSE based IA converges to local minima as well. Furthermore, due to the fact that the LS based IA is closely related to the MMSE based IA, it is straightforward to deduce that the LS based IA is also convergent to local minima.

E. On Diagonalized Subchannels

In this part, it is shown that the proposed LS based design leads to diagonalized subchannels for all SNR ranges. In other words, using LS based IA, $\mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{V}_k$, $k = 1, \ldots, K$ are diagonal matrices. This is in contrast to the previously proposed IA schemes wherein the decoupled subchannels are full matrices. This dissolve the need of SVD for the LS based IA in order to employ waterfilling since the decoupled subchannels have been already diagonalized. Therefore, in the sequel, it is shown that $\mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{V}_k$ is a diagonal matrix,

provided that $\mathbf{U}_k$ and $\mathbf{V}_k$ are obtained by LS based IA. Since, this condition is met under the assumption of perfect CSI, we assume that in the proposed LS based IA, τ has been set to zero and all imperfect CSI $\widehat{\mathbf{H}}_{k,j}$ have been replaced by the perfect CSI, i.e., $\mathbf{H}_{k,j}$.

With respect to Step 3 of LS based IA algorithm, we have

$$\left(\sum_{j=1}^{K} \mathbf{H}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \mathbf{H}_{k,j}^{\mathsf{H}} \right)^{-1} \mathbf{H}_{k,k} \mathbf{V}_k$$

$$= \left(\sum_{\substack{j=1 \\ j \neq k}}^{K} \mathbf{H}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \mathbf{H}_{k,j}^{\mathsf{H}} + \mathbf{H}_{k,k} \mathbf{V}_k \mathbf{V}_k^{\mathsf{H}} \mathbf{H}_{k,k}^{\mathsf{H}} \right)^{-1} \mathbf{H}_{k,k} \mathbf{V}_k$$

$$= \left(\boldsymbol{\mathcal{V}}_{-k} \boldsymbol{\mathcal{V}}_{-k}^{\mathsf{H}} + \boldsymbol{\mathcal{V}}_k \boldsymbol{\mathcal{V}}_k^{\mathsf{H}} \right)^{-1} \boldsymbol{\mathcal{V}}_k , \qquad (48)$$

where $\boldsymbol{\mathcal{V}}_k = \mathbf{H}_{k,k} \mathbf{V}_k \in \mathbb{C}^{M \times d}$ and

$$\boldsymbol{\mathcal{V}}_{-k} = \Big[\mathbf{H}_{k,1} \mathbf{V}_1, \ldots, \mathbf{H}_{k,k-1} \mathbf{V}_{k-1},$$

$$\mathbf{H}_{k,k+1} \mathbf{V}_{k+1}, \ldots, \mathbf{H}_{k,K} \mathbf{V}_K \Big] \in \mathbb{C}^{M \times (K-1)d} . \quad (49)$$

By considering the fact that the feasibility conditions of IA have been met [20], [32] and the interferences have been consequently aligned within the reduced subspace of the received signal space, the components of the desired space $\boldsymbol{\mathcal{V}}_k$ become linearly independent of the components of the interference space $\boldsymbol{\mathcal{V}}_{-k}$ such that [1]

$$\text{rank} \left(\boldsymbol{\mathcal{V}}_k \boldsymbol{\mathcal{V}}_k^{\mathsf{H}} \in \mathbb{C}^{M \times M} \right) = d , \qquad (50a)$$

$$\text{rank} \left(\boldsymbol{\mathcal{V}}_{-k} \boldsymbol{\mathcal{V}}_{-k}^{\mathsf{H}} \in \mathbb{C}^{M \times M} \right) = M - d . \qquad (50b)$$

To further continue, we consider the following lemma:

Lemma 6: Let $\mathbf{A}$ be a horizontal concatenation of two sub-matrices, i.e., $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2] \in \mathbb{C}^{M \times M}$ such that $\mathbf{A}_1 \in \mathbb{C}^{M \times a}$ and $\mathbf{A}_2 \in \mathbb{C}^{M \times (M-a)}$ have independent columns. In this case, the ZF condition implies that [33]

$$\mathbf{A}_1^{\mathsf{H}} \left(\mathbf{A} \mathbf{A}^{\mathsf{H}} \right)^{-1} \mathbf{A}_1 = \mathbf{A}_1^{\mathsf{H}} \left(\mathbf{A}_1 \mathbf{A}_1^{\mathsf{H}} + \mathbf{A}_2 \mathbf{A}_2^{\mathsf{H}} \right)^{-1} \mathbf{A}_1 = \mathbf{I}_{a \times a} . \qquad (51)$$

Hence, due to Lemma 6 and with respect to the fact that the interferences have been aligned within a $(M-d)$-dimensional subspace of the M-dimensional received signal space, we have

$$\boldsymbol{\mathcal{V}}_k^{\mathsf{H}} \left(\boldsymbol{\mathcal{V}}_{-k} \boldsymbol{\mathcal{V}}_{-k}^{\mathsf{H}} + \boldsymbol{\mathcal{V}}_k \boldsymbol{\mathcal{V}}_k^{\mathsf{H}} \right)^{-1} \boldsymbol{\mathcal{V}}_k$$

$$= \mathbf{V}_k^{\mathsf{H}} \mathbf{H}_{k,k}^{\mathsf{H}} \left(\sum_{\substack{j=1 \\ j \neq k}}^{K} \mathbf{H}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \mathbf{H}_{k,j}^{\mathsf{H}} + \mathbf{H}_{k,k} \mathbf{V}_k \mathbf{V}_k^{\mathsf{H}} \mathbf{H}_{k,k}^{\mathsf{H}} \right)^{-1}$$

$$\times \mathbf{H}_{k,k} \mathbf{V}_k = \mathbf{I}_{d \times d} . \qquad (52)$$

We further consider the QR decomposition of $\left(\boldsymbol{\mathcal{V}}_{-k} \boldsymbol{\mathcal{V}}_{-k}^{\mathsf{H}} + \boldsymbol{\mathcal{V}}_k \boldsymbol{\mathcal{V}}_k^{\mathsf{H}} \right)^{-1} \boldsymbol{\mathcal{V}}_k$ as

$$\left(\boldsymbol{\mathcal{V}}_{-k} \boldsymbol{\mathcal{V}}_{-k}^{\mathsf{H}} + \boldsymbol{\mathcal{V}}_k \boldsymbol{\mathcal{V}}_k^{\mathsf{H}} \right)^{-1} \boldsymbol{\mathcal{V}}_k = \mathbf{U}_k \boldsymbol{\mathcal{R}}_k , \qquad (53)$$

where $\mathbf{U}_k$ is the unitary part. Therefore with respect to (52), we have

$$\mathbf{U}_k^{\mathsf{H}} \boldsymbol{\mathcal{V}}_k = \mathbf{U}_k^{\mathsf{H}} \mathbf{H}_{k,k} \mathbf{V}_k = \left(\boldsymbol{\mathcal{R}}_k^{\mathsf{H}} \right)^{-1} . \qquad (54)$$

However, note that due to the properties of IA and LS criterion [3], [34], $\boldsymbol{\mathcal{R}}_k$ is now a diagonal matrix instead of an upper triangular. This diagonality can also be perceived with respect to the fact that by using IA, the K-use IC is decoupled into K parallel point-to-point MIMO systems, and as it is well-known, LS based design results in diagonalized point-to-point communications [34]. Furthermore, since the diagonal elements of the triangular part of the QR decomposition of any matrix are real numbers, this implies that $\mathbf{U}_k^{\mathsf{H}} \mathbf{H}_{k,k} \mathbf{V}_k$ is a diagonal matrix consists of real values.

It is also worthwhile to point out that since at high enough SNRs, MMSE based IA boils down to LS based IA, the resulted subchannels of MMSE based IA become diagonalized at sufficiently high SNRs.

VI. NUMERICAL RESULTS

In this section and by using numerical results, we corroborate the improved performance achieved by the proposed LS and MMSE based IA compared to standard IA techniques. To have a fair comparison, we compare LS based IA with Min-WLI since both of these algorithms do not consider the effect of noise to design beamformers. We also compare MMSE based IA with Max-SINR since both of these algorithms consider the effect of noise for beamformer design.

Without loss of generality, we just consider a symmetric constant MIMO IA with $K = 4$ users each with $d = 2$ independent data streams. To meet the sufficient conditions of feasibility for IA, we respectively set the number of receive and transmit antennas as $M = 4$, $N = 6$ [32].

Regarding the performance analysis under imperfect CSI, although the promised improvement of the proposed LS and MMSE based IA can be gleaned for various values of α, we focus on two representative cases [22], [24]: $\alpha = 0$ (which imitates the CSI feedback scenario), and $\alpha = 1$ (which mimics the reciprocal channels).

The performance trend of the proposed LS and MMSE based IA under perfect CSI can be obtained with respect to Remark 2 and Remark 4, respectively.

Plus, by considering i.i.d. Gaussian input signaling and uniform power allocation, we evaluate the achievable sum rates as [21], [22]

$$R = \sum_{k=1}^{K} \log_2 \det \left(\mathbf{I} + \left(\gamma^{-1} \mathbf{I} + \sum_{j=1, j \neq k}^{K} \boldsymbol{\Phi}_{k,j} \right)^{-1} \boldsymbol{\Phi}_{k,k} \right) , \qquad (55)$$

where $\boldsymbol{\Phi}_{k,j} = \mathbf{U}_k^{\mathsf{H}} \mathbf{H}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \mathbf{H}_{k,j}^{\mathsf{H}} \mathbf{U}_k$, such that in the case of imperfect CSI, all precoders and combiners are constructed based on erroneous channel estimations in (4).

In Fig. 2, we consider one fixed random CSI initialization for the case $K = 4$, $d = 2$, and a fixed SNR of 20 dB. We also assume that perfect CSI is available. For this scenario, we ran Min-WLI, LS and MMSE based IA algorithms 100 times,
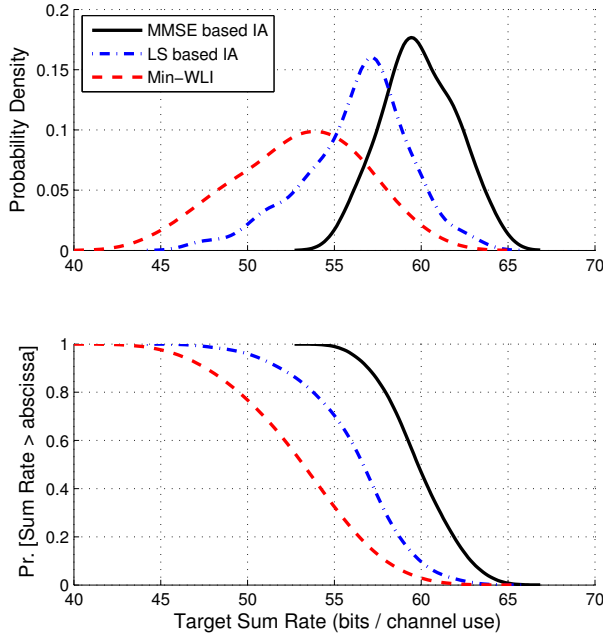
Fig. 2. a) Probability density and b) Complementary cumulative distribution of the sum rate for the solutions obtained from different random precoder initializations for Min-WLI, LS and MMSE based IA for the case $K = 4$, $d = 2$, $M = 4$, $N = 6$. The results are depicted at SNR of 20 dB and under one perfect CSI realization.

each run starting from different random unitary precoders. In other words, all these algorithms have been initialized over the same set of precoders. Without loss of generality and similar to [12], the number of iterations for each algorithm was 3000, which assures that the interferences are almost perfectly aligned within a reduced subspace of each received signal space.

Fig. 2.a depicts the probability density of the sum rate for three IA algorithm. As shown, although all algorithms have been initialized from the same unitary precoders, the final sum rates of the proposed LS and MMSE based algorithms are more concentrated around higher values, which implies that with the same initializations and under the same channel realization, the mean sum rate achieved by LS and MMSE based IA is higher than that of Min-WLI.

Fig. 2.b illustrates the complementary cumulative distributions (CCDs) of the sum rate for three IA algorithm. As revealed, although all algorithms have been initialized from the same unitary precoders, the proposed algorithms achieve better outage performance than Min-WLI.

Fig. 3 shows the convergence of the sum rate of Min-WLI, LS and MMSE based IA algorithms (averaged over 200 channel realizations) for the case $K = 4$, $d = 2$ at SNR of 20 dB, and under perfect CSI and imperfect CSI with $\beta = 10, \alpha = 1$. As revealed, under both perfect and imperfect CSI, the proposed LS and MMSE based IA have the same convergence rate as Min-WLI. Moreover, for any number of
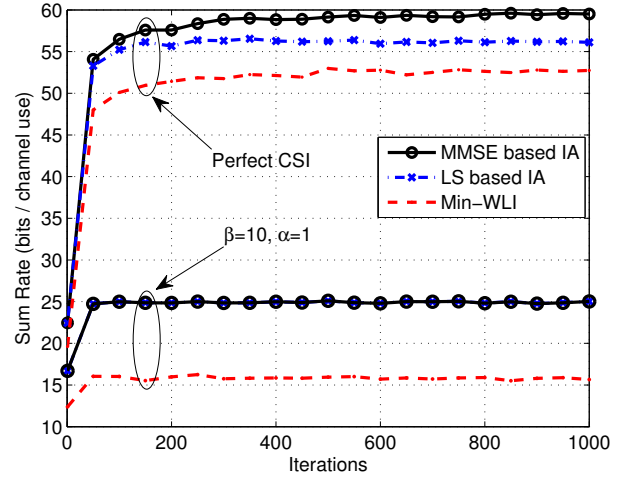

Fig. 3. Convergence of sum rate for Min-WLI, LS and MMSE based IA for the case $K = 4$, $d = 2$, $M = 4$, $N = 6$ at SNR of 20 dB averaged over 200 channel realizations. The results are related to both perfect CSI and imperfect CSI with $\beta = 10, \alpha = 1$.
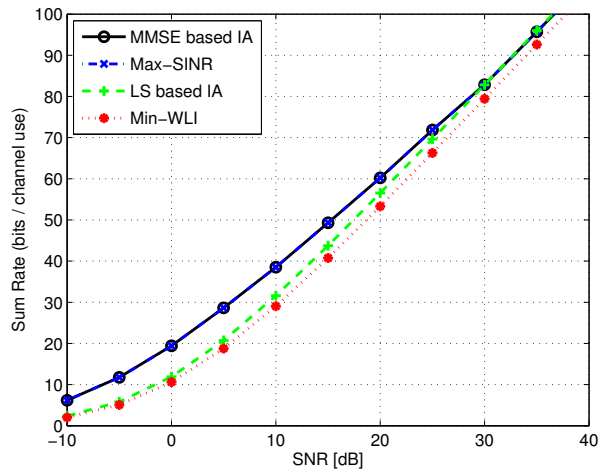

Fig. 4. Average sum rate for $K = 4$, $d = 2$, $M = 4$, $N = 6$ and under perfect CSI.

iterations, they achieve higher sum rates than Min-WLI.

Fig. 4 illustrates the sum rate under perfect CSI. As revealed, in the presence of perfect CSI, MMSE based IA achieves the same sum rate as Max-SINR while outperforming both LS based IA and Min-WLI. However, the proposed LS based IA outperforms Min-WLI such that the achieved gain in sum rate is 3 bits per channel use at intermediate SNRs.

Figs. 5 and 6, depict the average sum rate under imperfect CSI cases $\beta = 10$, $\alpha = 1$ and $\beta = 0.1$, $\alpha = 0$, respectively. As shown, while under perfect CSI, MMSE based IA achieves the same sum rate as Max-SINR, under imperfect CSI, MMSE based IA outperforms Max-SINR. Also the proposed LS based IA is able to achieve better performance than Min-WLI. For
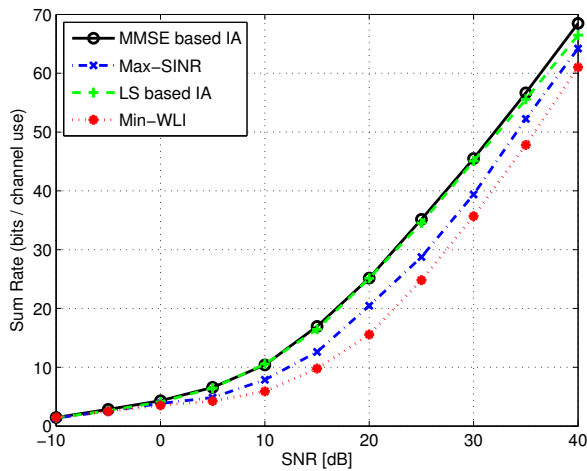
Fig. 5. Average sum rate for $K = 4$, $d = 2$, $M = 4$, $N = 6$ when $\beta = 10$, $\alpha = 1$.
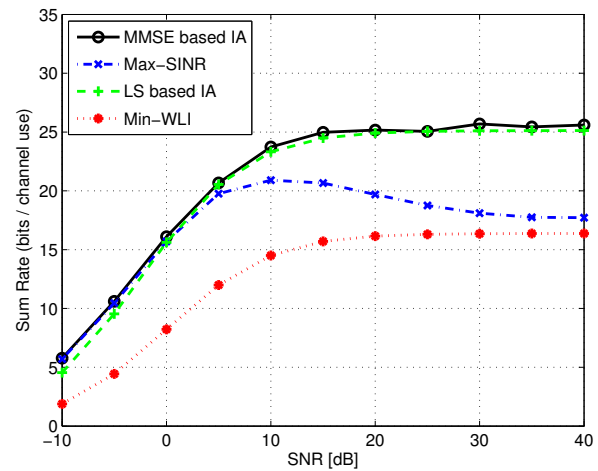

Fig. 6. Average sum rate for $K = 4$, $d = 2$, $M = 4$, $N = 6$ when $\beta = 0.1$, $\alpha = 0$.

example, for the case of $\beta = 10$, $\alpha = 1$ and at SNR of 30 dB, LS and MMSE based IA achieve 10 and 7 bits per channel use gain in sum rate compared to Min-WLI and Max-SINR, respectively. Similarly, for the case of $\beta = 0.1$, $\alpha = 0$, and at SNR of 30 dB, LS and MMSE based IA achieve 9 and 7 bits per channel use gain in sum rate compared to Min-WLI and Max-SINR algorithms, respectively.

As revealed in Figs. 5 and 6, the proposed LS based IA is able to achieve almost the same performance as the MMSE based IA under imperfect CSI. Also, it achieves better performance than Max-SINR for $\alpha = 1$. However, for $\alpha = 0$, while at low SNRs, Max-SINR outperforms LS based IA, the latter achieves better performance than the former at high SNRs.

Fig. 7 illustrates the average SER of Min-WLI, Max-SINR and the proposed MMSE based IA for $K = 4$, $d = 2$ under the imperfect CSI cases $\beta = 10$, $\alpha = 1$ and $\beta = 0.05$, $\alpha = 0$. We assumed that each transmitted block consists of 100 QPSK symbols. As seen, MMSE based IA outperforms both Min-WLI and Max-SINR. For example, when $\beta = 10$, $\alpha = 1$, MMSE based IA respectively achieves 18 dB and 14 dB gain compared to Min-WLI and Max-SINR to reach the SER of 10^{-3}. Also for the case $\beta = 0.05$, $\alpha = 0$, the MMSE based IA decreases the SER by a factor of at least $\frac{1}{10}$ compared to Min-WLI and Max-SINR at SNRs of larger than 30 dB.

From Figs. 6–7, one interesting observation is that, for $\alpha = 0$, the performance trend of Max-SINR is nonmonotonic whereas that of MMSE based IA is monotonic. The reason is that under CSI feedback and at high SNRs, the system becomes interference-limited, therefore by increasing the SNR, the performance trend does not constantly improve and eventually becomes saturated. This implies that for the Max-SINR and by increasing the nominal SNR, the performance first becomes improved at low-to-intermediate SNRs, but suddenly deteriorates at a saddle point and eventually becomes saturated at high SNRs, which causes a nonmonotonic behavior. For the
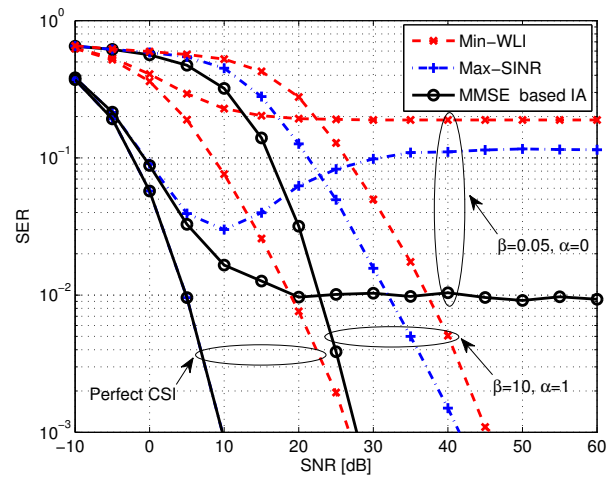

Fig. 7. Average SER for $K = 4$, $d = 2$, $M = 4$, $N = 6$ under the imperfect CSI cases $\beta = 10$, $\alpha = 1$ and $\beta = 0.05$, $\alpha = 0$. Each transmitted block consists of 100 QPSK symbols.

MMSE based IA, on the other hand, the performance trend does not suddenly deteriorate. This is due to the regularization parameter $\varepsilon_{\text{MMSE}}$ which is a function of the channel estimation error variance τ. Thus, in this case, the performance trend first improves at low-to-intermediate SNRs, and smoothly becomes saturated at high SNR, without any sudden turnover. It is worthy to note that the nonmonotonic behavior of Max-SINR algorithm and the monotonic behavior of MMSE based IA are respectively similar to those of conventionally regularized precoders [28] and adaptively regularized precoders [24] in downlink communications.

Also by considering the slope of the curves in both Figs. 5 and 7, another interesting point is that when $\alpha = 1$, i.e., the error variance scales with the inverse of SNR, full multiplexing

gain can be preserved since the curves related to the reciprocal channel have the same slope as those related to the perfect CSI at high SNRs. Also for the case of the CSI feedback, i.e., $\alpha = 0$, the achievable DoF is zero since the corresponding curves become saturated at high SNRs.

VII. Conclusions

In the presence of perfect CSI, interference alignment enables us to achieve full DoF. However, subject to imperfect CSI, full benefits of IA get compromised. In this paper, we proposed two novel IA algorithms such that the optimization criteria were set up based on the knowledge of imperfect CSI. Unlike the other robust algorithms, the proposed adaptive schemes do not impose extra computational complexity compared to the nonadaptive ones. While the LS based IA does not consider the effect of noise to design beamformers, the MMSE based IA does so. This causes the latter to outperform the former under both perfect and imperfect CSI. We also compared the proposed algorithms with standard IA methods. It was shown that the LS based IA outperforms interference leakage minimization algorithms under both perfect and imperfect CSI. However, while MMSE based IA achieves the same performance as Max-SINR under perfect CSI, the former outperforms the latter subject to imperfect CSI. We showed that even with this superior performance, the proposed MMSE based IA needs less CSI to be available and has less computational complexity compared to Max-SINR.

However, like many earlier IA algorithms, the proposed schemes (especially under perfect CSI) need hundreds of iterations to converge. Therefore one interesting direction for future work is to design IA algorithms that can converge with much fewer number of iterations, which eventually makes them more practical due to less computational complexity and the reduced running time.

References

[1] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[2] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3309–3322, Jun. 2011.

[3] S. W. Peters and R. W. Heath, "Cooperative algorithms for MIMO interference channels," *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, pp. 206–218, Jan. 2011.

[4] H. Yu and Y. Sung, "Least squares approach to joint beam design for interference alignment in multiuser multi-input multi-output interference channels," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4960–4966, Sep. 2010.

[5] D. S. Papailiopoulos and A. G. Dimakis, "Interference alignment as a rank constrained rank minimization," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4278–4288, Aug. 2012.

[6] H. Du, T. Ratnarajah, M. Sellathurai, and C. B. Papadias, "Reweighted nuclear norm approach for interference alignment," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3754–3765, Sep. 2013.

[7] S.-H. Park, H. Park, H. Sung, and I. Lee, "Regularized transceiver designs for multi-user MIMO interference channels," *IEEE Trans. Commun.*, vol. 60, no. 9, pp. 2571–2579, Sep. 2012.

[8] S. M. Razavi and T. Ratnarajah, "Interference alignment in K-user MIMO interference channels with partially coordinated receivers," *IET Commun.*, vol. 8, no. 1, pp. 50–57, Jan. 2014.

[9] S. W. Peters and R. W. Heath, "Interference alignment via alternating minimization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2009, pp. 2445–2448.

[10] K. R. Kumar and F. Xue, "An iterative algorithm for joint signal and interference alignment," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2010, pp. 2293–2297.

[11] D. A. Schmidt, C. Shi, R. A. Berry, M. L. Honig, and W. Utschick, "Minimum mean squared error interference alignment," in *Proc. Asilomar Conf. Signal Syst. Comput. (ACSSC)*, 2009, pp. 1106–1110.

[12] I. Santamaria, O. Gonzalez, R. W. Heath, and S. W. Peters, "Maximum sum-rate interference alignment algorithms for MIMO channels," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2010.

[13] S.-H. Park, H. Park, Y.-D. Kim, and I. Lee, "Regularized interference alignment based on weighted sum-MSE criterion for MIMO interference channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2010, pp. 1–5.

[14] J.-S. Kim, S.-H. Moon, S.-R. Lee, and I. Lee, "A new channel quantization strategy for MIMO interference alignment with limited feedback," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 358–366, Jan. 2012.

[15] X. Rao, L. Ruan, and V. K. N. Lau, "CSI feedback reduction for MIMO interference alignment," *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4428–4437, Sep. 2013.

[16] S. Cho, K. Huang, D. K. Kim, V. K. N. Lau, H. Chae, H. Seo, and B.-H. Kim, "Feedback-topology designs for interference alignment in MIMO interference channels," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6561–6575, Dec. 2012.

[17] J. Leithon, C. Yuen, H. A. Suraweera, and H. Gao, "A new opportunistic interference alignment scheme and performance comparison of MIMO interference alignment with limited feedback," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2012, pp. 1123–1127.

[18] X. Chen and C. Yuen, "Performance analysis and optimization for interference alignment over MIMO interference channels with limited feedback," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1785–1795, Apr. 2014.

[19] J. Park, Y. Sung, and H. V. Poor, "On beamformer design for multiuser MIMO interference channels." [Online]. Available: http://arxiv.org/abs/1011.6121v1

[20] C. M. Yetis, T. Gou, S. A. Jafar, and A. H. Kayran, "On feasibility of interference alignment in MIMO interference networks," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4771–4782, Sep. 2010.

[21] R. Tresch and M. Guillaud, "Cellular interference alignment with imperfect channel knowledge," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2009.

[22] S. M. Razavi and T. Ratnarajah, "Performance analysis of interference alignment under CSI mismatch," *IEEE Trans. Veh. Technol.*, vol. 63, no. 9, pp. 4740–4748, Nov. 2014.

[23] ——, "Performance analysis of Max-SINR algorithm under CSI mismatch," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 7729–7733, 2014.

[24] ——, "Adaptively regularized phase alignment precoding for multiuser multiantenna downlink," *IEEE Trans. Veh. Technol.* [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6963415

[25] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[26] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. New Jersey: Prentice-Hall, 1993.

[27] R. Tresch, M. Guillaud, and E. Riegler, "On the achievability of interference alignment in the K-user constant MIMO interference channel," in *Proc. IEEE/SP Workshop Stat. Signal Process.*, 2009, pp. 277–280.

[28] S. M. Razavi, T. Ratnarajah, and C. Masouros, "Transmit-power efficient linear precoding utilizing known interference for the multiantenna downlink," *IEEE Trans. Veh. Technol.*, vol. 63, no. 9, pp. 4383–4394, Nov. 2014.

[29] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.

[30] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore and London: The Johns Hopkins Univ. Press, 1996.

[31] C. Wilson and V. V. Veeravalli, "A convergent version of the Max SINR algorithm for the MIMO interference channel," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2952–2961, Jun. 2013.

[32] M. Razaviyayn, G. Lyubeznik, and Z.-Q. Luo, "On the degrees of freedom achievable through interference alignment in a MIMO interference channel," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 812–821, Feb. 2012.

[33] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[34] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3221–3231, Dec. 1994.

S. Morteza Razavi received the Ph.D. degree in Electrical Engineering from the University of Edinburgh, U.K., in 2014.

His research interests are in the broad areas of signal processing and wireless communications.

Tharmalingam Ratnarajah (A'96–M'05–SM'05) is currently with the Institute for Digital Communications, the University of Edinburgh, Edinburgh, U.K., as a Reader in signal processing and communications.

Since 1993, he has held various positions with the Queen's University Belfast, Belfast, U.K, University of Ottawa, Ottawa, ON, Canada, Nortel Networks, Ottawa, ON, Canada, McMaster University, Hamilton, ON, Canada, and Imperial College, London, U.K. His research interests include random matrices theory, information theoretic aspects of MIMO channels and ad hoc networks, wireless communications, signal processing for communication, statistical and array signal processing, biomedical signal processing, and quantum information theory. He has published over 200 publications in these areas and holds four U.S. patents. He is currently the coordinator of the FP7 FP7 projects HARP (3.2M€) in the area of highly distributed MIMO and ADEL (2.6M€) in the area of licensed shared access. Previously, he was the coordinator of FP7 Future and Emerging Technologies projects CROWN (2.3M€) in the area of cognitive radio networks and HIATUS (2.7M€) in the area of interference alignment. Dr. Ratnarajah is a member of the American Mathematical Society and Information Theory Society and Fellow of Higher Education Academy (FHEA), UK.