MMSE Receiver Design and SINR Calculation in MU-MIMO Systems with Imperfect CSI

Andrea Abrardo, Gábor Fodor, Marco Moretti, Miklós Telek

***Abstract*—The performance of the uplink of multiuser multiple input multiple output systems depends critically on the receiver architecture and on the quality of the acquired channel state information. A popular approach is to design linear receivers that minimize the mean squared error (MSE) of the received data symbols. Unfortunately, most of the literature does not take into account the presence of channel state information errors in the MSE minimization. In this letter we develop a linear minimum MSE (MMSE) receiver that employs the noisy instantaneous channel estimates to minimize the MSE, and highlight the dependence of the receiver performance on the pilot-to-data power ratio. By invoking the theory of random matrices, we calculate the users' signal-to-interference-plus-noise ratio as a function of the number of antennas and the pilot-to-data power ratio of all users. Numerical results indicate that this new linear receiver outperforms the classical mismatched MMSE receiver.**

I. Introduction

In single input single output (SISO) and multiple input multiple output (MIMO) systems, the trade-off between spending resources on channel state information (CSI) acquisition and data transmission is known to affect the performance in terms of spectral and energy efficiency [1]. Therefore, balancing the pilot-to-data power ratio (PDPR) and deriving the optimal pilot symbol assisted modulation (PSAM) signal detector are important design questions [2]. Specifically for SISO systems, the works reported in [1, 2] showed that the optimal PSAM detectors operate iteratively and exchange information between the channel estimator and decoder.

For MIMO systems, the effects of imperfect estimation when using maximum likelihood (ML)-based or iterative receivers that simultaneously process the pilot and data symbols are studied in the seminal papers [3, 4]. Similarly, reference [5] considers a single user MIMO system that uses symbol aided channel estimation and iterative (turbo) detection. None of the above referenced papers aims at designing a linear minimum mean squared error (MMSE) receiver for multiuser multiple input multiple output (MU-MIMO) systems. In MU-MIMO systems, determining the number of pilot and data symbols and allocating transmit power to pilot and data signals must take into account the number of antennas and the number of served users [6, 7].

A closely related and ongoing line of research develops and extends a general analytic framework for MIMO MMSE receivers when perfect channel state information is available at the receiver [8]. In contrast, the recent and closely related works reported in [9, 10] considered the case when only non-perfect channel estimates are available at the receiver. In our recent work [10], we have proposed a modified MU-MIMO MMSE receiver, which compensates for the channel estimation errors. However, the proposed MMSE receiver uses the covariance matrices of the interfering channels, rather than exploiting the knowledge of all instantaneous channel estimates.

The objective of the present paper is to develop a MU-MIMO linear MMSE receiver that takes advantage of the instantaneous channel estimates of all users and, accordingly, minimizes the mean squared error (MSE) of the received data symbols. Thus, our contribution – captured by Proposition 1 and Theorem 3 – to the existing literature summarized above is two-fold:

1) Proposition 1 derives the actual MMSE receiver that, in contrast to the classical mismatched or *naïve* formula [3, 9], minimizes the MSE of the estimated uplink data symbols in the presence of channel estimation errors.
2) Theorem 3 derives an equation, whose solution gives the asymptotic average signal-to-interference-plus-noise ratio (SINR) of any user as a function of not only the PDPR, but also the number of antennas.

II. Channel Estimation

We consider the uplink of a MU-MIMO system, in which K mobile stations (MSs) transmit orthogonal pilot sequences $\mathbf{s} = \left[s_1, ..., s_{\tau_p}\right]^T \in \mathbb{C}^{\tau_p \times 1}$, in which each pilot symbol is scaled as $|s_i|^2 = 1$, for $i = 1, .., \tau_p$. We assume a comb type arrangement of the pilot symbols: given F subcarriers in the coherence bandwidth, a fraction of τ_p subcarriers are allocated to the pilot and $F_d = F - \tau_p$ subcarriers are allocated to the data symbols. Each MS transmits at a constant power P_{tot} and the transmission power can be distributed unequally in each subcarrier. In particular, considering a transmitted power P_p for each pilot symbol and P for each data symbol transmission, the sum constraint of $\tau_p P_p + (F - \tau_p) P = P_{tot}$ is enforced. This model can represent a multicarrier Long Term Evolution (LTE) system, in which the available bandwidth is organized into physical resource blocks (PRBs) comprising $F = 12$ subcarriers. Each PRB represents a coherent bandwidth chunk and obtains its own channel estimate [11].

The base station (BS) is equipped with N_r antennas, so that the $N_r \times \tau_p$ matrix $\mathbf{Y}^{(p)}$ of the received pilot signal at the BS can be written as

$$\mathbf{Y}^{(p)} = \sum_{k=1}^{K} \alpha_k \sqrt{P_p} \mathbf{h}_k \mathbf{s}_k^T + \mathbf{N}, \tag{1}$$

where $\mathbf{h}_k$ models the small scale fading channel for user k, α_k accounts for the propagation loss and $\mathbf{N} \in \mathbb{C}^{N_r \times \tau_p}$ is the spatially and temporally additive white Gaussian noise with element-wise variance σ_p^2.

A. Abrardo is with Dipartimento di Ingegneria del'Informazione, University of Siena, Italy. G. Fodor is with Ericsson Research and KTH Royal Institute of Technology, Stockholm, Sweden. M. Moretti is with Dipartimento di Ingegneria del'Informazione, University of Pisa, Italy. M. Telek is with Budapest University of Technology and Economics, and with the MTA-BME Information Systems Research Group, Budapest, Hungary. M. Telek is partially supported by the OTKA K123914 grant. Corresponding author: G. Fodor

We assume that the large scale fading parameters α_k and the pilot transmit power P_p are known or can be well estimated at the BS. This is a reasonable assumption in cellular systems in which mobility management relies on measurement reporting and signaling protocols that facilitate the acquisition of the slow varying propagation parameters [11].

We assume that $\mathbf{h}_k \in \mathbb{C}^{N_r \times 1}$ is circularly symmetric complex normal distributed with mean $\mathbf{0}$ and $N_r \times N_r$ covariance matrix $\mathbf{C}_k$ and that the elements on the main diagonal of $\mathbf{C}_k$ are all equal to one. Exploiting the pilot sequence orthogonality and assuming to know the values of α_ℓ and P_p, the least squares estimate of the channel of user ℓ is

$$\hat{\mathbf{h}}_\ell = \frac{1}{\alpha_\ell \sqrt{P_p} \tau_p} \mathbf{Y}^{(p)} \mathbf{s}_\ell^* = \mathbf{h}_\ell + \frac{1}{\alpha_\ell \sqrt{P_p} \tau_p} \mathbf{N} \mathbf{s}_\ell^*, \tag{2}$$

where $\mathbf{s}_\ell \in \mathbb{C}^{\tau_p \times 1}$ is the vector of pilot symbols for user ℓ and $(\mathbf{s}_\ell^T \mathbf{s}_\ell^*) = \tau_p$. Since $\mathbf{h}_\ell \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_\ell)$, the estimated channel $\hat{\mathbf{h}}_\ell$ is a circular symmetric complex normal distributed vector $\hat{\mathbf{h}}_\ell \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_\ell)$, with

$$\mathbf{R}_\ell \triangleq \mathbb{E}\{\hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H\} = \mathbf{C}_\ell + \frac{\sigma_p^2}{\alpha_\ell^2 P_p \tau_p} \mathbf{I}_{N_r}, \tag{3}$$

where $\mathbf{I}_{N_r}$ is the identity matrix of size $N_r \times N_r$. As it was shown in [7], the distribution of the channel realization $\mathbf{h}_\ell$ conditioned on the estimate $\hat{\mathbf{h}}_\ell$ is normally distributed as

$$(\mathbf{h}_\ell \mid \hat{\mathbf{h}}_\ell) \sim \mathbf{D}_\ell \hat{\mathbf{h}}_\ell + \underbrace{\mathcal{CN}\left(\mathbf{0}, \mathbf{Q}_\ell\right)}_{\text{channel estimation noise}} \tag{4}$$

where $\mathbf{D}_\ell \triangleq \mathbf{C}_\ell \mathbf{R}_\ell^{-1}$ and $\mathbf{Q}_\ell \triangleq \mathbf{C}_\ell - \mathbf{C}_\ell \mathbf{R}_\ell^{-1} \mathbf{C}_\ell$. Note that the circular complex $\mathcal{CN}\left(\mathbf{0}, \mathbf{Q}_\ell\right)$ term, that characterizes the channel estimation error, can be thought of as the zero-mean estimation noise, which can be made small by sufficiently increasing the pilot power.

III. Minimum MSE Receiver

The $N_r \times 1$-dimensional received signal at the BS can be written as:

$$\mathbf{y} = \underbrace{\alpha_\ell \mathbf{h}_\ell \sqrt{P_\ell} x_\ell}_{\text{user } \ell} + \underbrace{\sum_{k \neq \ell}^{K} \alpha_k \mathbf{h}_k \sqrt{P_k} x_k}_{\text{other users}} + \mathbf{n}_d, \tag{5}$$

where x_k is the transmitted data symbol by the kth user and $\mathbf{n}_d$ is the thermal noise on the received data signal.

In this letter we assume that the BS applies a linear receiver $\mathbf{G}_\ell \in \mathbb{C}^{1 \times N_r}$ to estimate the transmitted data symbols of user ℓ. In the case of perfect knowledge of the channel gains, the MSE of the received data symbols, averaged over the transmitted symbols and the thermal noise, can then be written as [10]:

$$\begin{aligned}
&\text{MSE}\left(\mathbf{G}_\ell, \mathbf{H}\right) = \mathbb{E}_{x, \mathbf{n}_d} \left\{ |\mathbf{G}_\ell \mathbf{y} - x_\ell|^2 \right\} \\
&= \left| \mathbf{G}_\ell \alpha_\ell \mathbf{h}_\ell \sqrt{P_\ell} - 1 \right|^2 + \sum_{k \neq \ell}^{K} P_k |\mathbf{G}_\ell \alpha_k \mathbf{h}_k|^2 + \sigma_d^2 \mathbf{G}_\ell \mathbf{G}_\ell^H \\
&= 1 - \alpha_\ell \sqrt{P_\ell} \mathbf{G}_\ell \mathbf{h}_\ell - \alpha_\ell \sqrt{P_\ell} \mathbf{h}_\ell^H \mathbf{G}_\ell^H \\
&\quad + \mathbf{G}_\ell \left(\sum_{k=1}^{K} \alpha_k^2 P_k \mathbf{h}_k \mathbf{h}_k^H + \sigma_d^2 \mathbf{I}_{N_r} \right) \mathbf{G}_\ell^H,
\end{aligned}$$

where $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K] \in \mathbb{C}^{N_r \times K}$ is the matrix collecting the channel gains for all the K users. In order to determine the linear receiver structure that minimizes the MSE in presence of channel estimation errors, we need to express the MSE as a function of the receiver $\mathbf{G}_\ell$ and the estimated channel $\hat{\mathbf{H}}$, rather than the actual channel $\mathbf{H}$. To this end, we employ (4) for averaging over $\mathbf{h}_k | \hat{\mathbf{h}}_k$ and obtain:

$$\begin{aligned}
&\text{MSE}\left(\mathbf{G}_\ell, \hat{\mathbf{H}}\right) = \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} \left\{ \text{MSE}\left(\mathbf{G}_\ell, \mathbf{H}\right) \right\} \\
&= 1 - \alpha_\ell \sqrt{P_\ell} \mathbf{G}_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell - \alpha_\ell \sqrt{P_\ell} \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \mathbf{G}_\ell^H \\
&\quad + \mathbf{G}_\ell \left(\sum_{k=1}^{K} \alpha_k^2 P_k \left(\mathbf{D}_k \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H \mathbf{D}_k^H + \mathbf{Q}_k \right) + \sigma_d^2 \mathbf{I}_{N_r} \right) \mathbf{G}_\ell^H.
\end{aligned} \tag{6}$$

We can now state our first result, that establishes the true MMSE receiver that uses the instantaneous channel estimates of all users:

Proposition 1. *In the presence of channel estimation errors, the optimal $\mathbf{G}_\ell^\star$ can be derived as*

$$\mathbf{G}_\ell^\star = \alpha_\ell \sqrt{P_\ell} \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \mathbf{J}^{-1}, \tag{7}$$

where

$$\mathbf{J} \triangleq \sum_{k=1}^{K} \alpha_k^2 P_k \left(\mathbf{D}_k \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H \mathbf{D}_k^H + \mathbf{Q}_k \right) + \sigma_d^2 \mathbf{I}_{N_r}. \tag{8}$$

Proof. The proof is in Appendix I. ∎

IV. Average SINR of the Received Data Symbols

When a generic linear filter $\mathbf{G}_\ell$ is employed at the receiver, the estimated symbol of user ℓ is

$$\hat{x}_\ell = \mathbf{G}_\ell \mathbf{y}. \tag{9}$$

The energy of $\hat{x}_\ell$, averaged over $x, \mathbf{n}_d$ and $\mathbf{H}|\hat{\mathbf{H}}$, which are independent and zero mean, can be computed as

$$\begin{aligned}
\mathbb{E}_{x, \mathbf{n}_d, \mathbf{H}|\hat{\mathbf{H}}} \left\{ |\hat{x}_\ell|^2 \right\} =& \alpha_\ell^2 P_\ell |\mathbf{G}_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell|^2 + \sum_{k \neq \ell}^{K} \alpha_k^2 P_k |\mathbf{G}_\ell \mathbf{D}_k \hat{\mathbf{h}}_k|^2 \\
&+ \underbrace{\sum_{k=1}^{K} \alpha_k^2 P_k \mathbf{G}_\ell \mathbf{Q}_k \mathbf{G}_\ell^H + \sigma_d^2 \mathbf{G}_\ell \mathbf{G}_\ell^H}_{\text{channel estimation noise and thermal noise}}.
\end{aligned}$$

In order to determine the average SINR in the sequel, the following lemma will turn out to be useful.

Lemma 2. *When the receiver uses* MMSE *symbol estimation and the instantaneous channel estimates, $\gamma(\mathbf{G}_\ell^\star, \hat{\mathbf{H}})$ the instantaneous SINR of the data symbols, can be expressed as*

$$\gamma(\mathbf{G}_\ell^\star, \hat{\mathbf{H}}) = \alpha_\ell^2 P_\ell \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H {\mathbf{J}_\ell}^{-1} \mathbf{D}_\ell \hat{\mathbf{h}}_\ell, \tag{10}$$

where

$$\mathbf{J}_\ell \triangleq \mathbf{J} - \alpha_\ell^2 P_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H. \tag{11}$$

Proof. The proof is in Appendix II. ∎

A. Computing the Average SINR

If the N_r antennas at the BS are sufficiently spaced apart, we can assume that the generic correlation matrix $\mathbf{C}_k$ is diagonal, i.e. $\mathbf{C}_k = \mathbf{I}_{N_r}$, and, as a consequence $\mathbf{D}_k = d_k \mathbf{I}_{N_r}, \mathbf{R}_k =$

$r_k \mathbf{I}_{N_r}$ and $\mathbf{Q}_k = q_k \mathbf{I}_{N_r}$ are diagonal as well, with $d_k = 1/r_k$, $r_k = 1 + \frac{\sigma_p{}^2}{\alpha_k^2 P_p \tau_p}$ and $q_k = 1 - d_k$ [10].

By defining $\mathbf{v}_k \triangleq \alpha_k \sqrt{P_k} \mathbf{D}_k \hat{\mathbf{h}}_k$, it is $\mathbf{v}_k \sim \mathcal{CN}(0, \sigma_{v,k}^2 \mathbf{I}_{N_r})$, where $\sigma_{v,k}^2 = \alpha_k^2 P_k d_k$. Replacing $\mathbf{v}_k$ in (10) yields

$$\gamma_\ell(\mathbf{G}_\ell^\star, \hat{\mathbf{H}}) = \mathbf{v}_\ell^H \left(\sum_{k \neq \ell}^{K} \mathbf{v}_k \mathbf{v}_k^H + \beta \mathbf{I}_{N_r} \right)^{-1} \mathbf{v}_\ell, \tag{12}$$

where $\beta \triangleq \sum_{k=1}^{K} \alpha_k^2 P_k q_k + \sigma_d^2$. The average SINR for user ℓ is computed as

$$\bar{\gamma}_\ell = \mathbb{E}_{\mathbf{v}_k, k=1\ldots K} \left\{ \mathbf{v}_\ell^H \left(\sum_{k=1, k \neq \ell}^{K} \mathbf{v}_k \mathbf{v}_k^H + \beta \mathbf{I}_{N_r} \right)^{-1} \mathbf{v}_\ell \right\}, \tag{13}$$

To solve (13), we first introduce the normal Hermitian matrix $\mathbf{Y}_\ell \triangleq \sum_{k \neq \ell}^{K} \mathbf{v}_k \mathbf{v}_k^H$, which can be spectrally decomposed as $\mathbf{Y}_\ell = \mathbf{U}_\ell^H \mathbf{\Lambda}_\ell \mathbf{U}_\ell$, and then define $\mathbf{y}_\ell \triangleq \mathbf{U}_\ell^H \mathbf{v}_\ell$. Accordingly, (13) becomes

$$\begin{aligned} \bar{\gamma}_\ell &= \mathbb{E}_{\mathbf{y}_\ell, \lambda_i, i=1\ldots N_r} \left\{ \mathbf{y}_\ell^H \mathbf{U}_\ell \mathbf{U}_\ell^H (\mathbf{\Lambda}_\ell + \beta \mathbf{I}_{N_r})^{-1} \mathbf{U}_\ell \mathbf{U}_\ell^H \mathbf{y}_\ell \right\} \\ &= \mathbb{E}_{\mathbf{y}_\ell, \lambda_i, i=1\ldots N_r} \left\{ \sum_{i=1}^{N_r} \frac{|\mathbf{y}_{\ell i}|^2}{\lambda_i + \beta} \right\}, \end{aligned}$$

where $\mathbf{y}_{\ell i}$ is ith element of the vector $\mathbf{y}_\ell$ and λ_i is the ith eigenvalue of $\mathbf{Y}_\ell$.

Due to the fact that $\mathbf{U}$ is unitary, $\mathbf{y}_\ell$ has the same distribution as $\mathbf{v}_\ell$, so that is $\mathbf{y}_\ell \sim \mathcal{CN}(0, \sigma_{v,k}^2 \mathbf{I}_{N_r})$ and $\mathbb{E}_{\mathbf{y}_\ell} \left\{ |\mathbf{y}_{\ell i}|^2 \right\} = \sigma_{v,k}^2$. Moreover, since the interference matrix $\mathbf{Y}_\ell$ is independent of $\hat{\mathbf{h}}_\ell$, $\mathbf{y}_\ell$ is independent of the eigenvalues λ_i and hence

$$\bar{\gamma}_\ell = \alpha_\ell^2 P_\ell d_\ell \cdot \mathbb{E}_{\lambda_i, i=1\ldots N_r} \left(\sum_{i=1}^{N_r} \frac{1}{\lambda_i + \beta} \right). \tag{14}$$

To manage the complexity of this approach, we utilize results from random matrix theory (RMT) and, in particular, we make the assumption that $N_r, K \to \infty$, with K/N_r fixed. We can now state our main result regarding the asymptotic average SINR of the MU-MIMO system as a function of the transmit power when a total power budget per user is imposed:

Theorem 3. *The asymptotic average SINR of the tagged user-ℓ, denoted by $\bar{\gamma}_\ell$, satisfies the following equation:*

$$\sum_{k=1}^{K} \alpha_k^2 P_k q_k + \sigma_d^2 = \frac{N_r \alpha_\ell^2 P_\ell d_\ell}{\bar{\gamma}_\ell} - \sum_{k=1, k \neq \ell}^{K} \frac{\alpha_k^2 P_k d_k}{1 + \frac{\bar{\gamma}_\ell \alpha_k^2 P_k d_k}{\alpha_\ell^2 P_\ell d_\ell}}. \tag{15}$$

Proof. The proof is in Appendix III. ∎

V. Numerical Results and Concluding Remarks

In this section, we consider a single cell of a MU-MIMO system with $N_r = 16$ antennas, in which the BS serves $K = 10, 13$ and 16 mobile stations, with a power budget of $P_{tot} = 250$ mW each.

We set $F = 12$ subcarriers and $\tau_p = 2$, i.e., the resources allocated to pilot transmission are approximately 16% of the coherence budget. Note that in order to create a higher number of orthogonal pilot sequences, multiple slots within the coherence time of the channel may be used for channel estimation [6]. We assume the same noise power $\sigma^2 = \sigma_p^2 = \sigma_d^2$ for pilot and data symbols, and the same path loss for all users, with a maximum available $\text{SNR}_{max} = \alpha^2 P_{tot}/\sigma^2 = 20$ dB, computed as all the available power were transmitted over a single subcarrier.

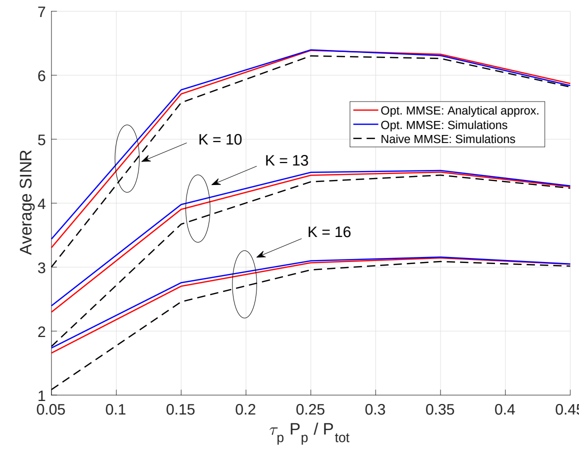

Figure 1: Average SINR versus the normalized pilot power when using the naïve and the MMSE receivers.

Fig. 1 compares the average SINR of the proposed receiver, obtained via simulations and solving (15), with the SINR obtained for the naïve receiver. The naïve receiver can be derived from (7) when $\mathbf{D}_k = \mathbf{I}_{N_r}$ and $\mathbf{Q}_k = \mathbf{0}$. Notice that the naïve receiver can be seen as the linear receiver counterpart of the mismatched receiver studied in [3]. The curves show that the analytical approximation (15) fits very well the simulation results, and that the proposed MMSE receiver that takes into consideration estimation errors outperforms the naïve receiver.

Acknowledgments

We would like to thank the Editor and the Reviewers for their insightful comments. G. Fodor acknowledges that part of this work has been performed in the framework of the H2020 project 5GCAR cofounded by the EU.

Appendix I: Proof of Proposition 1

Notice that (6) can be written in the quadratic form of:

$$\begin{aligned} &\text{MSE}\left(\mathbf{G}_\ell, \hat{\mathbf{H}}\right) = \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} \text{MSE}\left(\mathbf{G}_\ell, \mathbf{H}\right) \\ &= 1 - \underbrace{\mathbf{G}_\ell}_{\mathbf{x}} \underbrace{\alpha_\ell \sqrt{P_\ell} \mathbf{D}_\ell \hat{\mathbf{h}}_\ell}_{\mathbf{B}} - \underbrace{\alpha_\ell \sqrt{P_\ell} \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H}_{\mathbf{B}^H} \underbrace{\mathbf{G}_\ell^H}_{\mathbf{x}^H} \\ &\quad + \mathbf{G}_\ell \underbrace{\left(\sum_{k=1}^{K} \alpha_k^2 P_k \left(\mathbf{D}_k \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H \mathbf{D}_k^H + \mathbf{Q}_k \right) + \sigma_d^2 \mathbf{I}_{N_r} \right)}_{\triangleq \mathbf{J}} \mathbf{G}_\ell^H, \end{aligned}$$

whose solution is $\mathbf{x}^\star = \mathbf{B}^H \mathbf{J}^{-1}$, which gives $\mathbf{G}^\star$.

Appendix II: Proof of Lemma 2

From (10), the SINR is expressed as:

$$\begin{aligned} &\gamma(\mathbf{G}_\ell, \hat{\mathbf{H}}) = \\ &\frac{\alpha_\ell^2 P_\ell |\mathbf{G}_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell|^2}{\sum_{k \neq \ell} \alpha_k^2 P_k |\mathbf{G}_\ell \mathbf{D}_k \hat{\mathbf{h}}_k|^2 + \sum_k \alpha_k^2 P_k \mathbf{G}_\ell \mathbf{Q}_k \mathbf{G}_\ell^H + \sigma_d^2 \mathbf{G}_\ell \mathbf{G}_\ell^H} \\ &= \frac{\mathbf{G}_\ell (\mathbf{J} - \mathbf{J}_\ell) \mathbf{G}_\ell^H}{\mathbf{G}_\ell \mathbf{J}_\ell \mathbf{G}_\ell^H}. \end{aligned}$$

Substituting $\mathbf{G}_\ell^\star = \alpha_\ell\sqrt{P_\ell}\hat{\mathbf{h}}_\ell^H\mathbf{D}_\ell^H\mathbf{J}^{-1}$, we obtain:

$$\gamma(\mathbf{G}_\ell^\star,\hat{\mathbf{H}}) = \frac{\hat{\mathbf{h}}_\ell^H\mathbf{D}_\ell^H\mathbf{J}^{-1}(\mathbf{J}-\mathbf{J}_\ell)\mathbf{J}^{-1}\mathbf{D}_\ell\hat{\mathbf{h}}_\ell}{\hat{\mathbf{h}}_\ell^H\mathbf{D}_\ell^H\mathbf{J}^{-1}\mathbf{J}_\ell\mathbf{J}^{-1}\mathbf{D}_\ell\hat{\mathbf{h}}_\ell}$$

$$= \frac{\alpha_\ell^2 P_\ell\hat{\mathbf{h}}_\ell^H\mathbf{D}_\ell^H\mathbf{J}^{-1}\mathbf{D}_\ell\hat{\mathbf{h}}_\ell\hat{\mathbf{h}}_\ell^H\mathbf{D}_\ell^H\mathbf{J}^{-1}\mathbf{D}_\ell\hat{\mathbf{h}}_\ell}{\hat{\mathbf{h}}_\ell^H\mathbf{D}_\ell^H\mathbf{J}^{-1}\left(\mathbf{J}-\alpha_\ell^2P_\ell\mathbf{D}_\ell\hat{\mathbf{h}}_\ell\hat{\mathbf{h}}_\ell^H\mathbf{D}_\ell^H\right)\mathbf{J}^{-1}\mathbf{D}_\ell\hat{\mathbf{h}}_\ell}$$

$$= \frac{\alpha_\ell^2P_\ell\hat{\mathbf{h}}_\ell^H\mathbf{D}_\ell^H\mathbf{J}^{-1}\mathbf{D}_\ell\hat{\mathbf{h}}_\ell}{1-\alpha_\ell^2P_\ell\hat{\mathbf{h}}_\ell^H\mathbf{D}_\ell^H\mathbf{J}^{-1}\mathbf{D}_\ell\hat{\mathbf{h}}_\ell}. \tag{16}$$

Applying the Woodbury matrix identity gives the lemma.

Appendix III: Proof of Theorem 3

From equations (13) and (14) of [12] we have that:

$$\mathbb{E}_{\lambda_i, i=1\ldots N_r}\left\{\sum_{i=1}^{N_r}\frac{1}{\lambda_i+\beta}\right\} = N_r\mathbb{E}_\lambda\left\{\frac{1}{\lambda+\beta}\right\}, \tag{17}$$

where λ is a random eigenvalue (spectrum) of $\mathbf{Y}_\ell$. According to [13], to compute $\mathbb{E}_\lambda\left\{\frac{1}{\lambda+\beta}\right\}$, the Stieltjes transform according to:

$$G(s) \triangleq \int_x \frac{1}{x-s}dP(x) \tag{18}$$

can be used, since the Stieltjes transform of the distribution of λ at $s=-\beta$ is

$$G(-\beta) = \int_\lambda \frac{1}{\lambda+\beta}dP(\lambda) = \mathbb{E}_\lambda\left\{\frac{1}{\lambda+\beta}\right\} = \frac{\bar{\gamma}}{N_r\alpha_\ell^2P_\ell d_\ell} \tag{19}$$

For matrices like $\mathbf{Y}_\ell = \sum_{k\neq\ell}^K \mathbf{v}_k\mathbf{v}_k^H$ with $\mathbf{v}_k \sim \mathcal{CN}(0,\sigma_{k,v}^2\mathbf{I}_{N_r})$, the relationship between the R-transform and the G-transform [13] can be expressed as:

$$G\left(R(-w)-\frac{1}{w}\right) = w \tag{20}$$

In addition, under the assumption $N_r\to\infty$, we have that the family of matrices $\mathbf{v}_k\mathbf{v}_k^H$ is almost surely asymptotically free [13]. Accordingly, we can exploit the important result that the R-transform of the sum of random matrices belonging to a set of a free family is given by the sum of their individual R-transforms, where the R-transform of $\mathbf{v}_k\mathbf{v}_k^H$ is $R_k(w) = \frac{\alpha_k^2P_kd_k}{1-N_r\sigma_{k,v}^2w}$ [13]. Hence, we get:

$$R(w) = R_{\mathbf{Y}_\ell}(w) = \sum_{k=1,k\neq\ell}^K R_k(w) = \sum_{k=1,k\neq\ell}^K\frac{\alpha_k^2P_kd_k}{1-N_r\sigma_{k,v}^2w}. \tag{21}$$

Substituting (21) into (20) we have

$$G\left(\sum_{k=1,k\neq\ell}^K\frac{\alpha_k^2P_kd_k}{1+N_r\sigma_{k,v}^2w}-\frac{1}{w}\right) = w. \tag{22}$$

Comparing the RHS of (19) and (22) we have $w = \frac{\bar{\gamma}}{N_r\alpha_\ell^2P_\ell d_\ell}$, from which $\bar{\gamma}$ is the solution of the equation:

$$\beta\Bigg|_{\beta=\sum_{k=1}^K\alpha_k^2P_kq_k+\sigma_d^2} = \frac{1}{w}-\sum_{k=1,k\neq\ell}^K\frac{\alpha_k^2P_kd_k}{1+N_r\sigma_{k,v}^2w}\Bigg|_{w=\frac{\gamma}{N_r\alpha_\ell^2P_\ell d_\ell}},$$

which is equivalent with (15).

References

[1] Y. Chen and N. C.Beaulieu, "Optimum pilot symbol assisted modulation," *IEEE Trans. on Comm.*, vol. 55, no. 8, pp. 1536–1546, August 2007.

[2] Y. V. Zakharov, V. M. Baronkin, and J. Zhang, "Optimal and mismatched detection of QAM signals in fast fading channels with imperfect channel estimation," *IEEE Trans. on Wireless Comm.*, vol. 8, no. 2, pp. 617–621, February 2009.

[3] G. Taricco and E. Biglieri, "Space time decoding with imperfect channel estimation," *IEEE Trans. on Wireless Comm.*, vol. 4, no. 4, pp. 1874–1888, July 2005.

[4] G. Taricco, "Optimum receiver design and performance analysis of arbitrarily correlated Rician fading MIMO channels with imperfect channel state information," *IEEE Trans. Info. Theory*, vol. 56, no. 3, p. 2010, March 2010.

[5] S. M.-S. Sadough, M.-A. Khalighi, and P. Duhamel, "Improved iterative MIMO signal detection accounting for channel-estimation errors," *IEEE Trans. on Vehicular Techn.*, vol. 58, no. 7, pp. 3154–3167, September 2009.

[6] T. Marzetta, "How much training is needed for multiuser MIMO ?" *IEEE Asilomar Conf. on Signals, Systems and Computers (ACSSC)* , pp. 359–363, June 2006.

[7] G. Fodor, P. D. Marco, and M. Telek, "On the Impact of Antenna Correlation and CSI Errors on the Pilot-to-Data Power Ratio," *IEEE Trans. Comm.*, vol. 64, no. 6, pp. 2622 – 2633, April 2016.

[8] M. R. McKay, I. B. Collings, and A. M. Tulino, "Achievable sum rate of MIMO MMSE receivers: A general analytic framework," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 396 – 410, January 2010.

[9] E. Eraslan, B. Daneshrad, and C.-Y. Lou, "Performance indicator for MIMO MMSE receivers in the presence of channel estimation error," *IEEE W. Comm. Letters*, vol. 2, no. 2, pp. 211–214, April 2013.

[10] G. Fodor, P. D. Marco, and M. Telek, "On Minimizing the MSE in the Presence of Channel Information Errors," *IEEE Commun. Lett.*, vol. 19, no. 9, Sep. 2015.

[11] S. Sesia, I. Toufik, and M. Baker, *Long Term Evolution Systems: From Theory to Practice*. WILEY, 2012.

[12] G. Livan and P. Vivo, "Moments of Wishart-Laguerre and Jacobi ensembles of random matrices: Application to the quantum transport problem in chaotic cavities," *Acta Phys. Pol. B*, vol. 42, no. 5, pp. 1081–1104, 2011.

[13] R. R. Müller, G. Alfano, B. M. Zaidel, and R. de Miguel, "Applications of large random matrices in communications engineering," *arXiV*, vol. [cs.IT], no. 1310:5479, 2013.