This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the ICC 2008 proceedings.

Robust Precoder Design in MISO Downlink Based on Quadratic Channel Estimation[§]

Gan Zheng[‡] Shaodan Ma[*], Kai-Kit Wong[‡], and Tung-Sang Ng[*]
[‡]Adastral Park Research Campus, University College London
Martlesham Heath, IP5 3RE, UK, Email: {g.zheng,k.wong}@adastral.ucl.ac.uk
[*]Department of Electrical and Electronic Engineering, The University of Hong Kong
Pokfulam, Hong Kong, Email: {sdma,tsng}@eee.hku.hk

***Abstract*— In [1], it has been proposed that channel estimates in quadratic form can be obtained at the base station by sending training sequences to the mobiles where the received signals are forwarded back to the base for channel estimation. In this paper, we first examine the optimal training sequence design for such quadratic channel estimation and then analyze the error bound and statistics of the channel estimates in quadratic form. With the analytical results, two problems for a multiple-input single-output (MISO) antenna system in the downlink are constructed and optimally solved: Power minimization with individual users' 1) worst-case signal-to-interference plus noise ratio (SINR) and 2) average mean-square-error (MSE) constraints, through optimal multiuser MISO beamforming and power allocation.**

I. Introduction

Since the works in [2], [3], multiple-input multiple-output (MIMO) antenna has been well understood as an energy-and-spectral efficient solution to wireless communications. Recent studies on the use of MIMO for multiuser channels, e.g., [4]–[6], further reveals its extraordinary performance by allowing users to be shared in the spatial domain with the possible use of channel state information (CSI) at the transmitter.

To realize the benefit, in practice, CSI needs to be acquired from estimation but is never perfectly known, due to reasons such as noisy estimation, Doppler spread, and etc. In a time-division-duplex (TDD) system, the reciprocity of the up and downlink channels permits the CSI estimates in one link to be used for another link. However, this is not true for a frequency-division-duplex (FDD) system in which the two links occupy different frequency bands. For CSI to be available at the base station for downlink optimization, common techniques require the CSI to be estimated at the mobiles in the uplink and fed back to the base station.

Recently in [1], Dong and Ding showed that for the optimization of most metrics of interest, full CSI is not necessary and quadratic CSI will be sufficient. They further presented a downlink quadratic CSI estimation approach, which requires little processing at the mobiles to simply forward the received noisy training signals back to the base station for estimation.

In this paper, we approach the quadratic CSI estimation in [1] by deriving the mean-square-error (MSE) in the estimates and then designing the optimal training sequences minimizing the MSE. Using the optimal quadratic estimation, we analyze the error bound and statistics of the CSI estimates that allow us to formulate the robust beamforming design problems with consideration of the CSI errors. Power minimization problems subject to individual users' 1) signal-to-interference plus noise ratio (SINR) or 2) MSE constraints, in a multiple-input single-output (MISO) antenna system in the downlink with the aid of quadratic CSI estimates, are studied and optimally solved.

In the sequel, we shall use the following notations. Vectors are column vectors and denoted in lower case bold $\mathbf{x}$ while matrices are upper case bold $\mathbf{A}$. The superscripts † and $^\sharp$ stand for, respectively, the conjugate transposition and the Penrose-Moore pseudo-inversion. $\mathtt{tr}(\mathbf{A})$ is the trace of $\mathbf{A}$ and $\mathbf{I}$ is an identity matrix. $\mathtt{vec}(\mathbf{A})$ produces a column vector by stacking the entries of $\mathbf{A}$. The notation $\otimes$ denotes the tensor product. The complex number field is denoted by $\mathbb{C}$. $\mathtt{E}[\cdot]$ represents the expected value operator while $|\cdot|$ takes the modulus of a complex number and $\|\cdot\|$ returns the Frobenius norm of a vector or matrix. $\mathbf{x} \sim \mathcal{CN}(\mathbf{m}, \mathbf{V})$ means that $\mathbf{x}$ is a vector of complex Gaussian random variables and has a mean vector of $\mathbf{m}$ with a covariance matrix of $\mathbf{V}$.

II. System Model

A. MISO Antenna System in Downlink

Consider the mth user of an M-user MISO channel in the downlink with slow-fading

$$\hat{s}_m = \mu_m \mathbf{h}_m^\dagger \left(\sum_{n=1}^{M} \mathbf{t}_n s_n \right) + n_m, \quad \text{for } m = 1, 2, \ldots, M, \quad (1)$$

where

$\mathbf{h}_m^\dagger$	the channel seen at user m ($\mathbb{C}^{1\times n_T}$);
s_m	the symbol transmitted with unit power to user m;
$\mathbf{t}_m$	the precoding vector for user m ($\mathbb{C}^{n_T\times 1}$);
$\hat{s}_m$	the estimated signal for user m;
n_m	the noise with zero mean and variance of N_0;
n_T	the number of transmit antennas at the base;
μ_m	the real-valued scaling for the received signal.

Clearly, the total transmit power is $\sum_{n=1}^{M} \|\mathbf{t}_n\|^2$. For detection purpose, it is required that mobile receiver m has access to

[§]This work was supported in part by The Hong Kong Research Grants Council under grant HKU7175/03E and The University Research Committee of The University of Hong Kong, and in part by The Engineering and Physical Science Research Council under grant EP/D058716/1, United Kingdom.

the knowledge of $\mathbf{h}_m$. However, this is not necessary when the precoding vectors $\{\mathbf{t}_n\}$ are designed at the base station transmitter. In fact, for most metrics of interest, the CSI in quadratic form, i.e., $\{\mathbf{h}_n^\dagger \mathbf{h}_n\}$ will be sufficient. This is evident by observing, for instance, the SINR

$$\begin{aligned}\textsf{SINR at user } m &= \frac{\mathbf{t}_m^\dagger\left(\mathbf{h}_m\mathbf{h}_m^\dagger\right)\mathbf{t}_m}{\sum_{\substack{n=1\\n\neq m}}^M \mathbf{t}_n^\dagger\left(\mathbf{h}_m\mathbf{h}_m^\dagger\right)\mathbf{t}_n + N_0}\\ &= \frac{\mathbf{t}_m^\dagger\mathbf{Q}_m\mathbf{t}_m}{\sum_{\substack{n=1\\n\neq m}}^M \mathbf{t}_n^\dagger\mathbf{Q}_m\mathbf{t}_n + N_0},\end{aligned} \tag{2}$$

where $\{\mathbf{Q}_m \triangleq \mathbf{h}_m\mathbf{h}_m^\dagger \forall m\}$ are the so-called quadratic CSI. Since many other performance metrics depend on the SINR, such as bit-error-rate (BER), throughput and etc, the estimation of quadratic CSI covers a wide range of applications.

B. Downlink Quadratic Channel Estimation

In [1], it was proposed to acquire the quadratic CSI by the following procedure. First of all, the base station transmits a known training matrix $\mathbf{X}$ spanning in n_d symbol durations to a particular mobile user. Upon receiving at the mobile user (with user index omitted for convenience), the received signals

$$\mathbf{r}^\dagger = \mathbf{h}^\dagger\mathbf{X} + \mathbf{n}, \tag{3}$$

will be untouched and directly fed back to the base station in the uplink channel to give

$$\mathbf{Y} = \mathbf{g}\mathbf{r}^\dagger + \mathbf{W} = \mathbf{g}\mathbf{h}\dagger\mathbf{X} + \mathbf{g}\mathbf{n}^\dagger + \mathbf{W} \tag{4}$$

at the base station, where

$\mathbf{X}$	the training data sent in the downlink ($\mathbb{C}^{n_T\times n_d}$);
$\mathbf{Y}$	the signals received in the uplink ($\mathbb{C}^{n_T\times n_d}$);
$\mathbf{r}^\dagger$	the signals received in the downlink ($\mathbb{C}^{1\times n_d}$);
$\mathbf{h}^\dagger$	the downlink channel ($\mathbb{C}^{1\times n_T}$);
$\mathbf{g}$	the corresponding uplink channel ($\mathbb{C}^{n_T\times 1}$);
$\mathbf{n}^\dagger$	the downlink noise vector ($\mathbb{C}^{1\times n_d}$);
$\mathbf{W}$	the uplink noise matrix ($\mathbb{C}^{n_T\times n_d}$).

At the base station, the channel covariance matrix $\mathbf{E} = \mathbf{g}\mathbf{g}^\dagger$ can be easily estimated from the second-order statistics of the noisy receptions of the known sequences transmitted from the mobile station, which we assume to have performed perfectly. A slow-fading scenario is considered such that the channels, $\mathbf{h}$ and $\mathbf{g}$, are static during the channel estimation process. Note that $\mathbf{E}$ is a rank-1 matrix, so $\mathbf{g}$ can be found from

$$\mathbf{g} = e^{j\theta}\sqrt{\textsf{tr}(\mathbf{E})}\mathbf{e} \tag{5}$$

where $\mathbf{e}$ is the eigenvector of $\mathbf{E}$ and $\theta \in (0, 2\pi)$ takes into account the possible phase ambiguity between $\mathbf{g}$ and $\mathbf{e}$.

Define the matrix $\mathbf{F} \triangleq \mathbf{g}\mathbf{h} \in \mathbb{C}^{n_T\times n_T}$, which is crucial in estimating the quadratic CSI. In the absence of noise, $\mathbf{F}$ can be obtained by

$$\mathbf{F} = \mathbf{Y}\mathbf{X}^\sharp = \mathbf{g}\mathbf{h}^\dagger, \tag{6}$$

in which $\mathbf{X}\mathbf{X}^\sharp = \mathbf{I}$. Furthermore, the downlink quadratic CSI can be computed by

$$\mathbf{F}^\dagger\mathbf{E}^\sharp\mathbf{F} = \mathbf{h}\mathbf{g}^\dagger(\mathbf{g}\mathbf{g}^\dagger)^\sharp\mathbf{g}\mathbf{h}^\dagger = \mathbf{h}\mathbf{h}^\dagger. \tag{7}$$

Since $\mathbf{h}\mathbf{h}^\dagger$ is of rank one, the downlink channel vector can be further found as

$$\mathbf{h} = \frac{e^{j\phi}}{\textsf{tr}(\mathbf{E})}\mathbf{F}^\dagger\mathbf{e}, \tag{8}$$

where ϕ denotes the possible phase ambiguity. Consequently, estimation of $\mathbf{h}\mathbf{h}^\dagger$, or $\mathbf{h}$, reduces to knowing $\mathbf{F}$.

This CSI estimation algorithm, originally proposed in [1], is attractive because most of the processing and calculations are performed at the base station, where the CSI estimates are used to optimize the transmission in the downlink. Nonetheless, little is known about the MSE in the CSI estimates, which will have an implication on the downlink design.

In practice, noise is inevitable, which means that the estimates contain noise. For instance,

$$\begin{aligned}\hat{\mathbf{F}} &= \mathbf{Y}\mathbf{X}^\sharp\\ &= \mathbf{g}\mathbf{h}^\dagger + (\mathbf{g}\mathbf{n} + \mathbf{W})\mathbf{X}^\sharp\\ &\equiv \mathbf{F} \ + \ \Delta\mathbf{F}.\end{aligned} \tag{9}$$

The noise component, $\Delta\mathbf{F}$, will further affect the estimation of the quadratic CSI to give

$$\hat{\mathbf{F}}^\dagger\mathbf{E}^\sharp\hat{\mathbf{F}} = \mathbf{h}^\dagger\mathbf{h} + \underbrace{\hat{\mathbf{F}}^\dagger\mathbf{E}^\sharp\Delta\mathbf{F} + \Delta\mathbf{F}^\dagger\mathbf{E}^\sharp\hat{\mathbf{F}} + \Delta\mathbf{F}^\dagger\mathbf{E}^\sharp\Delta\mathbf{F}}_{\text{noise}}, \tag{10}$$

and

$$\frac{e^{j\phi}}{\textsf{tr}(\mathbf{E})}\hat{\mathbf{F}}^\dagger\mathbf{e} = \mathbf{h} + \underbrace{\frac{e^{j\phi}}{\textsf{tr}(\mathbf{E})}\Delta\mathbf{F}^\dagger\mathbf{e}}_{\text{noise}}. \tag{11}$$

The rest of the paper will be devoted to analyze the MSE of the quadratic CSI estimation method, i.e.,

$$\textsf{MSE} = \textsf{E}[\|\mathbf{h}\mathbf{h}^\dagger - \hat{\mathbf{F}}^\dagger\mathbf{E}^\sharp\hat{\mathbf{F}}\|^2], \tag{12}$$

which permits us to develop the optimal training sequences $\mathbf{X}$ for minimal MSE (see Section III), and later allows robust schemes to be devised based on the quadratic CSI estimates (see Sections IV & V).

III. Optimal Training Data Design

A. Derivation of MSE

Here, we derive the MSE (12) of the quadratic CSI estimates, which can be first done by

$$\begin{aligned}\textsf{MSE} =& \textsf{E}[\|\Delta\mathbf{F}^\dagger\mathbf{E}^\sharp\Delta\mathbf{F} + \Delta\mathbf{F}^\dagger\mathbf{E}^\sharp\mathbf{F} + \mathbf{F}^\dagger\mathbf{E}^\sharp\Delta\mathbf{F}\|^2]\\ =& \textsf{E}[\|\Delta\mathbf{F}^\dagger\mathbf{E}^\sharp\Delta\mathbf{F}\|^2] + 2\textsf{E}[\|\Delta\mathbf{F}^\dagger\mathbf{E}^\sharp\mathbf{F}\|^2]\\ =& \textsf{E}\left[\textsf{tr}\left(\Delta\mathbf{F}^\dagger\mathbf{E}^\sharp\Delta\mathbf{F}\Delta\mathbf{F}^\dagger\mathbf{E}^\sharp\Delta\mathbf{F}\right)\right]\\ &+ 2\textsf{E}\left[\textsf{tr}\left(\Delta\mathbf{F}^\dagger\mathbf{E}^\sharp\mathbf{F}\mathbf{F}^\dagger\mathbf{E}^\sharp\Delta\mathbf{F}\right)\right].\end{aligned} \tag{13}$$

Letting $\tilde{\mathbf{X}} = \mathbf{X}^\sharp \mathbf{X}^{\sharp\dagger}$, then the first term can be evaluated as

$$\begin{aligned}\mathsf{E}\left[\mathsf{tr}\left(\Delta\mathbf{F}^\dagger\mathbf{E}^\sharp\Delta\mathbf{F}\Delta\mathbf{F}^\dagger\mathbf{E}^\sharp\Delta\mathbf{F}\right)\right] = \\ \mathsf{E}\Big[\mathsf{tr}\Big((\mathbf{n}\mathbf{g}^\dagger + \mathbf{W}^\dagger)\mathbf{E}^\sharp(\mathbf{g}\mathbf{n}^\dagger + \mathbf{W})\tilde{\mathbf{X}} \\ (\mathbf{n}\mathbf{g}^\dagger + \mathbf{W}^\dagger)\mathbf{E}^\sharp(\mathbf{g}\mathbf{n}^\dagger + \mathbf{W})\tilde{\mathbf{X}}\Big)\Big], \quad (14)\end{aligned}$$

which can further be expanded as

$$\mathsf{tr}\begin{pmatrix} \mathsf{E}\left[\left(\mathbf{n}\mathbf{g}^\dagger\mathbf{E}^\sharp\mathbf{g}\mathbf{n}^\dagger\tilde{\mathbf{X}}\mathbf{n}\mathbf{g}^\dagger\mathbf{E}^\sharp\mathbf{g}\mathbf{n}^\dagger\tilde{\mathbf{X}}\right)\right] \\ +\mathsf{E}\left[\left(\mathbf{n}\mathbf{g}^\dagger\mathbf{E}^\sharp\mathbf{g}\mathbf{n}^\dagger\tilde{\mathbf{X}}\mathbf{W}^\dagger\mathbf{E}^\sharp\mathbf{W}\tilde{\mathbf{X}}\right)\right] \\ +\mathsf{E}\left[\left(\mathbf{W}^\dagger\mathbf{E}^\sharp\mathbf{W}\tilde{\mathbf{X}}\mathbf{n}\mathbf{g}^\dagger\mathbf{E}^\sharp\mathbf{g}\mathbf{n}^\dagger\tilde{\mathbf{X}}\right)\right] \\ +\mathsf{E}\left[\left(\mathbf{W}^\dagger\mathbf{E}^\sharp\mathbf{W}\tilde{\mathbf{X}}\mathbf{W}^\dagger\mathbf{E}^\sharp\mathbf{W}\tilde{\mathbf{X}}\right)\right] \\ +\mathsf{E}\left[\left(\mathbf{n}\mathbf{g}^\dagger\mathbf{E}^\sharp\mathbf{W}\tilde{\mathbf{X}}\mathbf{W}^\dagger\mathbf{E}\mathbf{g}\mathbf{n}^\dagger\tilde{\mathbf{X}}\right)\right] \\ +\mathsf{E}\left[\left(\mathbf{W}^\dagger\mathbf{E}\mathbf{g}\mathbf{n}^\dagger\tilde{\mathbf{X}}\mathbf{n}\mathbf{g}^\dagger\mathbf{E}^\sharp\mathbf{W}\tilde{\mathbf{X}}\right)\right] \end{pmatrix}$$

$$\begin{aligned}= \mathsf{tr}\Big(\sigma_n^4\left[\tilde{\mathbf{X}} + \mathsf{tr}(\tilde{\mathbf{X}})\mathbf{I}\right]\tilde{\mathbf{X}} + \frac{\sigma_w^4}{\mathsf{tr}(\mathbf{E})^2}\left[\tilde{\mathbf{X}} + \mathsf{tr}(\tilde{\mathbf{X}})\mathbf{I}\right]\tilde{\mathbf{X}} \\ + \frac{2\sigma_n^2\sigma_w^2}{\mathsf{tr}(\mathbf{E})}\left[\tilde{\mathbf{X}} + \mathsf{tr}(\tilde{\mathbf{X}})\mathbf{I}\right]\tilde{\mathbf{X}}\Big). \quad (15)\end{aligned}$$

Finally, we can have the first term equal

$$\begin{aligned}\mathsf{E}\left[\mathsf{tr}\left(\Delta\mathbf{F}^\dagger\mathbf{E}^\sharp\Delta\mathbf{F}\Delta\mathbf{F}^\dagger\mathbf{E}^\sharp\Delta\mathbf{F}\right)\right] \\ = \alpha\left(\mathsf{tr}(\tilde{\mathbf{X}})^2 + \mathsf{tr}(\tilde{\mathbf{X}}^2)\right), \quad (16)\end{aligned}$$

where $\alpha = \sigma_n^4 + \frac{\sigma_w^4}{\mathsf{tr}(\mathbf{E})^2} + \frac{2\sigma_n^2\sigma_w^2}{\mathsf{tr}(\mathbf{E})}$. On the other hand, the second term of (13) can be expressed as

$$\begin{aligned}&\mathsf{E}\left[\mathsf{tr}\left(\Delta\mathbf{F}^\dagger\mathbf{E}^\sharp\mathbf{F}\mathbf{F}^\dagger\mathbf{E}^\sharp\Delta\mathbf{F}\right)\right] \\ &= \mathsf{tr}\left[\left(\sigma_n^2\mathbf{g}^\dagger\mathbf{E}^\sharp\mathbf{F}\mathbf{F}^\dagger\mathbf{E}^\sharp\mathbf{g} + \sigma_w^2\mathsf{tr}(\mathbf{E}^\sharp\mathbf{F}\mathbf{F}^\dagger\mathbf{E}^\sharp)\right)\tilde{\mathbf{X}}\right] \\ &= \beta\mathsf{tr}(\tilde{\mathbf{X}}),\end{aligned} \quad (17)$$

where $\beta = \sigma_n^2\mathsf{tr}(\mathbf{E})\mathbf{e}^\dagger\mathbf{E}^\sharp\mathbf{F}\mathbf{F}^\dagger\mathbf{E}^\sharp\mathbf{e} + \sigma_w^2\mathsf{tr}(\mathbf{E}^\sharp\mathbf{F}\mathbf{F}^\dagger\mathbf{E}^\sharp)$. As a result,

$$\mathsf{MSE} = \alpha\left(\mathsf{tr}(\tilde{\mathbf{X}})^2 + \mathsf{tr}(\tilde{\mathbf{X}}^2)\right) + 2\beta\mathsf{tr}(\tilde{\mathbf{X}}). \quad (18)$$

Let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_{n_T})$ denote the eigenvalues of $\mathbf{X}\mathbf{X}^\dagger$. Then, the eigenvalues of $\tilde{\mathbf{X}}$ would be $(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \ldots, \frac{1}{\lambda_{n_T}})$ and (18) becomes

$$\mathsf{MSE}(\boldsymbol{\lambda}) = \alpha\left[\sum_{n=1}^{n_T}\frac{1}{\lambda_n^2} + \left(\sum_{n=1}^{n_T}\frac{1}{\lambda_n}\right)^2\right] + 2\beta\sum_{n=1}^{n_T}\frac{1}{\lambda_n}. \quad (19)$$

B. Minimization of MSE

As to the training signal design, the problem is to minimize the MSE with a given total training power, i.e.,

$$\min_{\boldsymbol{\lambda} > \mathbf{0}} \mathsf{MSE}(\boldsymbol{\lambda}) \quad \text{s.t.} \quad \sum_{n=1}^{n_T}\lambda_n \le n_T n_d. \quad (20)$$

Note that since $\mathsf{MSE}(\boldsymbol{\lambda})$ is a Schur-Convex function (a sum of convex function), and $\boldsymbol{\lambda}^* = (\frac{1}{n_T}, \frac{1}{n_T}, \cdots, \frac{1}{n_T}) \preceq \boldsymbol{\lambda}$ for any $\boldsymbol{\lambda} > \mathbf{0}$ and $\sum_{i=1}^{n_T}\lambda_i = n_T n_d$, we have

$$\mathsf{MSE}(\boldsymbol{\lambda}^*) \le \mathsf{MSE}(\boldsymbol{\lambda}) \ \forall\boldsymbol{\lambda} \quad (21)$$

and $\boldsymbol{\lambda}^*$ is therefore the optimal solution to (20). For details, we refer the readers to the majorization theory in [7]. This has also led to the condition for the optimal training sequence $\mathbf{X}$

$$\mathbf{X}\mathbf{X}^\dagger = n_d\mathbf{I}. \quad (22)$$

IV. Error Bound Analysis and a Robust Design Based on Worst-Case Constraints

A. Bounding $\|\Delta\mathbf{F}\|$

Based on the quadratic CSI estimates in Section III, we here proceed to develop a robust-optimal multiuser MISO system where the users' worst-case SINRs are ensured. To do so, we assume that the noise vector and matrix are bounded by

$$\|\mathbf{n}\| \le \rho_n \quad \text{and} \quad \|\mathbf{W}\| \le \rho_w \quad (23)$$

and they are uniformly distributed in the regions. The bound for $\|\Delta\mathbf{F}\|$ is important in the CSI quality and we derive this by first noting that $\|\Delta\mathbf{F}\|$ is maximum when

$$\mathbf{W} = \frac{\rho_w\mathbf{g}\mathbf{n}}{\sqrt{\mathsf{tr}(\mathbf{g}\mathbf{n}^\dagger\mathbf{n}\mathbf{g}^\dagger)}}. \quad (24)$$

On the other hand, we can obtain a bound for

$$\begin{aligned}\|\mathbf{g}\mathbf{n}\mathbf{X}^\sharp\|^2 &= \mathsf{tr}(\mathbf{g}\mathbf{n}^\dagger\tilde{\mathbf{X}}\mathbf{n}\mathbf{g}^\dagger) \\ &= \mathsf{tr}(\mathbf{n}^\dagger\tilde{\mathbf{X}}\mathbf{n}\mathbf{g}^\dagger\mathbf{g}) \\ &= \mathsf{tr}(\mathbf{E})\mathbf{n}^\dagger\tilde{\mathbf{X}}\mathbf{n} \le \mathsf{tr}(\mathbf{E})\rho_n^2\lambda_{\tilde{\mathbf{X}}},\end{aligned} \quad (25)$$

where $\lambda_{\tilde{\mathbf{X}}}$ is the maximal eigenvalue of $\tilde{\mathbf{X}}$ and this bound is achievable by having $\mathbf{n} = \rho_n\mathbf{v}$ in which $\mathbf{v}$ is the associated eigenvector. As a result, $\mathbf{w}$ can be written as

$$\mathbf{W} = \frac{\rho_w\mathbf{e}\mathbf{v}^\dagger}{\sqrt{\mathsf{tr}(\mathbf{e}\mathbf{v}^\dagger\mathbf{v}\mathbf{e}^\dagger)}}. \quad (26)$$

An upper bound for $\Delta\mathbf{F}$ can then be derived as

$$\begin{aligned}&\|\Delta\mathbf{F}\|^2 \le \\ &\left(1 + \frac{\rho_w}{\rho_n\sqrt{\mathsf{tr}(\mathbf{E})\mathsf{tr}\left(\mathbf{e}\mathbf{v}^\dagger\mathbf{v}\mathbf{e}^\dagger\right)}}\right)^2\mathsf{tr}(\mathbf{E})\rho_n^2\lambda_{\tilde{\mathbf{X}}} \triangleq \rho_{\Delta\mathbf{F}}^2. \quad (27)\end{aligned}$$

B. Robust Design with Worst-Case SINR Constraints

Given the error bound in $\Delta\mathbf{F}$, it becomes possible to design a robust multiuser MISO system. Mathematically, this can be achieved by including the worse-case SINR constraints of the users, i.e.,

$$\begin{aligned}&\min_{\{\mathbf{t}_m\}_{m=1}^M}\sum_{m=1}^M\|\mathbf{t}_m\|^2 \\ &\text{s.t.} \ \min_{\Delta\mathbf{F}_m}\frac{\mathsf{tr}(\mathbf{Q}_m\mathbf{t}_m\mathbf{t}_m^\dagger)}{\sum_{\substack{n=1\\n\neq m}}\mathsf{tr}(\mathbf{Q}_m\mathbf{t}_n\mathbf{t}_n^\dagger) + N_0} \ge \gamma_m, \\ &\|\Delta\mathbf{F}_m\| \le \rho_{\Delta\mathbf{F}_m} \quad \forall m,\end{aligned} \quad (28)$$

where $\Delta\mathbf{F}_m$ and $\rho_{\Delta\mathbf{F}_m}$ follow the same definitions as defined before, but with the user index m. This problem is challenging because it requires to evaluate $\min\mathsf{SINR}(\Delta\mathbf{F}_m)$ which is a complicated function of $\Delta\mathbf{F}_m$. We show it is possible to solve

(28) optimally by convex optimization with relaxation. This is done by first rewriting the constraint as

$$\begin{aligned}\mathsf{vec}(\Delta\mathbf{F}_m^\dagger)^\dagger\mathsf{vec}(\tilde{\mathbf{T}}_m\mathbf{F}_m^\dagger\mathbf{E}_m^\sharp)+\mathsf{vec}(\tilde{\mathbf{T}}_m\mathbf{F}_m^\dagger\mathbf{E}_m^\sharp)^\dagger\mathsf{vec}(\Delta\mathbf{F}_m^\dagger)+\\ \mathsf{vec}(\Delta\mathbf{F}_m^\dagger)^\dagger\mathsf{vec}(\mathbf{E}_m^\sharp\otimes\tilde{\mathbf{T}}_m)\mathsf{vec}(\Delta\mathbf{F}_m^\dagger)+\\ \mathsf{tr}\left(\mathbf{F}_m^\dagger\mathbf{E}_m^\sharp\mathbf{F}_m\tilde{\mathbf{T}}_m\right)-\gamma_m N_0\geq 0,\end{aligned}\quad (29)$$

where $\tilde{\mathbf{T}}_m \triangleq \mathbf{t}_m\mathbf{t}_m^\dagger - \gamma_m\sum_{n=1, n\neq m}^{M}\mathbf{t}_n\mathbf{t}_n^\dagger$. Note also that the error bound on $\|\Delta\mathbf{F}_m\|$ can be rewritten as $\mathsf{vec}(\Delta\mathbf{F}_m^\dagger)^\dagger\mathsf{vec}(\Delta\mathbf{F}_m^\dagger)\leq\rho_{\Delta\mathbf{F}_m}^2$. This problem can therefore be optimally solved with rank-relaxation. Due to space limitation, we refer the interested readers to [8] for the details of the derivation and optimization.

V. Stochastic Analysis and A Robust Design Based on Statistical Constraints

A. Derivation of MSE in Data Reception

In this section, based on the CSI estimation in Section III, we look at the statistics of $\Delta\mathbf{F}_m$ and use it to derive the MSE driven by the CSI error. To proceed, it is assumed that we have the ith column of $\mathbf{W}$, $\mathbf{w}_i\sim\mathcal{CN}(\mathbf{0},\sigma_w^2\mathbf{I})$ and $\mathbf{n}\sim\mathcal{CN}(\mathbf{0},\sigma_n^2\mathbf{I})$. In what follows, $\Delta\mathbf{F}_m=(\mathbf{g}_m\mathbf{n}+\mathbf{W})\mathbf{X}_m^\sharp$ is also a zero-mean matrix. The MSE between the estimated and the transmitted signals, averaged over data and noise, is given as

$$\begin{aligned}\mathsf{MSE}_m=&\mathsf{E}[|\hat{s}_m-s_m|^2]\\ =&\mu_m^2\mathbf{h}_m^\dagger\left(\sum_{n=1}^{M}\mathbf{t}_n\mathbf{t}_n^\dagger\right)\mathbf{h}_m\\ &-\mu_m\mathbf{h}_m^\dagger\mathbf{t}_m-\mu_m\mathbf{t}_m^\dagger\mathbf{h}_m+1+\mu_m^2\sigma_n^2.\end{aligned}\quad (30)$$

Then, we can evaluate the MSE by averaging over the channel estimation error $\Delta\mathbf{F}_m$ to yield

$$\begin{aligned}\mathsf{E}_{\Delta\mathbf{F}}[\mathsf{MSE}_m]=1+\mu_m^2\sigma_n^2+\mu_m^2\mathsf{E}\Bigg[\mathsf{tr}\left(\sum_{n=1}^{M}\mathbf{t}_n\mathbf{t}_n^\dagger\mathbf{h}_m\mathbf{h}_m^\dagger\right)\\ -\frac{\mu_m e^{j\phi_m}}{\sqrt{\mathsf{tr}(\mathbf{E}_m)}}\mathbf{e}_m^\dagger(\hat{\mathbf{F}}_m-\Delta\mathbf{F}_m)\mathbf{t}_m\\ -\frac{\mu_m e^{-j\phi_m}}{\sqrt{\mathsf{tr}(\mathbf{E}_m)}}\mathbf{t}_m^\dagger(\hat{\mathbf{F}}_m^\dagger-\Delta\mathbf{F}_m^\dagger)\mathbf{e}_m\Bigg]\end{aligned}\quad (31)$$

which can be further simplified to

$$\begin{aligned}\mathsf{E}_{\Delta\mathbf{F}}[\mathsf{MSE}_m]=\mu_m^2\mathsf{E}\left[\mathsf{tr}\left(\sum_{n=1}^{M}\mathbf{t}_n\mathbf{t}_n^\dagger\mathbf{h}_m\mathbf{h}_m^\dagger\right)\right]\\ -\frac{\mu_m e^{j\phi_m}}{\sqrt{\mathsf{tr}(\mathbf{E}_m)}}\mathbf{e}_m^\dagger\hat{\mathbf{F}}_m\mathbf{t}_m-\frac{\mu_m e^{-j\phi_m}}{\sqrt{\mathsf{tr}(\mathbf{E}_m)}}\mathbf{t}_m^\dagger\hat{\mathbf{F}}_m^\dagger\mathbf{e}_m+1+\mu_m^2\sigma_n^2,\end{aligned}\quad (32)$$

where

$$\begin{aligned}\mathsf{E}[\mathbf{h}_m\mathbf{h}_m^\dagger]&=\mathsf{E}_{\Delta\mathbf{F}}\left[(\hat{\mathbf{F}}_m-\Delta\mathbf{F}_m)^\dagger\mathbf{E}_m^\sharp(\hat{\mathbf{F}}_m-\Delta\mathbf{F}_m)\right]\\ &=\hat{\mathbf{F}}_m^\dagger\mathbf{E}_m^\sharp\hat{\mathbf{F}}_m\\ &\quad+\mathbf{X}_m^{\sharp\dagger}\mathsf{E}\left[(\mathbf{g}_m\mathbf{n}_m^\dagger+\mathbf{W})^\dagger\mathbf{E}_m^\sharp(\mathbf{g}_m\mathbf{n}_m^\dagger+\mathbf{W})\right]\mathbf{X}_m^\sharp,\end{aligned}\quad (33)$$

where the second term can be evaluated by noting that

$$\begin{aligned}\mathsf{E}\left[(\mathbf{g}_m\mathbf{n}_m^\dagger+\mathbf{W})^\dagger\mathbf{E}_m^\sharp(\mathbf{g}_m\mathbf{n}_m^\dagger+\mathbf{W})\right]\\ =\left(\sigma_n^2+\frac{\sigma_w^2}{\mathsf{tr}(\mathbf{E}_m)}\right)\mathbf{I}.\end{aligned}\quad (34)$$

As a result, the MSE in data reception (32) is given by

$$\begin{aligned}\mathsf{E}_{\Delta\mathbf{F}}[\mathsf{MSE}_m]=1+\mu_m^2\sigma_n^2\\ -\frac{\mu_m e^{j\phi_m}}{\sqrt{\mathsf{tr}(\mathbf{E}_m)}}\mathbf{e}_m^\dagger\hat{\mathbf{F}}_m\mathbf{t}_m-\frac{\mu_m e^{-j\phi_m}}{\sqrt{\mathsf{tr}(\mathbf{E}_m)}}\mathbf{t}_m^\dagger\hat{\mathbf{F}}_m^\dagger\mathbf{e}_m\\ +\mu_m^2\sum_{n=1}^{M}\mathbf{t}_n^\dagger\left[\hat{\mathbf{F}}_m^\dagger\mathbf{E}_m^\sharp\hat{\mathbf{F}}_m+\left(\sigma_n^2+\frac{\sigma_w^2}{\mathsf{tr}(\mathbf{E}_m)}\right)\mathbf{X}_m^{\sharp\dagger}\mathbf{X}_m^\sharp\right]\mathbf{t}_n.\end{aligned}\quad (35)$$

B. Robust Design with MSE Constraints

An interesting problem is to design the precoder matrix that ensures the users' MSE requirements. This can be done by

$$\min_{\{\mathbf{t}_n\}_{n=1}^M}\sum_{n=1}^{M}\|\mathbf{t}_n\|^2\quad\text{s.t.}\quad\mathsf{E}_{\Delta\mathbf{F}}[\mathsf{MSE}_m]\leq\varepsilon_m\ \forall m,\quad (36)$$

where $\{\varepsilon_m\}$ are the required users' MSEs. Note in (35) and (36) that $\{\phi_m\}$ are unimportant in the optimization and they can be absorbed into $\{\mathbf{t}_m\}$. Therefore, we can, without loss of generality, set $\phi_m=0\ \forall m$. An immediate difficulty is that (36) is non-convex and that the precoder vectors $\{\mathbf{t}_m\}$ and the scaling variables $\{\mu_m\}$ need to be jointly optimized.

To overcome this, we first note that $\{\mu_m\}$ are not coupled among the users and the MSE in (35) is a quadratic function in μ_m only. Therefore, the optimal μ_m in minimizing the individual MSE for any given $\{\mathbf{t}_m\}$ can be readily found as

$$\mu_m|_{\mathsf{opt}}=\frac{\frac{1}{\sqrt{\mathsf{tr}(\mathbf{E}_m)}}\mathbf{e}_m^\dagger\hat{\mathbf{F}}_m\mathbf{t}_m}{\sigma_n^2+\sum_{n=1}^{M}\mathbf{t}_n^\dagger\left(\hat{\mathbf{F}}_m^\dagger\mathbf{E}_m^\sharp\hat{\mathbf{F}}_m+\left(\sigma_n^2+\frac{\sigma_w^2}{\mathsf{tr}(\mathbf{E}_m)}\right)\mathbf{X}_m^{\sharp\dagger}\mathbf{X}_m^\sharp\right)\mathbf{t}_n}\quad (37)$$

which gives the optimized average MSE as

$$\mathsf{E}_{\Delta\mathbf{F}}[\mathsf{MSE}_m]|_{\mathsf{min}}=1-\frac{\frac{1}{\mathsf{tr}(\mathbf{E}_m)}|\mathbf{e}_m^\dagger\hat{\mathbf{F}}_m\mathbf{t}_m|^2}{\sigma_n^2+\sum_{n=1}^{M}\mathbf{t}_n^\dagger\left(\hat{\mathbf{F}}_m^\dagger\mathbf{E}_m^\sharp\hat{\mathbf{F}}_m+\left(\sigma_n^2+\frac{\sigma_w^2}{\mathsf{tr}(\mathbf{E}_m)}\right)\mathbf{X}_m^{\sharp\dagger}\mathbf{X}_m^\sharp\right)\mathbf{t}_n}.\quad (38)$$

Further manipulation rewrites the MSE constraint in (36) as

$$\sqrt{1-\varepsilon_m}\left\|\begin{bmatrix} \sigma_n \\ \sqrt{\frac{1}{\mathrm{tr}(\mathbf{G}_m)}}\mathbf{e}_m^{\dagger}\hat{\mathbf{F}}_m\mathbf{t}_1 \\ \vdots \\ \sqrt{\frac{1}{\mathrm{tr}(\mathbf{G}_m)}}\mathbf{e}_m^{\dagger}\hat{\mathbf{F}}_m\mathbf{t}_M \\ \sqrt{\left(\sigma_n^2+\frac{\sigma_w^2}{\mathrm{tr}(\mathbf{G}_m)}\right)}\mathbf{X}_m^{\sharp}\mathbf{t}_1 \\ \vdots \\ \sqrt{\sigma_n^2+\frac{\sigma_w^2}{\mathrm{tr}(\mathbf{G}_m)}}\mathbf{X}_m^{\sharp}\mathbf{t}_M \end{bmatrix}\right\| \le \sqrt{\frac{1}{\mathrm{tr}(\mathbf{G}_m)}}\mathbf{e}_m^{\dagger}\hat{\mathbf{F}}_m\mathbf{t}_m. \quad (39)$$

The above constraint is a Second-Order Cone Programming (SOCP) constraint. In addition, the cost function $\sum_{n=1}^{M}\|\mathbf{t}_n\|^2$ is a convex function, so the problem (36) can be reformulated to a convex one and the optimal solution can be easily obtained.

VI. Simulation Results

Simulations are conducted to assess the performance of the proposed robust optimization in Rayleigh flat-fading channels. For convenience, the notation (M, n_T) is used to denote a system with M users and n_T base station antennas, and we assume that $n_d = n_T$, $\sigma_w = \sigma_n$ and users have identical data reception MSE constraints, i.e., $\varepsilon_m = \varepsilon$. The average transmit signal-to-noise ratio (SNR), defined as $\frac{1}{\sigma_n^2}\sum_{m=1}^{M}\mathrm{E}[\|\mathbf{t}_m\|^2]$, is used as the performance metric while the training SNR, defined as $\frac{\mathrm{trace}(\mathbf{X}\mathbf{X}^{\dagger})}{n_d\sigma_n^2}$, is considered as the channel estimation cost. In the simulations, the beamforming design based on the estimated CSI is used as a benchmark and this method will be considered as the "non-robust design" as the channel error statistics are not exploited.

Fig. 1 illustrates the normalized MSE in the quadratic CSI estimates, i.e., $\frac{\|\mathbf{h}\mathbf{h}^{\dagger}-\hat{\mathbf{F}}\mathbf{G}^{\sharp}\hat{\mathbf{F}}\|^2}{\|\mathbf{h}^{\dagger}\mathbf{h}\|^2}$ for both systems using the optimal training sequences and randomly chosen training sequences. Results indicate that there is at least an order of magnitude reduction in the channel estimate MSE by using the optimal sequences and this reduction is more significant when the number of transmit antennas increases. In Fig. 2, results are provided for the user 1's output data reception MSE for various training SNR given that the target MSE is 10^{-3}. As can be seen, regardless of the training SNR, the target MSE is not met for the non-robust design, as opposed to the proposed robust method that the output MSE is exactly the target. In addition, the output MSE is violated more if the training SNR decreases.

VII. Conclusion

This paper has investigated the quadratic CSI estimation and the robust precoder design of a multiuser MISO antenna system in the downlink. The optimal training sequence design has been found. After analyzing the CSI error bound and its statistical property, the robust-optimal designs have been devised. Simulation results have shown the effect of the optimal training sequence and illustrated the robustness of the proposed design. The extension to multiuser MIMO case is not straightforward and needs more work.

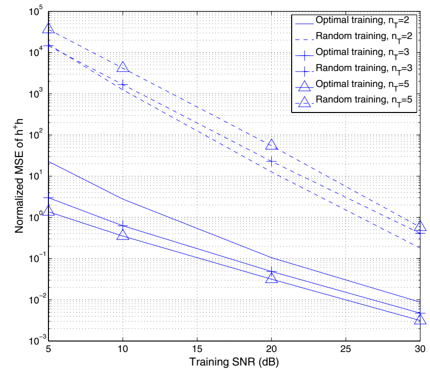

Fig. 1. Normalized MSE in the quadratic CSI estimates against the training SNR for systems using the optimal training sequences and random sequences.

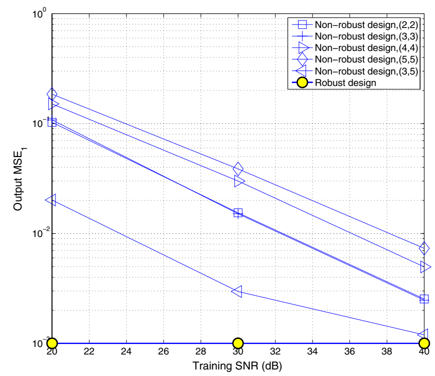

Fig. 2. User 1's MSE of the received data against the training SNR for both the proposed optimal robust algorithm and the non-robust design when the target MSE is $\varepsilon = 10^{-3}$.

References

[1] X. F. Dong, and Z. Ding, "Downlink MIMO channel estimation for transmission precoding," in Proc. *IEEE Global Tel. Conf. (GLOBECOM)*, San Francisco, CA, 27 Nov.-1 Dec. 2006.

[2] G. J. Foschini, and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Per. Commun.*, vol. 6, pp. 311–335, Mar. 1998.

[3] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *Tech. Rep., AT&T Bell Labs*, 1995.

[4] F. R. Farrokhi, K. J. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1437–1450, Oct. 1998.

[5] M. Bengtsson, and B. Ottersten, "Optimal and suboptimal transmit beamforming," *Handbook of Antennas Wireless Commun.*, Aug. 2001.

[6] Z. G. Pan, K. K. Wong, and T. S. Ng, "Generalized multiuser orthogonal space division multiplexing," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 1–5, Nov. 2004.

[7] A. W. Marshall, and I. Olkinr, *Inequality: Theory of Majorization and Its Applications*, New York: Academic, 1979.

[8] G. Zheng, K. K. Wong, and T. S. Ng, "Robust MIMO in the downlink: A worst-case optimization with ellipsoidal uncertainty regions," submitted to *EURASIP J. Advances Sig. Proc.*, 2008.