Optimizing the Rate of a Correlated MIMO Link Jointly over Channel Estimation and Data Transmission Parameters

Alkan Soysal

Department of Electrical and Electronics Engineering
Bahçeşehir University, Istanbul, Turkey
Email: alkan.soysal@bahcesehir.edu.tr

Sennur Ulukus

Department of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742
Email: ulukus@umd.edu

Abstract—We consider a single-user correlated MIMO channel with block fading, where each block is divided into training and data transmission phases. The receiver has a noisy CSI that it obtains through a channel estimation process, while the transmitter has partial CSI in the form of covariance feedback. We optimize the achievable rate jointly over the parameters of the training and data transmission phases. In particular, we first choose the training signal to minimize the channel estimation error, and then, develop an iterative algorithm to solve for the optimum training duration, the optimum allocation of power between training and data transmission phases, and the optimum allocation of power over the antennas during the data transmission phase.

I. INTRODUCTION

In wireless communication scenarios, the achievable rate of a system depends crucially on the amount of channel state information (CSI) available at the receivers and the transmitters. The CSI is observed only by the receiver, which can estimate it and feed the estimated CSI back to the transmitter.

There have been several different assumptions in the literature on the availability of the CSI at the receiver and the transmitter. With perfect CSI at the receiver and the transmitter, the optimum adaptation scheme is water-filling [1], [2]. However, in some cases, especially in multi-input multi-output (MIMO) links, feeding the instantaneous CSI back to the transmitter is not realistic. Therefore, some research assumes that there is perfect CSI at the receiver, but only partial CSI available at the transmitter [3]–[7].

Another line of research considers the actual estimation of the channel at the receiver, which is noisy. The capacity and the corresponding optimum signalling scheme for this case are not known. However, lower and upper bounds for the capacity can be obtained [8]–[10]. It is important to note that [8]–[10] do not consider optimizing the channel estimation process, because of the assumption of the existence of a separate channel that does not consume system resources for channel estimation. For a MIMO system with no CSI available at the transmitter, [11] considers optimizing the achievable rate as a function of both the training and the data transmission phases. In this case, optimizing the achievable rate involves finding the optimal power allocation between the training and data transmission phases.

In this paper, we consider a single-user correlated MIMO channel. The CSI feedback that we consider lies somewhere between perfect CSI [10] and no CSI [11], and it is similar to [3]–[7]. We consider the fact that the training phase uses communication resources, and we optimize the achievable rate of the data transmission phase over the parameters of the training and data transmission processes. Our model differs from [11] in that we consider a correlated channel, which requires a power allocation over the antennas, and we do not have a constraint on the training signal duration, which might result in shorter training signals.

Assuming that the receiver uses linear minimum mean square error (MMSE) detection to estimate the channel during the training phase, we first choose the training signal that minimizes the MMSE. This choice also increases the achievable rate of the data transmission phase [11]. However, unlike [11], our result does not necessarily allocate equal power over the antennas, and might not estimate all of the available channel variables. Then, we move to the data transmission phase, and maximize the achievable rate of the data transmission phase jointly over the rest of the training phase parameters, and data transmission phase parameters, i.e., we find the optimum partition of the given total transmitter power and the block length between the training and the data transmission phases, and we also find the optimum allocation of the data transmission power over the antennas.

II. SYSTEM MODEL

We consider a single-user channel, with n_R receive and n_T transmit antennas, that can be represented by a random matrix $\mathbf{H}$. The channel remains constant for a block (T symbols), and changes to an i.i.d. realization at the end of the block. In order to estimate the channel, the receiver performs a linear MMSE estimation using training symbols over T_t symbols. During the remaining $T_d = T - T_t$ symbols, data transmission occurs. While the receiver has a noisy estimate of the realization of the channel, the transmitter knows only the statistics of the channel. At time n, the transmitter sends a vector $\mathbf{x}_n$, and the received vector is

$$\mathbf{r}_n = \mathbf{H}\mathbf{x}_n + \mathbf{n}_n, \quad n = 1, \ldots, T \qquad (1)$$

where $\mathbf{n}_n$ is a zero-mean, identity-covariance complex Gaussian vector, and the entries of $\mathbf{H}$ are complex Gaussian random

variables. The transmitter has a power constraint P, averaged over T symbols.

The statistical model that we consider is the "partial CSI with covariance feedback" model where the transmitter knows the channel covariance information, in addition to the distribution of the channel. We will assume that the receiver does not have any physical restrictions and therefore, there is sufficient spacing between the antenna elements on the receiver such that the signals received at different antenna elements are uncorrelated. However, there exists correlation between the signals transmitted by different antenna elements. The channel is modeled as [12],

$$\mathbf{H} = \mathbf{Z}\boldsymbol{\Sigma}^{1/2} \qquad (2)$$

where the entries of $\mathbf{Z}$ are i.i.d., zero-mean, unit-variance complex Gaussian random variables. From this point on, we will refer to $\boldsymbol{\Sigma}$ as the channel covariance matrix. Similar covariance feedback models have been used in [3]–[7].

III. Joint Optimization

In our model, a coherence interval, over which the channel is fixed, is divided into two phases: training phase and data transmission phase. The transmitter uses P_t amount of power during the training phase, and P_d amount of power during the data transmission phase. Due to the conservation of energy, we have $PT = P_tT_t + P_dT_d$.

The optimization criterion that we consider is the achievable rate of the data transmission phase. Unlike the case with perfect channel estimation, the data rate here depends on the estimation parameters: training signal $\mathbf{S}$, training signal power P_t, and training signal duration T_t. Therefore, we need to optimize the rate jointly over these channel estimation parameters and the data transmission phase parameters. Intuitively, a longer training phase will result in a better channel estimate and therefore a larger achievable rate during the data transmission phase. However, we use channel resources such as time and power during the channel estimation process, which could otherwise be used for data transmission. A longer training phase implies a shorter data transmission phase, as the block length (coherence time) is fixed. A shorter data transmission phase, in turn, implies a smaller achievable rate. Similarly, the more the training power, the better the channel estimate will be. However, since the total power is fixed, a larger training power will imply a smaller data transmission power, which will decrease the achievable rate. Here, we will solve these trade-offs, and find the optimum training and data transmission parameters.

A. Training and Channel Estimation Phase

In practical communication scenarios, the channel is estimated at the receiver. One way of doing this is to use training symbols before the data transmission starts. The receiver estimates the channel using these training signals and the output of the channel. Since the channel stays the same during the entire block, we can write the input-output relationship

during the training phase in a matrix form as

$$\mathbf{R}_t = \mathbf{HS} + \mathbf{N}_t \qquad (3)$$

where $\mathbf{S}$ is an $n_T \times T_t$ dimensional training signal that will be chosen and known at both ends, $\mathbf{R}_t$ and $\mathbf{N}_t$ are $n_R \times T_t$ dimensional received signal and noise matrices, respectively. The nth column of the matrix equation in (3) represents the input-output relationship at time n. The power constraint for the training input signal is $\frac{1}{T_t}\text{tr}(\mathbf{SS}^{\dagger}) \leq P_t$.

Due to our channel model in (2), the entries in a row of $\mathbf{H}$ are correlated, and the entries in a column of $\mathbf{H}$ are uncorrelated, i.e., rows i and j of the channel matrix are i.i.d. Let us represent row i of $\mathbf{H}$ as $\mathbf{h}_i^{\dagger}$, with $E[\mathbf{h}_i\mathbf{h}_i^{\dagger}] = \boldsymbol{\Sigma}, i = 1, \ldots n_R$. Since rows are i.i.d., the receiver can estimate each of them independently using the same training signal. Now, the ith row of (3) can be written as

$$\mathbf{r}_{ti} = \mathbf{S}^{\dagger}\mathbf{h}_i + \mathbf{n}_{ti}. \qquad (4)$$

The receiver will estimate the ith row of the channel matrix using the received signal $\mathbf{r}_{ti}$, and the training signal $\mathbf{S}$ using linear MMSE estimation. We solve the following optimization problem with $\hat{\mathbf{h}}_i = \mathbf{Mr}_{ti}$ as the estimate of $\mathbf{h}_i$, and $\tilde{\mathbf{h}}_i = \mathbf{h}_i - \hat{\mathbf{h}}_i$ as the channel estimation error,

$$\min_{\mathbf{M}} E\left[\tilde{\mathbf{h}}_i^{\dagger}\tilde{\mathbf{h}}_i\right] = \min_{\mathbf{M}} E\left[\text{tr}\left(\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^{\dagger}\right)\right] \qquad (5)$$

$$= \min_{\mathbf{M}} E\left[\text{tr}\left((\mathbf{h}_i - \mathbf{Mr}_{ti})(\mathbf{h}_i - \mathbf{Mr}_{ti})^{\dagger}\right)\right] \quad (6)$$

Solving the optimum $\mathbf{M}$ from (6) is equivalent to solving $\mathbf{M}$ from the orthogonality principle for vector random variables, which is given as [13, page 91],

$$E\left[(\mathbf{h}_i - \mathbf{Mr}_{ti})\mathbf{r}_{ti}^{\dagger}\right] = \mathbf{0} \qquad (7)$$

where $\mathbf{0}$ is the $n_T \times T_t$ zero matrix. We can solve $\mathbf{M}$ from (7) as

$$\mathbf{M} = E\left[\mathbf{h}_i\mathbf{r}_{ti}^{\dagger}\right]\left(E\left[\mathbf{r}_{ti}\mathbf{r}_{ti}^{\dagger}\right]\right)^{-1}. \qquad (8)$$

By using (4), we calculate $E[\mathbf{h}_i\mathbf{r}_{ti}^{\dagger}] = \boldsymbol{\Sigma}\mathbf{S}$, and $E[\mathbf{r}_{ti}\mathbf{r}_{ti}^{\dagger}] = \mathbf{S}^{\dagger}\boldsymbol{\Sigma}\mathbf{S}+\mathbf{I}$. Then, the optimum $\mathbf{M}$ becomes $\mathbf{M} = \boldsymbol{\Sigma}\mathbf{S}(\mathbf{S}^{\dagger}\boldsymbol{\Sigma}\mathbf{S}+\mathbf{I})^{-1}$. Using this, and the matrix inversion lemma [14, page 19], the mean square error in (6) becomes,

$$\min_{\mathbf{M}} E\left[\tilde{\mathbf{h}}_i^{\dagger}\tilde{\mathbf{h}}_i\right] = \text{tr}\left(\left(\boldsymbol{\Sigma}^{-1} + \mathbf{SS}^{\dagger}\right)^{-1}\right) \qquad (9)$$

Note that the mean square error of the channel estimation process can be further decreased by choosing the training signal $\mathbf{S}$ to minimize (9). In addition, it is stated in [11] that the training signal $\mathbf{S}$ primarily affects the achievable rate through the so called *effective signal-to-noise ratio*, which is shown to be inversely proportional to the MMSE [11]. Therefore, choosing $\mathbf{S}$ to further minimize the MMSE, we also increase the achievable rate of the data transmission phase. The following theorem finds the optimal training signal for a given training power and training duration.

Theorem 1: For given $\boldsymbol{\Sigma} = \mathbf{U}_{\Sigma}\boldsymbol{\Lambda}_{\Sigma}\mathbf{U}_{\Sigma}^{\dagger}$, P_t, T_t, and the power constraint $\text{tr}(\mathbf{SS}^{\dagger}) \leq P_tT_t$, the optimum training input

that minimizes the power of the channel estimation error vector is $\mathbf{S} = \mathbf{U}_\Sigma \boldsymbol{\Lambda}_S^{1/2}$ with

$$\lambda_i^S = \left(\frac{1}{\mu_S} - \frac{1}{\lambda_i^\Sigma} \right)^+, \qquad i = 1, \ldots, \min(n_T, T_t) \quad (10)$$

where μ_S^2 is the Lagrange multiplier that satisfies the power constraint with $\mu_S = \frac{J}{P_t + \sum_{i=1}^J \frac{1}{\lambda_i^\Sigma}}$, and J is the largest index that has non-zero λ_i^S.

A proof of Theorem 1 can be found in [15].

It is important to note that for any given P_t, and $T_t > n_T$, the effect of training length is completely eliminated from the channel estimation problem, i.e., increasing T_t beyond n_T does not result in better channel estimates. However, larger T_t will result in smaller data transmission length, and decrease the achievable rate of the data transmission phase. Therefore, it is sufficient to consider only $T_t \leq n_T$, which we will assume through the rest of this paper.

Theorem 1 tells us that the columns of $\mathbf{S}$ are the weighted columns of a unitary matrix, and they are orthogonal. Since each column of $\mathbf{S}$ is transmitted at a channel use during the training phase, vectors that are transmitted at each channel use during the training phase are orthogonal to each other. This means that, at each channel use, it is optimal to train only one dimension of the channel along one eigenvector. Moreover, the optimum power allocation policy for the training power is to water-fill over the eigenvalues of the channel covariance matrix using (10). Depending on the power constraint and the training signal duration, some of the eigenvalues of the training signal might turn out to be zero. This means that some of the channels along the directions corresponding to zero eigenvalues of the training signal, are not even trained.

The value of T_t determines the total number of available parallel channels in the channel estimation problem, and the value of P_t determines the number of channels that will be estimated. The parametric values of P_t and T_t will appear in the achievable rate formula in the data transmission phase. After the rate maximization is performed, the optimum P_t and T_t will be found, and these in turn, will give us the optimum $\mathbf{S}$ through Theorem 1.

Plugging $\mathbf{S}$ into the covariance of the channel estimation error matrix, $\tilde{\boldsymbol{\Sigma}} = E\left[\tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^\dagger \right] = \left(\boldsymbol{\Sigma}^{-1} + \mathbf{S}\mathbf{S}^\dagger \right)^{-1}$, we find $\tilde{\boldsymbol{\Sigma}} = \mathbf{U}_\Sigma \left(\boldsymbol{\Lambda}_\Sigma^{-1} + \boldsymbol{\Lambda}_S \right)^{-1} \mathbf{U}_\Sigma^\dagger$, where the eigenvalues can be found using (10) as

$$\tilde{\lambda}_i^\Sigma = \begin{cases} \mu_S, & \mu_S < \lambda_i^\Sigma; \\ \lambda_i^\Sigma, & \mu_S > \lambda_i^\Sigma \end{cases} = \min\left(\lambda_i^\Sigma, \mu_S \right). \quad (11)$$

Note that along the directions that we send training signals, i.e., when the corresponding eigenvalues of the training signal are non-zero ($\mu_S < \lambda_i^\Sigma$), the variance of the channel estimation error is the same for all directions. Along the directions that we do not send training signals, the variance of the channel estimation error is equal to the variance of the channel along that direction. This is expected, since the channel is not estimated along that direction, the error in the channel

estimation process is the same as the realization of the channel itself.

Next, we will calculate the eigenvalues of the covariance of the channel estimate. Using the orthogonality property of the MMSE estimation, $\hat{\mathbf{h}}_i$ and $\tilde{\mathbf{h}}_i$ are uncorrelated [13, page 91]. The covariance matrix of the channel estimate $\hat{\boldsymbol{\Sigma}} = E\left[\hat{\mathbf{h}}_i \hat{\mathbf{h}}_i^\dagger \right]$ becomes

$$\hat{\boldsymbol{\Sigma}} = \mathbf{U}_\Sigma \boldsymbol{\Lambda}_\Sigma \mathbf{U}_\Sigma^\dagger - \mathbf{U}_\Sigma \tilde{\boldsymbol{\Lambda}}_\Sigma \mathbf{U}_\Sigma^\dagger \quad (12)$$

$$= \mathbf{U}_\Sigma \left(\boldsymbol{\Lambda}_\Sigma - \tilde{\boldsymbol{\Lambda}}_\Sigma \right) \mathbf{U}_\Sigma^\dagger \triangleq \mathbf{U}_\Sigma \hat{\boldsymbol{\Lambda}}_\Sigma \mathbf{U}_\Sigma^\dagger \quad (13)$$

which has the same eigenvectors as the covariance matrix of the actual channel, however, their eigenvalues are different. We can write each eigenvalue of the covariance matrix of the estimated channel as $\hat{\lambda}_i^\Sigma = \min\left(0, \lambda_i^\Sigma - \mu_S \right)$. Along the directions that we do not send training signals, the value of the channel estimate itself is zero. Therefore, as expected, the power of the estimated channel is zero as well, along those channels with $\mu_S > \lambda_i^\Sigma$.

In the next section, we will plug in these values into the rate formula and develop an algorithm that solves the rate maximization problem of the data transmission phase jointly in terms of the training signal power P_t, training signal duration T_t, and the covariance of the data carrying input signal $\mathbf{Q}$. When the joint optimization problem is solved, the resulting P_t and T_t will determine the optimum training sequence $\mathbf{S}$ through Theorem 1.

B. Data Transmission Phase

When the CSI at the receiver is noisy, the optimum input signaling that achieves the capacity is not known. Following [8]–[11], we derive a lower bound (i.e., an achievable rate) on the capacity for our model, and find the training and data transmission parameters that result in the largest such achievable rate. Using the channel estimation error, $\tilde{\mathbf{H}} = \mathbf{H} - \hat{\mathbf{H}}$, we can write (1) as

$$\mathbf{r} = \hat{\mathbf{H}}\mathbf{x} + \tilde{\mathbf{H}}\mathbf{x} + \mathbf{n}. \quad (14)$$

where $\mathbf{x}$ is the information carrying input, $\mathbf{n}$ is a zero-mean, identity-covariance complex Gaussian vector. Let $\mathbf{Q} = E[\mathbf{x}\mathbf{x}^\dagger]$ be the transmit covariance matrix, which has an average power constraint of P_d, $\text{tr}(\mathbf{Q}) \leq P_d$. Although the optimum input distribution is not known, we achieve the following rate with Gaussian $\mathbf{x}$ for a MIMO channel [10],

$$C_{lb} = I(\mathbf{r}; \mathbf{x}|\hat{\mathbf{H}}) = E\left[\log\left| \mathbf{I} + \mathbf{R}_{\tilde{\mathbf{H}}\mathbf{x}+\mathbf{n}}^{-1} \hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^\dagger \right| \right] \quad (15)$$

where $\mathbf{R}_{\tilde{\mathbf{H}}\mathbf{x}+\mathbf{n}}$ is the covariance matrix of the effective noise, $\tilde{\mathbf{H}}\mathbf{x} + \mathbf{n}$, which is equal to

$$\mathbf{R}_{\tilde{\mathbf{H}}\mathbf{x}+\mathbf{n}} = E\left[\tilde{\mathbf{H}}\mathbf{x}\mathbf{x}^\dagger \tilde{\mathbf{H}}^\dagger \right] + \mathbf{I} = E\left[\tilde{\mathbf{H}}\mathbf{Q}\tilde{\mathbf{H}}^\dagger \right] + \mathbf{I}. \quad (16)$$

By denoting each row of $\tilde{\mathbf{H}}$ as $\tilde{\mathbf{h}}_i^\dagger$, we can write the $(i,j)^{th}$ entry of $E\left[\tilde{\mathbf{H}}\mathbf{Q}\tilde{\mathbf{H}}^\dagger \right]$ as,

$$E\left[\tilde{\mathbf{h}}_i^\dagger \mathbf{Q}\tilde{\mathbf{h}}_j \right] = \begin{cases} \text{tr}(\mathbf{Q}\tilde{\boldsymbol{\Sigma}}), & \text{when } i = j \\ 0, & \text{when } i \neq j \end{cases} \quad (17)$$

which results in $E\left[\tilde{\mathbf{H}}\mathbf{Q}\tilde{\mathbf{H}}^\dagger\right] = \text{tr}(\mathbf{Q}\tilde{\boldsymbol{\Sigma}})\mathbf{I}$. Since our goal is to find the largest such achievable rate, the rate maximization problem over the entire block becomes

$$R = \max_{\substack{(\mathbf{Q},P_t,T_t)\in\mathcal{S} \\ \text{tr}(\mathbf{Q})\le P_d}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^\dagger}{1+\text{tr}(\mathbf{Q}\tilde{\boldsymbol{\Sigma}})}\right|\right] \quad (18)$$

where $\mathcal{S} = \left\{(\mathbf{Q},P_t,T_t)\middle|\text{tr}(\mathbf{Q})T_d + P_tT_t = PT\right\}$, and the coefficient $\frac{T-T_t}{T}$ reflects the amount of time spent during the training phase. The maximization is over the training parameters P_t, and T_t, and the data transmission parameter $\mathbf{Q}$, which can be decomposed into its eigenvectors, i.e., the transmit directions, and eigenvalues, i.e., powers along the transmit directions.

While solving this optimization problem, we will first find the optimum transmit directions of the data transmission phase, which are given by the eigenvectors of $\mathbf{Q}$. We will then focus on the joint optimization of the rate over the eigenvalues (i.e., power distribution over the transmit directions) of $\mathbf{Q}$, the transmit power and the duration of the training phase.

1) Transmit Directions: Unlike the case with no-CSI at the transmitters [11], the optimum transmit covariance matrix is not equal to the identity matrix. In this case, the problem becomes that of choosing the eigenvectors, i.e., the transmit directions, and the eigenvalues, i.e., the powers allocated to the transmit directions, of the transmit covariance matrix $\mathbf{Q} = \mathbf{U}_Q\boldsymbol{\Lambda}_Q\mathbf{U}_Q^\dagger$, to maximize (18).

When the CSI at the receiver is perfect, [4] showed that the eigenvectors of the transmit covariance and the channel covariance matrices must be equal, i.e., $\mathbf{U}_Q = \mathbf{U}_\Sigma$. In the next theorem, we show that this is also true when there is channel estimation error at the receiver.

Theorem 2: Let $\boldsymbol{\Sigma} = \mathbf{U}_\Sigma\boldsymbol{\Lambda}_\Sigma\mathbf{U}_\Sigma^\dagger$ be the spectral decomposition of the covariance feedback matrix of the channel. Then, the optimum transmit covariance matrix $\mathbf{Q}$ has the form $\mathbf{Q} = \mathbf{U}_\Sigma\boldsymbol{\Lambda}_Q\mathbf{U}_\Sigma^\dagger$.

A proof of Theorem 2 can be found in [15]. Using Theorem 2, we can write the optimization problem in (18) as,

$$R = \max_{(\boldsymbol{\lambda}^Q,P_t,T_t)\in\mathcal{P}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\sum_{i=1}^{n_T}\lambda_i^Q\hat{\lambda}_i^\Sigma\hat{\mathbf{z}}_i\hat{\mathbf{z}}_i^\dagger}{1+\sum_{i=1}^{n_T}\lambda_i^Q\tilde{\lambda}_i^\Sigma}\right|\right] \quad (19)$$

where $\mathcal{P} = \left\{\left(\boldsymbol{\lambda}^Q,P_t,T_t\right)\middle|\left(\sum_{i=1}^{n_T}\lambda_i^Q\right)T_d + P_tT_t = PT\right\}$, $\boldsymbol{\lambda}^Q = [\lambda_1^Q,\ldots,\lambda_{n_T}^Q]$, and $\hat{\mathbf{z}}_i$, which is an $n_R\times 1$ dimensional i.i.d., zero-mean, identity-covariance Gaussian random vector, is the ith column of $\hat{\mathbf{Z}}$.

2) Power Allocation Policy: In a MIMO system, a transmit strategy is a combination of a transmit direction strategy, and a transmit power allocation strategy, which is the set of optimum eigenvalues of the transmit covariance matrix, $\boldsymbol{\lambda}^Q$, that solves (19). Although Theorem 2 gives us a very simple closed form solution for the optimum transmit directions, solving (19) for $\boldsymbol{\lambda}^Q$ in a closed form does not seem to be feasible due to the expectation operation in the objective function. Therefore, we will develop an iterative algorithm that finds the optimum $\boldsymbol{\lambda}^Q$.

For a single-user MIMO system with perfect CSI at the receiver and partial CSI at the transmitter in the form of covariance feedback, an algorithm that finds the optimum power allocation policy is proposed in [6]. In this section, we extend the algorithm in [6] to the case when there is channel estimation error at the receiver, or in other words, when we have the training signal power and the training signal duration in the sum-rate expression. The algorithm in [6] cannot be trivially generalized to the model in this paper, since, here we have the training power P_t, and the training duration T_t as additional parameters.

By plugging $\hat{\lambda}_i^\Sigma$ and $\tilde{\lambda}_i^\Sigma$ into (19), and noting that λ_i^Q, for $i = J+1,\ldots,n_T$ does not contribute to the numerator, we have

$$R = \max_{(\boldsymbol{\lambda}^Q,P_t,T_t)\in\mathcal{P}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\sum_{i=1}^J\lambda_i^Q(\lambda_i^\Sigma - \mu_S)\hat{\mathbf{z}}_i\hat{\mathbf{z}}_i^\dagger}{1+\mu_SP_d}\right|\right] \quad (20)$$

From Theorem 1, we know that $J \le T_t$. We further claim that while optimizing the rate, it is sufficient to search over those (P_t,T_t) pairs that result in $J = T_t$ [15]. This creates an additional constraint $P_t > \sum_{i=1}^{T_t}\left(\frac{1}{\lambda_{T_t}^\Sigma} - \frac{1}{\lambda_i^\Sigma}\right)$ that guarantees that, using the pair (P_t,T_t), all T_t channels are filled, i.e., $J = T_t$.

Note that the parameters that we want to optimize (20) over are discrete valued T_t, and continuous valued P_t, and $\boldsymbol{\lambda}^Q$. Since T_t is discrete, and $1 \le T_t \le n_T$, we can perform an exhaustive search over T_t and solve n_T reduced optimization problems with fixed T_t at each one. Then, we take the solution that results in the maximum rate, i.e.,

$$R = \max_{1\le T_t\le n_T} R_{T_t} \quad (21)$$

where R_{T_t} is found by fixing the value of the free parameter T_t in (20). While solving R_{T_t}, we define $f_i(P_t) = \frac{\lambda_i^\Sigma - \mu_S}{1+\mu_SP_d}$, for $i = 1,\ldots,T_t$. In this case, the inner optimization problem becomes

$$R_{T_t} = \max_{(\boldsymbol{\lambda}^Q,P_t)\in\mathcal{R}_{T_t}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \sum_{i=1}^{T_t}\lambda_i^Qf_i(P_t)\hat{\mathbf{z}}_i\hat{\mathbf{z}}_i^\dagger\right|\right] \quad (22)$$

Note that, for the inner optimization problem, in addition to T_t, if P_t was fixed, $f_i(P_t)$ would also be fixed. In this case, the problem in (22) would become exactly the same as the corresponding problem with perfect CSI assumption at the receiver [6], where here, $f_i(P_t)$ replaces λ_i^Σ in [6, equation (8)]. In the optimization problem in (22), we have $T_t + 1$ optimization variables, $\lambda_1^Q,\ldots,\lambda_{T_t}^Q$, and P_t. The KKT conditions can be written as

$$\frac{T_d}{T}f_i(P_t)E\left[\mathbf{z}_i^\dagger\mathbf{A}^{-1}\mathbf{z}_i\right] \le \mu T_d, \quad i = 1,\ldots,T_t \quad (23)$$

$$\frac{T_d}{T}\sum_{i=1}^{T_t}\lambda_i^Q E\left[\mathbf{z}_i^\dagger\mathbf{A}^{-1}\mathbf{z}_i\right]\frac{\partial f_i(P_t)}{\partial P_t} = \mu T_t \quad (24)$$

where $\mathbf{A} = \mathbf{I} + \sum_{i=1}^{T_t}\lambda_i^Qf_i(P_t)\hat{\mathbf{z}}_i\hat{\mathbf{z}}_i^\dagger$, and the equality of

the last equation follows from the complementary slackness condition, which says $P_t > \sum_{i=1}^{T_t}\left(\frac{1}{\lambda_{T_t}^\Sigma} - \frac{1}{\lambda_i^\Sigma}\right)$.

Note that when the optimum λ_i^Q is non-zero, the corresponding inequality in (23) will be satisfied with equality due to its corresponding complementary slackness condition. Therefore, we pull the expectation terms from (23) for those equations with non-zero λ_i^Q's, and insert them into (24). Since those indices with $\lambda_i^Q = 0$ do not contribute to (24), we have

$$\sum_{i=1}^{T_t} \lambda_i^Q \frac{f_i'(P_t)}{f_i(P_t)} = \frac{T_t}{T_d} \tag{25}$$

where we canceled μ's out on both sides. Now, we have a fixed-point equation which does not include any expectation terms. We can use this to solve P_t in terms of λ_i^Q's. Also note that the structure of (23) is the same as the KKT conditions in [6]. Therefore, we propose to update λ_i^Q in the same way as in [6], and between the iterations solve (25) to update P_t. At any given iteration, our algorithm first solves $P_t(n+1)$ from

$$\sum_{i=1}^{T_t} \lambda_i^Q(n) \frac{f_i'(P_t(n+1))}{f_i(P_t(n+1))} = \frac{T_t}{T_d} \tag{26}$$

and then, updates $\lambda_i^Q(n+1)$ using

$$\lambda_i^Q(n+1) = \frac{\lambda_i^Q(n) f_i(P_t(n+1)) E\left[\mathbf{z}_i^\dagger \mathbf{A}^{-1} \mathbf{z}_i\right]}{\sum_j \lambda_j^Q(n) f_j(P_t(n+1)) E\left[\mathbf{z}_j^\dagger \mathbf{A}^{-1} \mathbf{z}_j\right]} P_d(n+1) \tag{27}$$

where $P_d(n+1) = \frac{PT - P_t(n+1)T_t}{T_d}$. This algorithm finds the solution for the training power P_t, and the eigenvalues of the transmit covariance matrix $\lambda_1^Q, \ldots, \lambda_{T_t}^Q$, for a fixed T_t, for $1 \le T_t \le n_T$. We run n_T such algorithms, and the solution of (20) is found by taking the one that results in the largest rate, which gives us the optimum T_t.

As a result, we solved the joint channel estimation and resource allocation problem that we considered in this paper. Through the optimum T_t and P_t, we find the optimum allocation of available time and power over the training and data transmission phases, since total block length and power is fixed. Through Theorem 2, we find the optimum transmit directions, and through $\lambda_1^Q, \ldots, \lambda_{T_t}^Q$, we find the optimum allocation of data transmission power over these transmit directions. Finally, the optimum training signal $\mathbf{S}$ is determined by the optimum T_t and P_t through Theorem 1.

IV. NUMERICAL ANALYSIS

Analytical proof of the convergence of this algorithm seems to be more complicated than the proof in the case when there is no channel estimation error [6], and seems to be intractable for now. However, in our extensive simulations, we observed that the algorithm always converges.

We first consider a system having $n_T = n_R = 2$ with 10 dB total average power and block length $T = 4$. In Figure 1, we plot the eigenvalues of the data transmit matrix and the
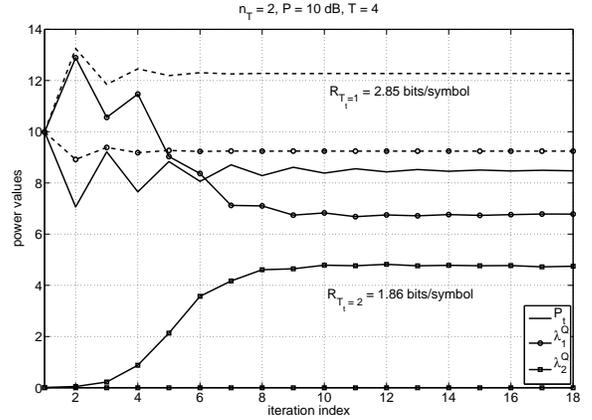
Fig. 1. The convergence of the single-user algorithm with $n_T = n_R = 2$, 10 dB total average power and $T = 4$. The dashed curves correspond to one symbol long training, $T_t = 1$, and solid curves correspond to two symbols long training, $T_t = 2$.
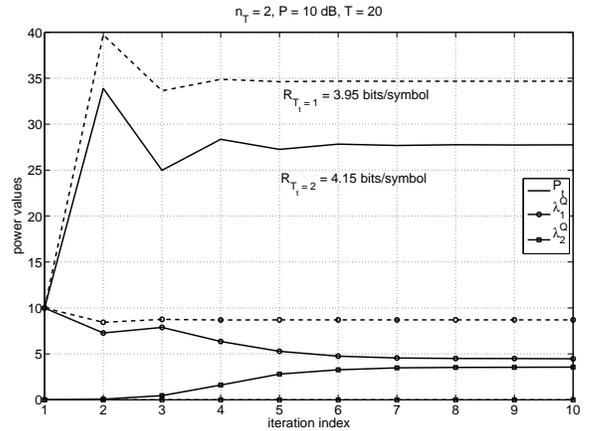
Fig. 2. The convergence of the single-user algorithm with $n_T = n_R = 2$, 10 dB total average power and $T = 20$. The dashed curves correspond to one symbol long training, $T_t = 1$, and solid curves correspond to two symbols long training, $T_t = 2$.

training power as a function of the iteration index for both possible values of the training signal duration. We observe that when the training duration is one symbol period, we achieve a higher rate. Therefore, for this set of given system parameters, estimating only one dimension of the channel is optimum.

Next, we investigate the effect of the block length, in Figure 2, we consider 10 dB total average power, and block length $T = 20$. We observe that having two symbols long training phase is optimum. We repeat this experiment with different numbers of antennas, and channel eigenvalues for long block lengths, and we see that moderate block lengths are sufficient in order to use more than one symbol of training. Therefore, we conclude that for very fast changing channels where the coherence interval and therefore the block length is short, estimating only one dimension of the channel results in higher achievable rates. In this case, we cannot take advantage of the multiple dimensions that the MIMO channel provides,
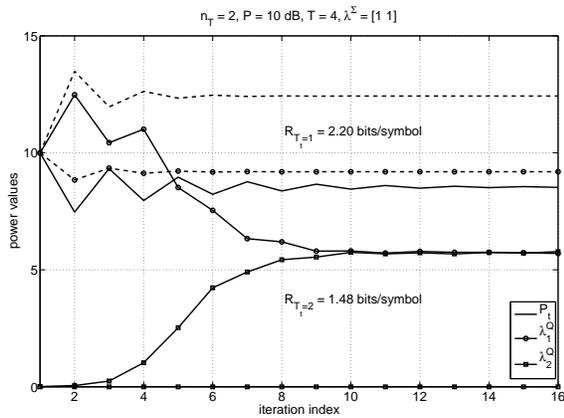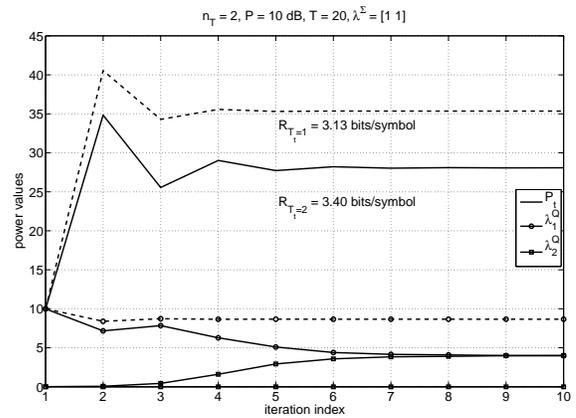
Fig. 3. The convergence of the single-user algorithm with $n_T = n_R = 2$, 10 dB total average power, and channel eigenvalues $\boldsymbol{\lambda}^\Sigma = [1, 1]$, where dashed curves correspond to one symbol long training, $T_t = 1$, and solid curves correspond to two symbols long training, $T_t = 2$: (a) $T = 4$; (b) $T = 20$.

because the amount of time required to estimate those channels cancels the data rate advantage brought by having multiple channels.

We next analyze the effects of different channel covariance matrices. In Figure 3, we consider 10 dB average power, and a channel covariance matrix that has both eigenvalues equal. Note that this case is exactly the case considered [11]. However, in this paper, we do not assume the restriction that $T_t \geq n_T$ as it was assumed in [11] by reasoning that one needs at least $T_t \geq n_T$ measurements in order to estimate n_T variables. Although this reasoning is valid, we relax this restriction by pointing out that in some cases, we might not want to estimate n_T variables. If the resources are limited, estimating some of the variables and saving the resources for data transmission is more useful. As a result, in this paper, we find that the duration of the training signal is equal to the number of variables to be estimated rather than the total number of variables. Figure 3 supports our findings, by showing that, for a short block length $T = 4$ with 10 dB total power, not estimating one of the dimensions results in a higher data transmission rate. This advantage disappears when the block length is long enough.

V. CONCLUSIONS

We analyzed the joint optimization of the channel estimation and data transmission parameters of a single-user MIMO block-fading channel where the receiver has a noisy estimate of the channel and the transmitter has the partial CSI in the form of covariance feedback. We first found the optimum training signal to minimize the MMSE, and then, we formulated the joint optimization problem over the eigenvalues of the transmit covariance matrix and the channel estimation parameters. We solved this problem by introducing a number of reduced optimization problems, each of which can be solved efficiently using the proposed iterative algorithm. Through simulations, we observed that the proposed iterative algorithm converges and it converges to the same point regardless of the

initial point of the iterations.

REFERENCES

[1] A. J. Goldsmith and P. P. Varaiya. Capacity of fading channels with channel side information. *IEEE Transactions on Information Theory*, 43(6):1986–1992, November 1997.

[2] İ. E. Telatar. Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunication*, 10(6):585–596, November 1999.

[3] E. Visotsky and U. Madhow. Space-time transmit precoding with imperfect feedback. *IEEE Transactions on Information Theory*, 47(6):2632–2639, September 2001.

[4] S. A. Jafar and A. Goldsmith. Transmitter optimization and optimality of beamforming for multiple antenna systems. *IEEE Transactions on Wireless Communications*, 3(4):1165–1175, July 2004.

[5] H. Boche and E. Jorswieck. On the optimality range of beamforming for MIMO systems with covariance feedback. *IEICE Trans. Commun.*, E85-A(11):2521–2528, November 2002.

[6] A. Soysal and S. Ulukus. Optimum power allocation for single-user MIMO and multi-user MIMO-MAC with partial CSI. *IEEE Journal on Selected Areas in Communications*, 25(7):1402–1412, September 2007.

[7] A. Soysal and S. Ulukus. Optimality of beamforming in fading MIMO multiple access channels. *IEEE Transactions on Communications*, 2008. Under review.

[8] M. Médard. The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel. *IEEE Transactions on Information Theory*, 46(3):933–946, May 2000.

[9] T. E. Klein and R. G. Gallager. Power control for the additive white Gaussian noise channel under channel estimation errors. In *ISIT*, June 2001.

[10] T. Yoo and A. Goldsmith. Capacity and power allocation for fading MIMO channels with channel estimation error. *IEEE Transactions on Information Theory*, 52(5):2203–2214, May 2006.

[11] B. Hassibi and B. M. Hochwald. How much training is needed in multiple-antenna wireless links? *IEEE Transactions on Information Theory*, 49(4):951–963, April 2003.

[12] C. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela. Capacity scaling in MIMO wireless systems under correlated fading. *IEEE Transactions on Information Theory*, 48(3):637–650, March 2002.

[13] E. W. Kamen and J. Su. *Introduction to Optimal Estimation*. Springer, 1999.

[14] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[15] A. Soysal and S. Ulukus. Joint channel estimation and resource allocation for MIMO systems–Part I: Single-user analysis. *IEEE Transactions on Wireless Communications*. Submitted.