On Optimal Training and Beamforming in Uncorrelated MIMO Systems with Feedback

Francisco Rubio*, Dongning Guo†, Michael L. Honig† and Xavier Mestre*
*Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)
(PMT) Av. Canal Olímpic s/n, 08860 Castelldefels (Barcelona), Spain
E-mail:{francisco.rubio,xavier.mestre}@cttc.es
†Dept. of Electrial Engineering and Computer Science, Northwestern University
2145 Sheridan Road, Evanston, IL 60208, USA
E-mail:{dguo,mh}@northwestern.edu

***Abstract*—This paper studies the design and analysis of optimal training-based beamforming in uncorrelated multiple-input multiple-output (MIMO) channels with known Gaussian statistics. First, given the response of the MIMO channel to a finite sequence of training vectors, the beamforming vector which maximizes the average received signal-to-noise ratio (SNR) over all channel realizations is found. Secondly, the question of what consists of optimal training for a given amount of training is addressed. Upper and lower bounds for the maximum achievable SNR using beamforming are established. Furthermore, optimal training sequences are conjectured to satisfy the Welch bound. The conjecture is supported by the evidence that such sequences achieve close to the upper bound with moderate to large amount of trainings.**

I. Introduction

The performance of multiple-input multiple-output (MIMO) channels can be significantly enhanced if the channel state is known to the transmitter, the receiver, or both. Alternative blind techniques applied in order to avoid channel training may often incur in a nonnegligible loss of performance and a fairly increased computational complexity. In practice, the coefficients of a MIMO channel often vary over time and need to be estimated. The impact of the realistic availability of an imprecise channel state information in the capacity gains predicted for MIMO systems are summarized in [1]. If the channel state varies slowly, one may carry out some measurements in order to learn the channel statistics and estimate (or predict) its instantaneous realization. Typically, the channel coefficients are measured at the receiver by having the transmitter send known training vectors. Knowledge of the channel at the receiver can be sent to the transmitter via feedback channels [2].

The tradeoff between the time and the power allocated to training operation and data transmission was evaluated in [3]. In particular, they provided the optimum number of pilots and training power allocation of a training-based MIMO system in the sense of maximizing a lower-bound on the Shannon capacity over the class of ergodic block-fading (memoryless and uncorrelated) channels, as a function of the number of transmit and receive antennas, the received signal-to-noise ratio (SNR) and the length of the fading coherence time. Earlier related contributions include [4], as well as [5], [6], where the number of channel uses available for training and the optimal input distribution achieving capacity at high SNR over unknown block-fading uncorrelated MIMO channels with a finite coherence time interval is investigated. Furthermore, the effects of pilot-assisted channel estimation on achievable data rates over frequency-flat time-varying channels is analyzed in [7]. Along with the time-division multiplexing training scheme considered in the previous works, a tight lower-bound on the maximum mutual information of a MIMO system using superimposed pilots is derived in [8].

One simple use of the channel state information at the transmitter is to modulate the transmitted symbol onto a beamforming vector matched to the channel in order to improve the received SNR. In this paper, we focus on the previous beamforming approach with the aim of exploiting the diversity gain achieved over the MIMO channel. If the MIMO channel is completely known to the transmitter, the evident choice of the beamforming vector is the right eigenvector of the channel matrix corresponding to the maximum singular value in amplitude, which maximizes the received SNR.

Consider training-based beamforming for uncorrelated MIMO channels with known statistics. We assume that a sequence of known training vectors are sent by the transmitter so that the receiver can estimate the channel. In [9], the capacity of beamforming over a block-fading MIMO channel is maximized with respect to the finite amount of training available for channel estimation as well as of limited feedback. Instead of following the generally suboptimal approach consisting of obtaining an intermediate estimate of the channel matrix to be used in further post-processing, we pursue the direct estimation of the optimal (channel-adapted) beamformer vector that needs to be fed back. In particular, the following questions are addressed in this paper:

1) Given the channel response to training, what is the optimal beamforming vector which maximizes the received SNR averaged over all possible realizations of the channel?

This work was supported by the Catalan Government (DURSI) under grant BE-2007, the U.S. Army Research Office under grant DAAD19-99-1-0288, the NSF CAREER Award under grant CCF-0644344, and the DARPA IT-MANET program Grant W911NF-07-1-0028.

2) Given a total budget for training purposes, such as the total energy consumed by training, what is the optimal sequence of training vectors to use?

We first describe the MIMO channel in Section II. Section III then provides an exact answer to question 1 in above. Section IV presents a partial answer to question 2, where an optimal choice of training vector is conjectured and shown to achieve close to an upper bound on the performance of optimal training.

II. Channel Model

Consider a linear channel model corresponding to a MIMO transmission system with M receive antennas and K transmit antennas, namely, the received signal is expressed as

$$\mathbf{y}(t) = \mathbf{H}\,\mathbf{x}(t) + \mathbf{n}(t), \quad t = 1, 2, \ldots \tag{1}$$

where $\mathbf{x}(t) \in \mathbb{C}^K$ represents the transmitted signal, $\mathbf{n}(t) \in \mathbb{C}^M$ the background noise samples, and $\mathbf{H} \in \mathbb{C}^{M\times K}$ the MIMO channel matrix. It is assumed that the entries of $\mathbf{H}$, $\{H\}_{m,k}, m = 1, \ldots, M, k = 1, \ldots, K$, are independent identically distributed (i.i.d.) circularly symmetric complex Gaussian with zero mean and unit variance. The noise process is modeled as wide-sense stationary, with standarized complex Gaussian vector entries such that $\mathbb{E}\left[\mathbf{n}(s)\mathbf{n}(t)^H\right] = \delta_{s,t}\mathbf{I}_M$, where $\delta_{s,t}$ is the Kronecker delta function.

The purpose of using multiple antennas here is to enhance the received SNR through beamforming. Specifically, one wishes to modulate a sequence of transmitted symbols $x(t)$ onto a unit-norm beamforming vector $\mathbf{v} \in \mathbb{C}^K$ ($\mathbf{x}(t) = \mathbf{v}x(t)$), so that the received signal becomes

$$\mathbf{y}(t) = \mathbf{H}\,\mathbf{v}\,x(t) + \mathbf{n}(t), \quad t = 1, 2, \ldots \tag{2}$$

Assuming $\mathbb{E}\left\{|x(t)|^2\right\} = P$, the received SNR is then

$$\mathsf{SNR} = \mathbb{E}\left\{\mathbf{v}^\dagger\mathbf{H}^\dagger\mathbf{H}\mathbf{v}\right\} P. \tag{3}$$

Suppose the channel $\mathbf{H}$ is known completely or partially to the transmitter, then one wishes to find a beamforming vector matched to the channel, so that the SNR is maximized in some sense.

In order to construct the beamforming vector, a set of training beams are assumed to be available. Then, given a time interval of certain length is available for the purposes of training, we are interested in the optimal construction of a transmit beamforming vector maximizing the SNR during the allocated period of time. Additionally, as the achieved performance clearly depends on the available training beams, we are also concerned about the optimality of the training sequence. These problems are addressed in the following.

III. Optimal Beamforming

A. Estimation of Optimal Beamforming Vector

In this section, we seek the optimum transmit beamforming vector for a MIMO system where a statistical characterization of the channel is available through training. Suppose a sequence of T fixed beams are used for training. The posterior mean and covariance of the (random) channel entries conditioned on the received observations allow us to formulate the optimal beamforming problem as that of finding the unit length (deterministic) vector $\mathbf{v}$ such that the SNR over the training interval is maximized. Then, this problem can be formulated as

$$\max_{\mathbf{v}:\|\mathbf{v}\|=1} \mathbb{E}\left[\|\mathbf{H}\mathbf{v}\|^2 \middle| \mathbf{y}(1), \ldots, \mathbf{y}(T)\right] \tag{4}$$

with the optimal solution given by the top eigenvector of

$$\mathbb{E}\left[\mathbf{H}^\dagger\mathbf{H}\middle|\mathbf{y}(1), \ldots, \mathbf{y}(T)\right] = \mathbb{E}\left[\mathbf{H}^\dagger\mathbf{H}\middle|\mathbf{Y}\right] \tag{5}$$

where

$$\mathbf{Y} = \mathbf{H}\mathbf{B} + \mathbf{N} \tag{6}$$

with

$$\mathbf{Y} = [\mathbf{y}(1), \ldots, \mathbf{y}(T)] \tag{7}$$

$$\mathbf{B} = [\mathbf{b}(1), \ldots, \mathbf{b}(T)], \tag{8}$$

$$\mathbf{N} = [\mathbf{n}(1), \ldots, \mathbf{n}(T)]. \tag{9}$$

Now, we have

Lemma 1: The optimal training vector is the eigenvector associated with the maximum eigenvalue of the matrix $\mathbb{E}\left[\mathbf{H}^\dagger\mathbf{H}\middle|\mathbf{Y}\right]$, which is given by

$$M\left(\mathbf{B}\mathbf{B}^\dagger + \mathbf{I}_K\right)^{-1} + \left(\mathbf{B}\mathbf{B}^\dagger + \mathbf{I}_K\right)^{-1}\mathbf{B}\mathbf{Y}^\dagger\mathbf{Y}\mathbf{B}^\dagger\left(\mathbf{B}\mathbf{B}^\dagger + \mathbf{I}_K\right)^{-1}. \tag{10}$$

Proof: Observe first that the entries of $\mathbb{E}\left[\mathbf{H}^\dagger\mathbf{H}\middle|\mathbf{Y}\right]$ can be found as

$$\left\{\mathbb{E}\left[\mathbf{H}^\dagger\mathbf{H}\middle|\mathbf{Y}\right]\right\}_{i,j} = \sum_{m=1}^{M}\mathbb{E}\left[H_{m,i}^*H_{m,j}\middle|\mathbf{Y}\right]. \tag{11}$$

Now, note that both (independent) entries $H_{m,i}$ and $H_{m,j}$ belong to the same (mth-)row of the channel matrix $\mathbf{H}$. Thus, obtaining the posterior correlation of the random variables $(H_{m,i}, H_{m,j})$, for $i, j = 1, \ldots, K$, is equivalent to obtaining the posterior correlation of this row vector (note that, by assumption, the covariance does not depend on the index m). Additionally, consider the model

$$\bar{\mathbf{Y}} = \bar{\mathbf{B}}\bar{\mathbf{H}} + \bar{\mathbf{N}} \tag{12}$$

where $\bar{\mathbf{H}} = \left[\bar{\mathbf{h}}_1, \ldots, \bar{\mathbf{h}}_M\right] \equiv \mathbf{H}^\dagger \in \mathbb{C}^{K\times M}$, and $\bar{\mathbf{Y}} = [\bar{\mathbf{y}}(1), \ldots, \bar{\mathbf{y}}(M)]$, $\bar{\mathbf{B}} = \left[\bar{\mathbf{b}}(1), \ldots, \bar{\mathbf{b}}(K)\right]$ and $\bar{\mathbf{N}} = [\bar{\mathbf{n}}(1), \ldots, \bar{\mathbf{n}}(M)]$ are equivalently defined. Using these definitions, we observe that the previous problem reduces to the following Bayesian LMMSE estimation problem, namely,

$$\bar{\mathbf{y}}(m) = \bar{\mathbf{B}}\bar{\mathbf{h}}_m + \bar{\mathbf{n}}(m) \tag{13}$$

From (13), it is clear that the columns of $\bar{\mathbf{Y}}$ (or, accordingly, the rows of $\mathbf{Y}$) are independent. (In fact, note that $\bar{\mathbf{y}}(m)$ is the only contribution in the received observations helping to describe the posterior statistics of $\bar{\mathbf{h}}_m$.) Since $\bar{\mathbf{y}}(m)$ and $\bar{\mathbf{h}}_m$ are jointly Gaussian, we can use the well-known fact that

$$\mathbb{E}\left[\bar{\mathbf{h}}_m|\bar{\mathbf{y}}(m)\right] = \mathbf{C}_{\bar{\mathbf{h}}_m\bar{\mathbf{y}}(m)}\mathbf{C}_{\bar{\mathbf{y}}(m)}^{-1}\bar{\mathbf{y}}(m) \tag{14}$$

and

$$\mathbf{C}_{\bar{\mathbf{h}}_m|\bar{\mathbf{y}}(m)} = \mathbf{C}_{\bar{\mathbf{h}}_m} - \mathbf{C}_{\bar{\mathbf{h}}_m\bar{\mathbf{y}}(m)}\mathbf{C}^{-1}_{\bar{\mathbf{y}}(m)}\mathbf{C}_{\bar{\mathbf{y}}(m)\bar{\mathbf{h}}_m} \tag{15}$$

where the covariances are

$$\mathbf{C}_{\bar{\mathbf{h}}_m} = \mathbf{I}_K, \tag{16}$$
$$\mathbf{C}_{\bar{\mathbf{y}}(m)} = \mathbf{B}^\dagger\mathbf{B} + \mathbf{I}_T \tag{17}$$
$$\mathbf{C}_{\bar{\mathbf{h}}_m\bar{\mathbf{y}}(m)} = \mathbf{C}^\dagger_{\bar{\mathbf{y}}(m)\bar{\mathbf{h}}_m} = \mathbf{B}. \tag{18}$$

Thus, we have

$$\mathbb{E}\left[\bar{\mathbf{h}}_m|\bar{\mathbf{y}}(m)\right] = \mathbf{B}\left(\mathbf{B}^\dagger\mathbf{B} + \mathbf{I}_T\right)^{-1}\bar{\mathbf{y}}(m) \tag{19}$$
$$= \left(\mathbf{B}\mathbf{B}^\dagger + \mathbf{I}_K\right)^{-1}\mathbf{B}\bar{\mathbf{y}}(m) \tag{20}$$

and

$$\mathbf{C}_{\bar{\mathbf{h}}_m|\bar{\mathbf{y}}(m)} = \mathbf{I}_K - \mathbf{B}\left(\mathbf{B}^\dagger\mathbf{B} + \mathbf{I}_T\right)^{-1}\mathbf{B}^\dagger \tag{21}$$
$$= \left(\mathbf{B}\mathbf{B}^\dagger + \mathbf{I}_K\right)^{-1}. \tag{22}$$

Hence, we can find the posterior correlation as

$$\mathbb{E}\left[\mathbf{H}^\dagger\mathbf{H}\middle|\,\mathbf{Y}\right]$$
$$= \sum_{m=1}^{M}\mathbb{E}\left[\bar{\mathbf{h}}_m\bar{\mathbf{h}}_m^\dagger\middle|\,\mathbf{Y}\right] \tag{23}$$
$$= \sum_{m=1}^{M}\mathbb{E}\left[\bar{\mathbf{h}}_m\bar{\mathbf{h}}_m^\dagger\middle|\,\bar{\mathbf{y}}(m)\right] \tag{24}$$
$$= \sum_{m=1}^{M}\mathbf{C}_{\bar{\mathbf{h}}_m|\bar{\mathbf{y}}(m)} + \mathbb{E}\left[\bar{\mathbf{h}}_m|\bar{\mathbf{y}}(m)\right]\mathbb{E}\left[\bar{\mathbf{h}}_m|\bar{\mathbf{y}}(m)\right]^\dagger \tag{25}$$

which, after some algebra, becomes (10).

Note that Lemma 1 admits also the following alternative proof. We can directly use the fact that the matrices $\bar{\mathbf{Y}}$ and $\bar{\mathbf{H}}$ are jointly matrix-variate normal distributed with mean zero and covariances[1] $\mathbf{C}_{\bar{\mathbf{h}}} = \mathbf{I}_M \otimes \mathbf{I}_K = \mathbf{I}_{KM}$ and $\mathbf{C}_{\bar{\mathbf{y}}} = \mathbf{I}_M \otimes \bar{\mathbf{B}}\bar{\mathbf{B}}^\dagger + \mathbf{I}_{TM} = \mathbf{I}_M \otimes \bar{\mathbf{B}}\bar{\mathbf{B}}^\dagger + \mathbf{I}_M \otimes \mathbf{I}_T = \mathbf{I}_M \otimes \left(\bar{\mathbf{B}}\bar{\mathbf{B}}^\dagger + \mathbf{I}_T\right)$, respectively, with $\otimes$ denoting the Kronecker tensor-product. In order to find $\mathbb{E}\left[\mathbf{H}^\dagger\mathbf{H}\middle|\,\mathbf{Y}\right]$, we just need to obtain the distribution of $\bar{\mathbf{H}}$ conditioned on $\bar{\mathbf{Y}}$, or, equivalently, the conditional distribution of $\bar{\mathbf{h}} = \mathrm{vec}\left(\bar{\mathbf{H}}\right)$ given $\bar{\mathbf{y}} = \left(\mathbf{I}_M \otimes \bar{\mathbf{B}}\right)\bar{\mathbf{h}} + \tilde{\mathbf{n}}$. Clearly, $\bar{\mathbf{h}}$ and $\bar{\mathbf{y}}$ are jointly Gaussian distributed, so that, using

$$\mathbf{C}_{\bar{\mathbf{y}},\bar{\mathbf{h}}} = \mathbb{E}\left[\left(\left(\mathbf{I}_M \otimes \bar{\mathbf{B}}\right)\bar{\mathbf{h}} + \tilde{\mathbf{n}}\right)\bar{\mathbf{h}}^\dagger\right] = \mathbf{I}_M \otimes \bar{\mathbf{B}} \tag{26}$$

the posterior mean of $\bar{\mathbf{h}}$ conditioned on the observation of $\bar{\mathbf{y}}$ can be found to be given by

$$\mathbb{E}\left[\bar{\mathbf{h}}\middle|\,\bar{\mathbf{y}}\right] = \mathbf{C}^\dagger_{\bar{\mathbf{y}},\bar{\mathbf{h}}}\mathbf{C}^{-1}_{\bar{\mathbf{y}}}\bar{\mathbf{y}} \tag{27}$$
$$= \left(\mathbf{I}_M \otimes \bar{\mathbf{B}}\right)^\dagger\left(\mathbf{I}_M \otimes \left(\bar{\mathbf{B}}\bar{\mathbf{B}}^\dagger + \mathbf{I}_T\right)\right)^{-1}\bar{\mathbf{y}} \tag{28}$$
$$= \left(\mathbf{I}_M \otimes \bar{\mathbf{B}}^\dagger\left(\bar{\mathbf{B}}\bar{\mathbf{B}}^\dagger + \mathbf{I}_T\right)^{-1}\right)\bar{\mathbf{y}} \tag{29}$$
$$= \left(\mathbf{I}_M \otimes \left(\bar{\mathbf{B}}^\dagger\bar{\mathbf{B}} + \mathbf{I}_K\right)^{-1}\bar{\mathbf{B}}^\dagger\right)\bar{\mathbf{y}} \tag{30}$$

[1]The covariance of a random matrix is defined as the covariance of the vector obtained by lining up all of its columns.

and the conditional variance given by

$$\mathbf{C}_{\bar{\mathbf{h}}|\bar{\mathbf{y}}} = \mathbf{C}_{\bar{\mathbf{h}}} - \mathbf{C}^\dagger_{\bar{\mathbf{y}},\bar{\mathbf{h}}}\mathbf{C}^{-1}_{\bar{\mathbf{y}}}\mathbf{C}_{\bar{\mathbf{y}},\bar{\mathbf{h}}} \tag{31}$$
$$= \mathbf{I}_{KM} - \left(\mathbf{I}_M \otimes \bar{\mathbf{B}}^\dagger\left(\bar{\mathbf{B}}\bar{\mathbf{B}}^\dagger + \mathbf{I}_T\right)^{-1}\bar{\mathbf{B}}\right) \tag{32}$$
$$= \mathbf{I}_M \otimes \left(\mathbf{I}_K - \bar{\mathbf{B}}^\dagger\left(\bar{\mathbf{B}}\bar{\mathbf{B}}^\dagger + \mathbf{I}_T\right)^{-1}\bar{\mathbf{B}}\right) \tag{33}$$
$$= \mathbf{I}_M \otimes \left(\mathbf{I}_K + \bar{\mathbf{B}}^\dagger\bar{\mathbf{B}}\right)^{-1} \tag{34}$$

where we have used the matrix inversion lemma in order to obtain the last equality. Accordingly, the channel coefficients $\bar{\mathbf{H}}$ conditioned on $\bar{\mathbf{Y}}$ is again matrix-variate normal distributed with mean $\mathbf{M} = \left(\bar{\mathbf{B}}^\dagger\bar{\mathbf{B}} + \mathbf{I}_K\right)^{-1}\bar{\mathbf{B}}^\dagger\bar{\mathbf{Y}}$ and covariance $\mathbf{C}_{\bar{\mathbf{h}}|\bar{\mathbf{y}}} = \mathbf{\Psi} \otimes \mathbf{\Sigma}$, where we have defined $\mathbf{\Psi} = \mathbf{I}_M$ and $\mathbf{\Sigma} = \left(\bar{\mathbf{B}}^\dagger\bar{\mathbf{B}} + \mathbf{I}_K\right)^{-1}$. From the properties of the matrix-normal distribution [10], we have that

$$\mathbb{E}\left[\bar{\mathbf{H}}\bar{\mathbf{H}}^\dagger\middle|\,\mathbf{Y}\right] = \mathrm{Tr}\left[\mathbf{\Psi}\right]\mathbf{\Sigma} + \mathbf{M}\mathbf{M}^\dagger \tag{35}$$

which is again equivalent to (10). ■

B. Beamforming Based on Channel Estimation

Note that an indirect (suboptimal) approach would consist in constructing the quadratic form $\mathbf{H}^\dagger\mathbf{H}$ using an estimate of the MIMO channel matrix. In particular, the MMSE estimator and the LS estimator (here also the maximum-likelihood estimator) of $\mathbf{H}$ are, respetively,

$$\hat{\mathbf{H}}_{\mathsf{MMSE}} = \mathbf{Y}\mathbf{B}^\dagger\left(\mathbf{B}\mathbf{B}^\dagger + \mathbf{I}_K\right)^{-1}, \tag{36}$$
$$\hat{\mathbf{H}}_{\mathsf{LS}} = \mathbf{Y}\mathbf{B}^\dagger\left(\mathbf{B}\mathbf{B}^\dagger\right)^{-1}, \tag{37}$$

and the solution for the beamforming vector can be accordingly obtained as the top eigenvector of, respectively,

$$\hat{\mathbf{H}}^\dagger_{\mathsf{MMSE}}\hat{\mathbf{H}}_{\mathsf{MMSE}} = \left(\mathbf{B}\mathbf{B}^\dagger + \mathbf{I}_K\right)^{-1}\mathbf{B}\mathbf{Y}^\dagger\mathbf{Y}\mathbf{B}^\dagger\left(\mathbf{B}\mathbf{B}^\dagger + \mathbf{I}_K\right)^{-1}, \tag{38}$$
$$\hat{\mathbf{H}}^\dagger_{\mathsf{LS}}\hat{\mathbf{H}}_{\mathsf{LS}} = \left(\mathbf{B}\mathbf{B}^\dagger\right)^{-1}\mathbf{B}\mathbf{Y}^\dagger\mathbf{Y}\mathbf{B}^\dagger\left(\mathbf{B}\mathbf{B}^\dagger\right)^{-1}. \tag{39}$$

In Section V we compare the performance of the beamforming schemes described in Section III-A and III-B.

IV. Optimal training

We now consider the optimal design of the training sequence used to construct the optimal transmit beamformer. In particular, we are interested in a set of training beams such that

$$\max_{\mathbf{B}:\|\mathbf{B}\|_F \le E}\mathbb{E}_{\mathbf{H},\mathbf{N}}\left[\lambda_{\max}\left\{\mathbf{H}^\dagger\mathbf{H}\middle|\,\mathbf{Y}\right\}\right], \tag{40}$$

where E determines the constraint on the total energy consumed by training.

A. A Constraint Optimization Problem

From the definition of $\mathbf{Y}$ in (6), the estimator provided by Lemma 1 is equivalently distributed as

$$\mathbb{E}\left[\mathbf{H}^\dagger\mathbf{H}\middle|\,\mathbf{Y}\right] = M\left\{\left(\mathbf{B}\mathbf{B}^\dagger + \mathbf{I}_K\right)^{-1} + \mathbf{\Xi}\mathbf{\Xi}^\dagger\right\} \tag{41}$$

where

$$\mathbf{\Xi} = \left(\mathbf{B}\mathbf{B}^\dagger + \mathbf{I}_K\right)^{-1}\mathbf{B}\left[\begin{array}{cc}\mathbf{B}^\dagger & \mathbf{I}_T\end{array}\right]\mathbf{W} \tag{42}$$

with the entries of $\mathbf{W} \in \mathbb{C}^{(K+T)\times M}$ being i.i.d. circularly symmetric complex Gaussian with mean zero and variance $1/M$ (with some abuse of notation, in the sequel $\mathbf{W}$ will denote a matrix distributed as in here with pertinent dimensions). From the last expression, using the matrix inversion lemma, we recognize $\mathbb{E}\left[\mathbb{E}\left[\mathbf{H}^\dagger\mathbf{H}\middle|\,\mathbf{Y}\right]\right] = M\mathbf{I}_K$, as it is straightforward from the properties of the conditional expectation, namely, by the law of iterated expectations, $\mathbb{E}\left[\mathbb{E}\left[\mathbf{H}^\dagger\mathbf{H}\middle|\,\mathbf{Y}\right]\right] = \mathbb{E}\left[\mathbf{H}^\dagger\mathbf{H}\right]$.

Furthermore, from the SVD of $\mathbf{B}$, i.e., $\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger$, with $\mathbf{\Sigma}$ being a K-dimensional square matrix, and the unitary invariance property of Wishart matrices, we have

$$\mathbb{E}\left[\mathbf{H}^\dagger\mathbf{H}\middle|\,\mathbf{Y}\right] = M\mathbf{U}\left\{\left(\mathbf{\Sigma}^2 + \mathbf{I}_K\right)^{-1} + \mathbf{\Theta}\mathbf{\Theta}^\dagger\right\}\mathbf{U}^\dagger \tag{43}$$

where

$$\mathbf{\Theta} = \left(\mathbf{\Sigma}^2 + \mathbf{I}_K\right)^{-1}\mathbf{\Sigma}\left[\begin{array}{cc}\mathbf{\Sigma} & \mathbf{I}_K\end{array}\right]\mathbf{W}. \tag{44}$$

Since we are interested in the eigenvalues of the estimator (in particular in the maximum eigenvalue), we may directly focus on the matrix $\left(\mathbf{\Sigma}^2 + \mathbf{I}_K\right)^{-1} + \mathbf{\Theta}\mathbf{\Theta}^\dagger$, which is clearly distributed equivalently as

$$\left(\mathbf{\Sigma}^2 + \mathbf{I}_K\right)^{-1} + \left(\mathbf{\Sigma}^2 + \mathbf{I}_K\right)^{-\frac{1}{2}}\mathbf{\Sigma}\mathbf{W}\mathbf{W}^\dagger\mathbf{\Sigma}\left(\mathbf{\Sigma}^2 + \mathbf{I}_K\right)^{-\frac{1}{2}} \tag{45}$$

and has the same eigenvalues as

$$\left(\mathbf{\Sigma}^2 + \mathbf{I}_K\right)^{-1} + \left(\mathbf{\Sigma}^2 + \mathbf{I}_K\right)^{-1}\mathbf{\Sigma}^2\mathbf{W}\mathbf{W}^\dagger. \tag{46}$$

Then, the problem of optimum training sequence design reduces to the following optimization problem, namely,

$$\max_{\mathbf{\Sigma}^2:\,\mathrm{tr}\,\mathbf{\Sigma}^2\le E}\mathbb{E}\left[\lambda_{\max}\left\{\left(\mathbf{\Sigma}^2 + \mathbf{I}_K\right)^{-1} + \left(\mathbf{\Sigma}^2 + \mathbf{I}_K\right)^{-1}\mathbf{\Sigma}^2\mathbf{W}\mathbf{W}^\dagger\right\}\right]. \tag{47}$$

Furthermore, define

$$\mathbf{D} = (\mathbf{\Sigma}^2 + \mathbf{I}_K)^{-1}\mathbf{\Sigma}^2, \tag{48}$$

and write

$$\mathbb{E}\left[\mathbf{H}^\dagger\mathbf{H}\middle|\,\mathbf{Y}\right] = M\mathbf{U}\left\{(\mathbf{I}_K - \mathbf{D}) + \mathbf{D}^{1/2}\mathbf{W}\mathbf{W}^\dagger\mathbf{D}^{1/2}\right\}\mathbf{U}^\dagger.$$

Then, the constraint on the training power can be written as $\mathrm{tr}\,\mathbf{\Sigma}^2 = \mathrm{tr}\left(\mathbf{D}\left(\mathbf{I}_K - \mathbf{D}\right)^{-1}\right) \le E$, and the training matrix $\mathbf{B}$, or, equivalently the matrix $\mathbf{D}$, should be chosen to achieve

$$\max_{\mathbf{D}:\,\mathrm{tr}\left(\mathbf{D}(\mathbf{I}_K-\mathbf{D})^{-1}\right)\le E}\mathbb{E}\left[\lambda_{\max}\left\{(\mathbf{I}_K - \mathbf{D}) + \mathbf{D}^{1/2}\mathbf{W}\mathbf{W}^\dagger\mathbf{D}^{1/2}\right\}\right]. \tag{49}$$

Conjecture 1: The solution to (49) in $\mathbf{D}$ has equal diagonal elements.

In other words, $\mathbf{\Sigma} = \sqrt{E/K}\mathbf{I}_K$ and hence the training vectors $\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger$ satisfies that $\mathbf{B}\mathbf{B}^\dagger = \mathbf{I}$. Therefore, the conjectured optimal training sequences (colums of $\mathbf{B}$) satisfy the Welch-bound equality (WBE) [11], [12]. In the multiuser detection literature, WBE signature sequences are known to maximize the sum capacity achieved by overloaded symbol-synchrous code-division multiple-access channels with equal average-input-energy constraints [13], [14].

B. An Equivalent Constraint

Not being able to show Conjecture 1, we distill it to an equivalent but more concise mathematical formulation in the following. We rewrite (49) by shedding $\mathbf{I}_K$ and rearranging the order of the matrices in the product, so that it is equivalent to finding $\mathbf{D}$ achieving

$$\max_{\mathbf{D}:\,\mathrm{tr}\left(\mathbf{D}(\mathbf{I}_K-\mathbf{D})^{-1}\right)\le E}\mathbb{E}\left[\lambda_{\max}\left\{\mathbf{D}\left(\mathbf{W}\mathbf{W}^\dagger - \mathbf{I}_K\right)\right\}\right] \tag{50}$$

The constraint $\mathrm{tr}\left(\mathbf{D}\left(\mathbf{I}_K - \mathbf{D}\right)^{-1}\right) \le E$ (i.e., $\mathrm{tr}\,\mathbf{\Sigma}^2 \le E$) is not convenient. We show how to relax the constraint $\mathrm{tr}\,\mathbf{\Sigma}^2 \le E$ to a constraint on $\mathrm{tr}\,\mathbf{D}$. For the purpose of showing Conjecture 1, it is enough to show

Conjecture 2: The solution to the following over K dimensional diagonal $\mathbf{D}$ is $\mathbf{D} = \mathbf{I}_K$,

$$\max_{\mathbf{D}:\,\mathrm{tr}\,\mathbf{D}\le K}\mathbb{E}\left[\lambda_{\max}\left\{\mathbf{D}\left(\mathbf{W}\mathbf{W}^\dagger - \mathbf{I}_K\right)\right\}\right]. \tag{51}$$

Remark 1: The constraint $\mathrm{tr}\,\mathbf{D} \le K$ can be replaced by $\mathrm{tr}\,\mathbf{D} \le E'$ for any $E' > 0$ because the maximum eigenvalue is linear in E', i.e., one can always scale $\mathbf{D}$ by $1/\mathrm{tr}\,\mathbf{D} \le K$ so that the trace is limited to 1.

Next, we claim that the original constraint

$$\mathrm{tr}\left(\mathbf{D}\left(\mathbf{I}_K - \mathbf{D}\right)^{-1}\right) \le E \tag{52}$$

is tighter than

$$\mathrm{tr}\,\mathbf{D} \le \frac{K^2}{E+K} = E'. \tag{53}$$

The reason is that $x/(1-x)$ is a concave function. In fact the nontrivial boundary of the two sets touches at the point $D_i \equiv K/(E+K)$, i.e., $\mathbf{D} = (K/(E+K))\mathbf{I}_K$.

Now, if the maximum under the relaxed constraint (53) is achieved by $\mathbf{D} = (K/(E+K))\mathbf{I}_K$, which in fact satisfies the original constraint (52), the maximization under the original constraint is also achieved by $\mathbf{D} = (K/(E+K))\mathbf{I}_K$, which implies Conjecture 1.

Note, on the other hand that Conjecture 1, if true for all $E > 0$, also implies Conjecture 2. The reason is that the constraint (52) reduces to $\mathrm{tr}\,\mathbf{D} \le E$ as E becomes small (D_i is small so that $(1 - D_i) \approx 1$), as well as the fact that scaling of $\mathbf{D}$ does not change the solution to Conjecture 2.

Based on the above, it is enough to show Conjecture 2, which is the hardest version of Conjecture 1, where $E \to 0$.

C. An Upper-bound on the Optimal SNR

The conjectured optimal power allocation over time consists of a training covariance matrix equal to a scaled identity meeting the energy constraint. More specifically, the diagonals of $\mathbf{D}$, or, equivalently $\mathbf{\Sigma}^2$, are all equal and given by $\mathbf{D} = (E/(E+K))\mathbf{I}_K$. This special choice of training leads to the following received SNR:

$$\mathbb{E}\left[\lambda_{\max}\left\{\mathbf{D}\left(\mathbf{W}\mathbf{W}^\dagger - \mathbf{I}_K\right)\right\}\right] = \\ = \frac{E}{E+K}\left(\mathbb{E}\left[\lambda_{\max}\left\{\mathbf{W}\mathbf{W}^\dagger\right\}\right] - 1\right). \quad (54)$$

Note that all elements of $\mathbf{D}$ are constrained to be less than 1, so that using $\mathbf{D} = \mathbf{I}_K$ leads to an upper-bound, namely,

$$\mathbb{E}\left[\lambda_{\max}\left\{\mathbf{D}\left(\mathbf{W}\mathbf{W}^\dagger - \mathbf{I}_K\right)\right\}\right] = \mathbb{E}\left[\lambda_{\max}\left\{\mathbf{W}\mathbf{W}^\dagger\right\}\right] - 1. \quad (55)$$

Using results in the literature on the moment generating function of the maximum eigenvalue of a complex central Wishart matrix [15], the expectation defining equations (54) and (55) can be obtained as

$$\mathbb{E}\left[\lambda_{\max}\left\{\mathbf{W}\mathbf{W}^\dagger\right\}\right] = 1 + \sum_{l=1}^{K}\xi_l, \quad (56)$$

with ξ_l defined as

$$\sum_{\{\alpha_{l,K}\}}\sum_{\{\beta\}}\frac{(-1)^{l+\mathrm{per}(\beta)}\hat{\Gamma}_K(M,\beta)}{\Gamma_K(K)\Gamma_K(M)}\sum_{s=0}^{T(\beta,\alpha_{l,K})}\sum_{\{s_1,\ldots,s_l\}}\binom{s}{s_1,\ldots,s_l}\frac{1}{l^2} \quad (57)$$

where $\mathrm{per}(\cdot)$ denotes the number of permutations, $\beta = \{\beta_1,\ldots,\beta_K\}$ is a permutation of the index set $[K] = \{1,\ldots,K\}$, $\alpha_{l,K} = \{\alpha_1,\ldots,\alpha_l\}$ is a subset of $[K]$ with $\alpha_1 < \ldots < \alpha_l$; $\Gamma(\cdot)$ is the usual gamma function and

$$\Gamma_K(M) = \prod_{i=1}^{K}\Gamma(M-i+1) \quad (58)$$

$$\hat{\Gamma}_K(M,\beta) = \prod_{j=1}^{K}\Gamma(\tau+j+\beta_j-1), \quad (59)$$

where we have defined $\tau = M-K$. Moreover,

$$T(\beta,\alpha_{l,K}) = \sum_{i=1}^{K}\left(g(\alpha_i,\beta_{\alpha_i})-1\right) \quad (60)$$

where $g(i,j) = \tau+i+j-1$, and the last sum in (57) is over all integer partitions of $s = s_1+\cdots+s_l$ satisfying $0 \le s_i < g(\alpha_i,\beta_{\alpha_i})$, $i = 1,\ldots,l$.

Thus, we have that

$$U_{\mathsf{bound}} = \mathbb{E}\left[\lambda_{\max}\left\{\mathbf{D}\left(\mathbf{W}\mathbf{W}^\dagger - \mathbf{I}_K\right)\right\}\right] = \sum_{l=1}^{K}\xi_l. \quad (61)$$

The upper-bound in (61) corresponds to the asympotic regime defined by either an infinite training period length or an infinite available power. Denoting $P = \|\mathbf{b}(n)\|^2$, the problem of allocating power over time can be related to the energy limitation problem as $E = P \times T$ (or $E = \mathsf{SNR} \times T$ assuming unit additive noise variance). Note that the gap between the upper-bound and the conjectured solution is small in percentage if $E >> K$, which is almost always the case in practice. In particular, we can establish the following rate of convergence for the conjectured optimal SNR to achieve the SNR upper-bound, namely,

$$\frac{MU_{\mathsf{bound}}}{\mathbb{E}\left[\lambda_{\max}\left\{\mathbf{H}^\dagger\mathbf{H}|\,\mathbf{Y}\right\}\right]} = 1 + \mu_{tx}, \quad (62)$$

where $\mu_{tx} = K/E$ is a function of the total energy budget (depending on the SNR and the training length) as well as the number of transmit antennas. For instance, note that for $\mu_{tx} = 1$, a loss of 3dB is to be expected.

D. Lower-bound

Clearly, a lower-bound is obtained when trying to estimate the instantaneous realization of the Gram matrix $\mathbf{H}^\dagger\mathbf{H}$ with only one sample. In our framework, this is equivalent to a matrix $\mathbf{D}$ with a certain diagonal entry equal to $E/(E+1)$ and all others fixed to zero, leading to

$$\mathbb{E}\left[\lambda_{\max}\left\{\mathbf{D}\left(\mathbf{W}\mathbf{W}^\dagger - \mathbf{I}_K\right)\right\}\right] = \mathbb{E}\left[\max\{0,Q\}\right], \quad (63)$$

where $Q = \frac{1}{M}\sum_{m=1}^{M}|w_m|^2 - 1$, with w_m, $m = 1,\ldots,M$, being standarized complex Gaussian random variables and $M(Q+1)$ a random variable following a chi-square law with M degrees of freedom. In particular, from the probability density function of the chi-square law and the pertinent transformations of random variables, the expectation in (63) can be found as

$$\mathbb{E}\left[\max\{0,Q\}\right] = \frac{\left(se^{-1}\right)^s}{\Gamma(s)}\sum_{k=0}^{s-1}\binom{s-1}{k}\frac{(k+1)!}{s^{k+2}}, \quad (64)$$

where we have defined $s = M/2$.

The lower-bound in (64) corresponds to the suboptimal situation in which all the energy is concentrated as the power allocated to only a certain time instant, whereas the transmitter is switched off during the rest of the training period. This scheme is equivalent to a strategy according to which only a certain direction is trained during the entire training phase.

V. Numerical results

In this section, we numerically evaluate the performance of the optimal training vector and the analysis of the optimality of different training beams. In the simulations, we fix $P = \mathsf{SNR} = 1$, and, without loss of generality, relate an increase in the energy limitations to the availability of a longer training period (larger sample-size).

Figure 1 illustrates the averaged conditional SNR, i.e., $\mathbb{E}\left[\mathsf{SNR}(T+1)|\,\mathbf{y}(1),\ldots,\mathbf{y}(T)\right] = \hat{\mathbf{v}}^\dagger\mathbf{H}^\dagger\mathbf{H}\hat{\mathbf{v}}$, where $\hat{\mathbf{v}}$ is the beamformer estimate obtained as the top eigenvector of three different estimators, namely the conditional mean estimator (MMSE estimator) of the quadratic form $\mathbf{H}^\dagger\mathbf{H}$, derived in Section III, as well as the approximation of the previous Gram matrix constructed using the MMSE estimator and the LS

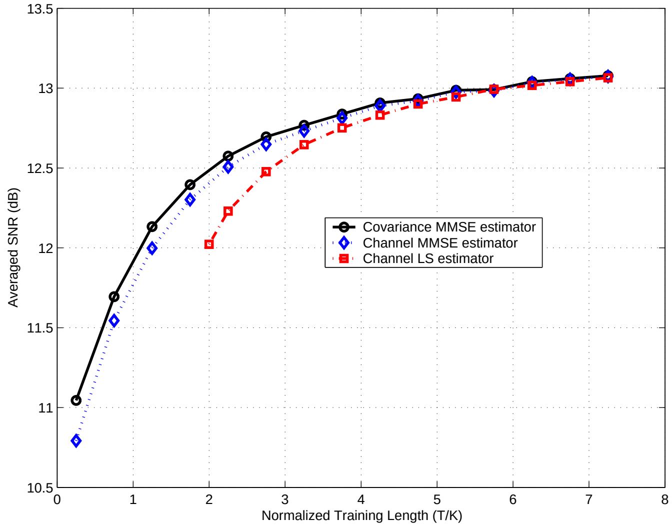

Fig. 1. Performance measure in (3) for different transmit beamforming estimators. $M = 12, K = 4$

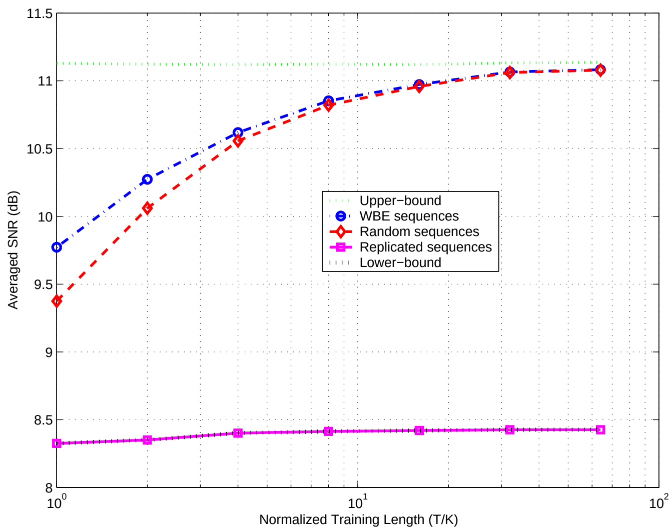

Fig. 2. Performance measure in (40) for different training strategies estimators. $M = 6, K = 4$

estimator of the MIMO channel matrix, respetively. Random training has been assumed. Note that the three estimates coincide for the case of the conjectured optimal training design, as they all produce the same principal eigenvector.

In Figure 2, the SNR approximation in (40) is shown for different training strategies determining the allocation of the (limited) power throughout the training interval so as to meet the energy constraint. Binary complex sequences of normalized (unit-norm) vectors were used. In particular, the selected choices correspond to the conjectured optimal training covariance with all diagonals equal (obtained by sequences meeting the Welch-bound equality, for which $\mathbf{\Sigma}^2 = T/K\mathbf{I}_K$, or, accordingly, $\mathbf{D} = T/\left(T+K\right)\mathbf{I}_K$), randomly drawn training beams (with norm one and complex binary entries), and a training matrix $\mathbf{B}$ with columns consisting of a random training beam replicated throught the training interval. Along with the performance for the different choices of $\mathbf{B}$, the upper and lower bounds are also shown. As mentioned above, the upper-bound corresponds to the high SNR regime or an infinite training sample-support and the lower-bound corresponds to the suboptimal situation in which all available energy is concentrated on a single mode, such that only one certain direction is trained. Clearly, the lower-bound is met by the training strategy consisting of a set of replicated training beams. In order to straightforwardly analyze the system performance with WBE sequences meeting the assumptions above, the length of the training period is chosen as $T = \{4, 8, 16, 32, 64, 128, 256\}$.

VI. Concluding Remarks

Given the response of a MIMO channel to a sequence of training vectors, we have shown that the optimal beamforming vector is the top eigenvector of an appropriately defined conditional covariance of the channel conditioned on the training response. We also formulated the problem of finding the optimal training sequences for maximizing the received signal-to-noise ratio, and conjectured that such sequences satisfy the Welch bound. The conjecture is distilled to a concise open mathematical problem. Using upper and lower bounds for the optimal received SNR, we show that the Welch-bound sequences achieves close to the upper bound with moderate to large amount of training.

References

[1] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, June 2003.
[2] D. Love, J. R.W. Heath, W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 54–59, October 2004.
[3] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, April 2003.
[4] T. Marzetta, "BLAST training: Estimating channel characteristics for high-capacity space-time wireless," in *Proc. 37th Annual Allerton Conference on Communications, Control, and Computing*, September 1999.
[5] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139–157, January 1999.
[6] L. Zheng and D. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. on Information Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.
[7] D. Samardzija and N. Mandayam, "Pilot-assisted estimation of MIMO fading channel response and achievable data rates," *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2882–2890, November 2003.
[8] M. Coldrey and P. Bohlin, "Training-based MIMO systems Part I: Performance comparison," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5464–5476, November 2007.
[9] W. Santipach and M. Honig, "Optimization of training and feedback for beamforming over a MIMO channel," in *Proc. IEEE Wireless Commun. and Networking Conf.*, March 2007.
[10] T. Kollo and D. von Rosen, *Advanced Multivariate Statistics with Matrices*. Springer, 2005.
[11] L. Welch, "Lower bounds on the maximum cross correlation of signals," *IEEE Trans. on Information Theory*, vol. 20, pp. 397–399, May 1974.
[12] S. Verdú, *Multiuser Detection*. NY: Cambridge University Press, 1998.
[13] J. L. Massey and T. Mittelholzer, "Welch's bound and sequence sets for code-division multiple access systems," in *in Sequences II, Methods in Commun., Security and Computer Science*, R. Capocelli, A. D. Santis, and U. Vaccaro, Eds. New York: Springer-Verlag, 1993, pp. 200–266.
[14] M. Rupf and J. Massey, "Optimum sequence multisets for synchronous code-division multiple-access channels," *IEEE Transactions on Information Theory*, vol. 40, pp. 1261–1266, July 1994.
[15] M. McKay and I. Collings, "Improved general lower bound for spatially-correlated Rician MIMO capacity," *IEEE Communications Letters*, vol. 10, no. 3, pp. 951–963, March 2006.