Large System Analysis of Beamforming for MIMO Systems with Limited Training

Francisco Rubio*, Dongning Guo†, Michael L. Honig† and Xavier Mestre*
*Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)
(PMT) Av. Carl Friedrich Gauss, 7, 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*—We consider multiple-input multiple-output (MIMO) systems exploiting the full diversity order of a MIMO fading channel via optimal beamforming and combining. Specifically, an analytical characterization of the transient regime of a training-based MIMO system over arbitrarily correlated channels is presented. No channel state information is assumed to be available at either the transmitter or the receiver side, so that the design of the optimal transmit and receive beamformers is necessarily based on a finite collection of samples observed during a training phase. The focus is on practical scenarios where the length of the training sequence is comparable in magnitude to the system size. In these situations, the performance of the MIMO system can be expected to suffer from a considerable degradation. In order to characterize the actual performance under the previous realistic conditions, a large-system performance analysis is proposed that builds upon some new results on the convergence of the eigenvectors of large information-plus-noise covariance matrices.**

***Index Terms*—MIMO channel, training, optimal beamforming, full diversity gain, sample covariance matrix, random matrix theory, asymptotic eigenvector**

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. In practice, the coefficients of a MIMO channel often vary over time and need to be estimated. 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 [1].

The impact of the practical availability of imprecise channel state information (CSI) in the capacity gains achieved in MIMO spatial multiplexing systems is summarized in [2]. On the other hand, the tradeoff between the time and the power allocated to training operation and data transmission was evaluated in [3]. In particular, the authors provide 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.

Much less effort has been placed on understanding the consequences of the lack of CSI on the achieved diversity gain of an unknown MIMO channel that is learned by means of a training sequence of finite length. Indeed, perfect knowledge of the channel realization can be used in general to modulate each transmitted symbol onto a beamforming vector matched to the channel in order to improve the received SNR. In particular, if the MIMO channel is completely known to the transmitter, the evident choice of the beamforming vector is the right singular vector of the channel matrix corresponding to the maximum singular value in amplitude, which maximizes the received SNR. In [7], the problem of optimal transmit beamforming maximizing the received SNR over unknown MIMO channels with given Gaussian statistics is addressed.

In this paper, we focus on the problem of achieving full diversity gain over an unknown, arbitrary block-fading MIMO channel by optimal transmit beamforming and receive combining. The problem formulation here builds upon the work in [8], where least-squares filtering with a limited number of training symbols is analyzed for the suppression of multiple-access interference at reception (without feedback). In particular, we assume a certain given amount of channel uses is allocated for training purposes at the beginning of each coherence interval, such that both sides can learn the channel from a sequence of known training beams. Instead of following the generally suboptimal approach consisting of obtaining an intermediate estimate of the channel matrix to be used for further processing, we pursue the direct estimation of both optimal (channel-adapted) beamformer vector and receive combiner using the sequence of pilots during the so-called

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.

training phase. In particular, we are interested in the actual empirical performance obtained from a limited number of training samples per degree-of-freedom. More specifically, we provide a large-system analysis of the transient estimation regime of such a training-based MIMO scheme in which the number of transmit and receive antennas as well as the length of the training sequence are considered, as in practice, to be comparable in magnitude. For that purpose, we investigate the asymptotic convergence of the eigenvectors of large-dimensional information-plus-noise covariance matrices by relying on existing results from random matrix theory.

The paper is organized as follows. In Section II, the problem of pilot-aided MIMO transmitter and receiver estimation is addressed. In Section III, we provide a brief overview of the existing convergence results regarding the asymptotic behavior of the eigenvectors of sample covariance matrices. Moreover, the main mathematical result of the paper concerning information-plus-noise-type covariance matrices, on the basis of which our findings are grounded, is introduced without proof at the end of the section. In Section IV, we present a large system performance analysis of the transient estimation regime of a MIMO system with limited training. Finally, the proposed approximation of the practically achieved diversity gain is numerically validated in Section V.

II. Channel model and transceiver estimation

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

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

where $\mathbf{x}(n) \in \mathbb{C}^K$ represents the transmitted signal, $\mathbf{n}(n) \in \mathbb{C}^M$ is the background noise, and $\mathbf{H} \in \mathbb{C}^{M\times K}$ models an arbitrary MIMO channel matrix. The noise process is assumed to be wide-sense stationary, with independent and identically distributed (i.i.d.) standardized complex Gaussian[1] vector entries such that $\mathbb{E}\left[\mathbf{n}(l)\,\mathbf{n}(m)^H\right] = \sigma_{\mathbf{n}}^2 \delta_{l,m}\mathbf{I}_M$, where $\delta_{l,m}$ is the Kronecker delta function. Without loss of generality, we will assume in the following $\sigma_{\mathbf{n}}^2 = 1$. Specifically, one wishes to modulate a sequence of transmitted symbols $x(n)$ onto a (unit-norm) beamforming vector $\mathbf{v} \in \mathbb{C}^K$ ($\mathbf{x}(n) = \mathbf{v}x(n)$), so that the received signal in (1) becomes

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

As mentioned above, the purpose of using multiple antennas here is to enhance through beamforming the SNR at the receiver side and after matched filtering, namely,

$$\mathsf{SNR} = \left|\mathbf{u}^H\mathbf{H}\mathbf{v}\right|^2, \tag{3}$$

where $\mathbf{u} \in \mathbb{C}^M$ represents the receiver matched to the MIMO channel. In particular, the receiver and transmitter vectors maximizing the SNR are respectively the right and left top singular vectors of $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, henceforth denoted by $\mathbf{u}_1$ and $\mathbf{v}_1$. Accordingly, if $\mathsf{SNR}_k = [\mathbf{\Sigma}]_{k,k}^2$ is the signal-to-noise ratio associated with the kth channel eigenmode, it is clear from (3) that the maximum achievable SNR is $\mathsf{SNR}_1 = \max_k \mathsf{SNR}_k$.

We assume that no CSI is available at either the transmitter or the receiver side, and that a sequence of N fixed pilot beams $\mathbf{b}(n) \in \mathbb{C}^K$ consuming a certain given amount of training energy is available for transceiver estimation purposes. In this work, we will consider $K \leq N$. Accordingly, the received signal becomes ($\mathbf{x}(n) = \mathbf{b}(n)$)

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

By collecting the column vector observations in (2) at different instants of time in a matrix $\mathbf{Y} \in \mathbb{C}^{M\times N}$, we can write

$$\mathbf{Y} = [\mathbf{y}(1), \ldots, \mathbf{y}(N)] = \mathbf{H}\mathbf{B} + \mathbf{N},$$

where we have defined

$$\mathbf{B} = [\mathbf{b}(1), \ldots, \mathbf{b}(N)],$$
$$\mathbf{N} = [\mathbf{n}(1), \ldots, \mathbf{n}(N)].$$

In the following, we consider the problem of empirical estimation of the optimal transceiver given a fixed training energy budget (i.e., power allocation strategy across pilot beams and length of training phase), namely,

$$\{\mathbf{u}, \mathbf{v}\} = \arg \max_{\mathbf{u},\mathbf{v}:\|\mathbf{B}\|_F^2 \leq E} \mathbb{E}\left[\mathsf{SNR}|\, \mathbf{y}(1), \ldots, \mathbf{y}(N)\right],$$

where E determines the constraint on the total energy consumed by training. In particular, note that the total energy constraint will be related to the power allocated to the beamvector pilots sent during the training phase, as well as the length of this training window (i.e., number of training beams). Furthermore, for estimation purposes, observe that $\mathbf{u}_1$ is the top eigenvector of $\mathbf{H}\mathbf{H}^H$, whereas $\mathbf{v}_1$ is the top eigenvector of $\mathbf{H}^H\mathbf{H}$. The achieved system performance based on pilot-assisted transceiver estimation clearly depends on the selection of training beams. In this work, we will focus on the more relevant case in practice of orthogonal training. In particular, we assume that the training phase is defined by a set of orthogonal (unitary) beams satisfying the training budget constraint, such that $\mathbf{B}\mathbf{B}^H = E/K\mathbf{I}_K$. In other words, the training sequences (column vectors of $\mathbf{B}$) satisfy the Welch-bound equality (WBE) [9], [10]. In the multiuser detection literature, WBE signature sequences are known to maximize the sum capacity achieved by overloaded symbol-synchronous code-division multiple-access channels with equal average-input-energy constraints [11], [12]. The optimality of WBE sequences for transmit beamforming schemes maximizing the received SNR is discussed in [7]. For the purpose of statistically analyzing the effect of limited training in the performance of pilot-assisted MIMO systems, it will be in order to assume in the sequel, and with some abuse of notation, the following model for the training matrix, namely, $\mathbf{B} = \sqrt{E/K}\mathbf{U}^H$, where the columns of $\mathbf{U} \in \mathbb{C}^{N\times K}$ are orthogonal, such that $\mathbf{U}^H\mathbf{U} = \mathbf{I}_K$.

[1] A complex random variable is standarized complex Gaussian if its real and imaginary parts are i.i.d. Gaussian distributed with mean zero and variance 1/2.

A. Receiver estimation

Since the top eigenvector of $\mathbf{H}\mathbf{H}^H$ is equal to the principal eigenvector of the covariance matrix of the received observations, namely,

$$\mathbf{R} = \mathbb{E}\left[\mathbf{y}\left(n\right)\mathbf{y}^H\left(n\right)\right] = E/\left(KN\right)\mathbf{H}\mathbf{H}^H + \mathbf{I}_M,$$

the problem of estimating $\mathbf{u}_1$ can be directly approached by equivalently finding an estimator of the top eigenvector of $\mathbf{R}$. To that effect, we may use the sample estimate of the latter, namely the sample covariance matrix (SCM), i.e.,

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{y}\left(n\right)\mathbf{y}^H\left(n\right) = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H. \tag{4}$$

From the strong law of large numbers, the SCM $\hat{\mathbf{R}}$ is a consistent estimator of the theoretical covariance matrix. In fact, the SCM is the minimum variance unbiased estimator of $\mathbf{R}$ [13]. Moreover, for Gaussian observations, the maximum-likelihood (ML) estimator of the principal eigenvector of $\mathbf{R}$ is the corresponding eigenvector of $\hat{\mathbf{R}}$.

B. Transmitter estimation

In order to find an estimator of the optimum transmitter (i.e., the top eigenvector of $\mathbf{H}^H\mathbf{H}$), consider the following construction based on the (known) training vectors, namely,

$$\hat{\mathbf{C}} = \frac{1}{N}\left(\mathbf{B}^{\#}\right)^H\mathbf{Y}^H\mathbf{Y}\mathbf{B}^{\#}, \tag{5}$$

where $(\cdot)^{\#}$ denotes the Moore-Penrose pseudoinverse, i.e., $\mathbf{B}^{\#} = \mathbf{B}^H\left(\mathbf{B}\mathbf{B}^H\right)^{-1}$. Indeed, note that, as N goes to infinity, almost surely, $\hat{\mathbf{C}} \to \mathbf{C}$, where

$$\mathbf{C} = \mathbb{E}\left[\hat{\mathbf{C}}\right] = \frac{1}{N}\mathbf{H}^H\mathbf{H} + \frac{MK}{EN}\mathbf{I}_K.$$

The following section provides an overview of the existing results concerning the asymptotic behavior of the eigenvectors of sample covariance matrices. In particular, we present the main mathematical tool of this paper, namely a convergence result characterizing the limit of the sample eigenvectors of information-plus-noise covariance matrices.

III. Asymptotic eigenvectors of signal-plus-noise sample covariance matrices

Due to the relevance of the eigenvalue spectrum of certain covariance matrix models in statistical signal processing and wireless communications, the theory of the spectral analysis of large-dimensional random matrices, or random matrix theory (RMT), has proved very useful in the tasks of both performance analysis and system design. For a monograph exposition on the subject, we refer the reader to [14]. Indeed, the characterization of the asymptotic behavior of the eigenvalues of certain random matrix models is of unquestionable practical interest, as it can be drawn from the vast engineering literature based on RMT results. However, in many problems, the study of an objective function is required that depends upon not only the eigenvalues but also the eigenvectors of the random matrix model.

While there are many results in the RMT literature about the eigenvalues of random matrices of increasing dimensions, not much has been reported about the asymptotic behavior of their eigensubspaces since Silverstein's work in [15] (see also references therein to his earlier contributions on the topic). In particular, consider the matrix $\mathbf{B} = \mathbf{X}\mathbf{X}^H$, with $\mathbf{X}$ being an $M \times N$ random matrix such that the entries of $\sqrt{N}\mathbf{X}$ are i.i.d. complex random variables with mean zero, variance one and finite fourth-order moment. Let $\mathbf{B} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ be the eigendecomposition of the previous matrix. Then, building on the fact that the matrix of eigenvectors of Wishart matrices follows the Haar distribution, i.e., the uniform distribution over the group of unitary matrices, Silverstein showed that, for any nonrandom vector $\mathbf{x}$ of appropriate dimensions, whose entries are either $-1/\sqrt{N}$ or $+1/\sqrt{N}$, the random vector $\mathbf{a} = \mathbf{U}^H\mathbf{x}$ is asymptotically isotropic (in [16] the same is proved for a random vector $\mathbf{x}$ with i.i.d. entries independent of $\mathbf{U}$). Furthermore, in order to study sample covariance matrices of the form $\mathbf{B} = \mathbf{R}^{1/2}\mathbf{X}\mathbf{X}^H\mathbf{R}^{1/2}$, where $\mathbf{R}^{1/2}$ is the positive square-root of an $M \times M$ Hermitian positive definite matrix $\mathbf{R}$ with uniformly bounded spectral norm, the following empirical distribution function was considered in [17], namely,

$$H_{\mathbf{B}}^{M}\left(\lambda\right) = \sum_{m=1}^{M}\left|a_m\right|^2\mathcal{I}_{\left(\lambda_m(\mathbf{B})\leq\lambda\right)},$$

where $\mathcal{I}_{\Omega}$ denotes the indicator function over the set Ω and a_m is the mth entry of the vector $\mathbf{a}$. Clearly, $H_{\mathbf{B}}^{M}$ is a random probability distribution function with Stieltjes transform given by

$$m_H\left(z\right) = \int_{\mathbb{R}}\frac{dH_{\mathbf{B}}^{M}\left(\lambda\right)}{\lambda - z} = \mathbf{x}^H\left(\mathbf{B} - z\mathbf{I}_M\right)^{-1}\mathbf{x}. \tag{6}$$

In this context, from the connection between vague convergence of distributions and pointwise convergence of Stieltjes transforms, almost sure convergence of the (random) distribution function $H_{\mathbf{B}}^{M}$ can be established by showing convergence of $m_H\left(z\right)$. In [18], an asymptotic deterministic equivalent of the Stieltjes transform in (6) was proposed for the more general case $\mathbf{B} = \mathbf{A} + \mathbf{R}^{1/2}\mathbf{X}\mathbf{T}\mathbf{X}^H\mathbf{R}^{1/2}$, where $\mathbf{A}$ is Hermitian and $\mathbf{T}$ real diagonal and positive definite, both having appropriate dimensions. Specifically, it is shown that $m_H\left(z\right)$ converges with probability one, for each $z \in \mathbb{C}^+$, as

$$\mathbf{x}^H\left(\mathbf{B} - z\mathbf{I}_M\right)^{-1}\mathbf{x} - \mathbf{x}^H\left(\mathbf{A} + x_M\mathbf{R} - z\mathbf{I}_M\right)^{-1}\mathbf{x} \to 0,$$

as $M, N \to \infty$ with $M/N \to c < +\infty$, where $x_M = x_M\left(e_M\right)$ is defined as

$$x_M = \frac{1}{N}\operatorname{Tr}\left[\mathbf{T}\left(\mathbf{I}_N + ce_M\mathbf{T}\right)^{-1}\right],$$

and $e_M = e_M\left(z\right)$ is the unique solution in $\mathbb{C}^+$ of the equation

$$e_M = \frac{1}{M}\operatorname{Tr}\left[\mathbf{R}\left(\mathbf{A} + x_M\mathbf{R} - z\mathbf{I}_M\right)^{-1}\right].$$

In this paper, we study the asymptotic behavior of $m_H(z)$ for information-plus-noise covariance matrix models. In particular, we present the following result:

Theorem 1: Let $\mathbf{X}$, $\mathbf{R}$ and $\mathbf{T}$ be defined as above, and consider the matrix $\mathbf{Y} = \mathbf{R}^{1/2}\mathbf{X}\mathbf{T}^{1/2}$. Furthermore, let $\mathbf{\Sigma} = \mathbf{Y} + \mathbf{A}$. Then, for each $z \in \mathbb{C}^+$, almost surely, as $M, N \to \infty$ with $M/N \to c < +\infty$,

$$\mathbf{x}^H \left(\mathbf{\Sigma}\mathbf{\Sigma}^H - z\mathbf{I}_M\right)^{-1}\mathbf{x} - \mathbf{x}^H \mathbf{\Upsilon}(z)\,\mathbf{x} \to 0,$$
$$\mathbf{x}^H \left(\mathbf{\Sigma}^H\mathbf{\Sigma} - z\mathbf{I}_N\right)^{-1}\mathbf{x} - \mathbf{x}^H \tilde{\mathbf{\Upsilon}}(z)\,\mathbf{x} \to 0,$$

where we have defined

$$\mathbf{\Upsilon}(z) = \left(-z\left(\mathbf{I} + \mathbf{R}\tilde{\delta}\right) + \mathbf{A}\left(\mathbf{I} + \mathbf{T}\delta\right)\mathbf{A}^H - z\mathbf{I}_M\right)^{-1},$$
$$\tilde{\mathbf{\Upsilon}}(z) = \left(-z\left(\mathbf{I} + \mathbf{T}\delta\right) + \mathbf{A}^H\left(\mathbf{I} + \mathbf{R}\tilde{\delta}\right)^{-1}\mathbf{A} - z\mathbf{I}_N\right)^{-1},$$

and $\delta = \delta(z)$ and $\tilde{\delta} = \tilde{\delta}(z)$ are the unique solution to the following system of equations:

$$\begin{cases} \delta(z) = \frac{1}{M}\operatorname{Tr}\left[\mathbf{R}\mathbf{\Upsilon}(z)\right] \\ \tilde{\delta}(z) = \frac{1}{M}\operatorname{Tr}\left[\mathbf{T}\tilde{\mathbf{\Upsilon}}(z)\right]. \end{cases}$$

Proof: The special case for $\mathbf{R} = \mathbf{I}_M$ and $\mathbf{T} = \sigma^2\mathbf{I}_N$, with σ^2 an arbitrary positive scalar was handled in [19, Proposition 1.1][2] as an extension of Theorem 1.1 in [20] on the asymptotic eigenvalue distribution of information-plus-noise covariance matrices. For the proof of Theorem 1, we follow the main stream in the proof of [21]. The proof is omitted due to lack of space (see [22]). ■

In the following section, we provide an analytical characterization of the performance of a training-based MIMO system under the realistic assumption of a training phase length comparable in magnitude with the system dimension.

IV. Large system performance analysis

In this section, we are interested in assessing the performance of a training-based MIMO system under a limited training budget. In particular, we will concentrate on the effect of a bounded ratio between training sample-size and number of degrees of freedom. In this work, in order to study the effect of the energy budget limitation as essentially due to a finite training sequence length, we assume a fixed power allocation across training beams given by $\|\mathbf{b}(n)\|^2 = 1$, $n = 1, \ldots, N$.

Using the principal eigenvectors of $\hat{\mathbf{R}}$ and $\hat{\mathbf{C}}$, denoted in the sequel by $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{v}}_1$, respectively, as the estimators of the optimum receiver and transmitter achieving the maximum SNR, namely given by SNR_1, we are interested in evaluating the performance loss incurred in practice by the use of the estimated solutions, i.e.,

$$\widehat{\mathsf{SNR}} = \left|\sum_{k=1}^{K\wedge M}\sqrt{\mathsf{SNR}_k}\;\hat{\mathbf{u}}_1^H\mathbf{u}_k\mathbf{v}_k^H\hat{\mathbf{v}}_1\right|^2, \quad (7)$$

where $\wedge$ denotes the minimum of the two quantities. Observe that the lack of an accurate estimate will contribute to the spread of power over the different orthogonal subchannels (similar to a linear programming suboptimal solution to the power allocation problem). In order to analytically characterize the performance measure in (7), it is enough to characterize the projection of the transceiver estimate obtained from a finite training sample-support onto the eigensubspaces spanned by the different right and left singular vectors. Indeed, for an unlimited training energy budget, as $N \to \infty$ (infinite training phase length), we clearly have $\hat{\mathbf{u}}_1^H\mathbf{u}_k\mathbf{v}_k^H\hat{\mathbf{v}}_1 \to 1\delta_{1,k}$, and, accordingly, $\widehat{\mathsf{SNR}} \to \mathsf{SNR}_1$.

The (finite-dimensional) statistical analysis of the quantity in (7) for finite system-size and limited training energy is rather intricate (for some related work based on a similar model and using finite RMT techniques see [23]). On the other hand, from an asymptotic characterization in the large-sample regime based on classical limiting results from the multivariante analysis of sample covariance matrices (see, e.g., [13]), no further insights can be gained for comparable training length and system size. Here, we focus on a large-system analysis of (7) and let not only the number of training samples (N), but also both the number transmit (K) and receive (M) antennas (i.e., the system dimension) go to infinity at a constant rate, defined by $\alpha = M/N$ and $\beta = K/N$. Since the previous asymptotic framework better matches realistic deployment conditions in practice, we may expect our results to more appropriately model the system performance in a practical setting characterized by a small number of training beams per degree-of-freedom.

Regarding the projections in the sum in (7) involving the estimates $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{v}}_1$, we may rely on the following procedure based on the power method for finding the eigenvalues and associated eigenvectors of an arbitrary Hermitian matrix. In particular, let us concentrate for instance on the top eigenvector of $\hat{\mathbf{R}}$ as the estimate of the optimal receiver. Then, consider the following quantity, namely,

$$\frac{\upsilon^H\left(\hat{\mathbf{R}} - \xi\mathbf{I}_M\right)^{-1}\mathbf{u}_k}{\left(\upsilon^H\left(\hat{\mathbf{R}} - \xi\mathbf{I}_M\right)^{-2}\upsilon\right)^{1/2}}, \quad (8)$$

where $\upsilon \in \mathbb{C}^K$ is any vector with a non-zero component in the direction of $\hat{\mathbf{u}}_1$ and $\xi = \lambda_1\left(\hat{\mathbf{R}}\right) + \epsilon$, with $\lambda_1\left(\hat{\mathbf{R}}\right)$ being the maximum eigenvalue of $\hat{\mathbf{R}}$ and ϵ being a small strictly positive constant. Indeed, $\hat{\mathbf{u}}_1^H\mathbf{u}_k$ can be arbitrarily well approximated by the expression in (8) for an arbitrarily small $\epsilon > 0$ (this follows from a limiting argument by letting ϵ vanish). For the purpose of analysis, we can use $\upsilon = \mathbf{u}_k$ (as M, N go to infinity, with probability one, $\hat{\mathbf{u}}_1$ has a non-zero component in the direction of $\mathbf{u}_k$, for each k). Then, we finally have

$$\frac{\mathbf{u}_k^H\left(\hat{\mathbf{R}} - \xi\mathbf{I}_M\right)^{-1}\mathbf{u}_k}{\left(\mathbf{u}_k^H\left(\hat{\mathbf{R}} - \xi\mathbf{I}_M\right)^{-2}\mathbf{u}_k\right)^{1/2}} \equiv \frac{\hat{U}_{1,k}(\xi)}{\hat{U}_{2,k}^{1/2}(\xi)}. \quad (9)$$

[2] We would like to thank Philippe Loubaton for his remarks on the correctness of the proof.

Note that an equivalent procedure follows for the optimal combiner at the transmitter side by replacing the sample covariance matrix $\hat{\mathbf{R}}$ with the matrix $\hat{\mathbf{C}}$ in (5), and $\mathbf{u}_k$ with $\mathbf{v}_k$. In particular, using the previous procedure, an arbitrarily accurate approximation of the SNR estimate in (7) can be obtained as

$$\widehat{\mathsf{SNR}}\left(\xi_r,\xi_c\right)=\left|\sum_{k=1}^{K\wedge M}\sqrt{\mathsf{SNR}_k}\frac{\hat{U}_{1,k}\left(\xi_r\right)}{\hat{U}_{2,k}^{1/2}\left(\xi_r\right)}\frac{\hat{V}_{1,k}\left(\xi_c\right)}{\hat{V}_{2,k}^{1/2}\left(\xi_c\right)}\right|^2,$$

where $\hat{V}_{1,k}\left(\xi\right)=\mathbf{v}_k^H\left(\hat{\mathbf{C}}-\xi\mathbf{I}_K\right)^{-1}\mathbf{v}_k$ and $\hat{V}_{2,k}\left(\xi\right)=\mathbf{v}_k^H\left(\hat{\mathbf{C}}-\xi\mathbf{I}_K\right)^{-2}\mathbf{v}_k$, and $\xi_r=\lambda_1\left(\hat{\mathbf{R}}\right)+\epsilon_r$ and $\xi_c=\lambda_1\left(\hat{\mathbf{C}}\right)+\epsilon_c$, with ϵ_r and ϵ_c being two two arbitrarily small strictly positive constants.

For the purposes of validating the proposed analytical characterization, we consider a Rayleigh MIMO channel matrix with particularly low-rank, such that the highest eigenmode alone essentially characterizes the full diversity gain that can be achieved over the channel. Note that, apart from simplifying the numerical validation, such a scenario renders especially relevant the accurate analysis and estimation of the diversity gain achieved by a MIMO system. Thus, as an approximation of $\widehat{\mathsf{SNR}}\left(\xi_r,\xi_c\right)$, we consider

$$\widetilde{\mathsf{SNR}}\left(\xi_r,\xi_c\right)=\left|\sqrt{\mathsf{SNR}_1}\;\frac{\hat{U}_{1,1}\left(\xi_r\right)}{\hat{U}_{2,1}^{1/2}\left(\xi_r\right)}\frac{\hat{V}_{1,1}\left(\xi_c\right)}{\hat{V}_{2,1}^{1/2}\left(\xi_c\right)}\right|^2. \quad (10)$$

Our analysis builds upon the fact that the expression in (9) is given in terms of the resolvent of $\hat{\mathbf{R}}$. In particular, an asymptotic deterministic equivalent of the empirical performance measure in (10) can be provided by using the result on the convergence of Stieltjes transforms of the type in (6) given by Theorem 1. Concretely, we use the following[3]:

Corollary 1: In Theorem 1, let $\mathbf{x}=\mathbf{u}_1$, $\mathbf{R}=\mathbf{I}_M$, $\mathbf{T}=\mathbf{I}_N$, $\mathbf{X}=1/\sqrt{N}\mathbf{N}$ and $\mathbf{A}=1/\sqrt{N}\mathbf{HB}$, and fix $z=\xi$, with $\xi=\lambda_1\left(\hat{\mathbf{R}}\right)+\epsilon$, $\epsilon>0$. Then, as $M,N\to\infty$, $M/N\to c<+\infty$

$$\frac{\hat{U}_{1,1}\left(\xi\right)}{\hat{U}_{2,1}^{1/2}\left(\xi\right)}\asymp\frac{\bar{U}_{1,1}\left(\xi\right)}{\bar{U}_{2,1}^{1/2}\left(\xi\right)},$$

with

$$\bar{U}_{1,1}\left(\xi\right)=\mathbf{u}_1^H\mathbf{\Psi}\left(z\right)\mathbf{u}_1,\ \bar{U}_{2,1}\left(\xi\right)=\frac{\partial}{\partial z}\left\{\mathbf{u}_1^H\mathbf{\Psi}\left(z\right)\mathbf{u}_1\right\}\Big|_{z=\xi},$$

where now

$$\mathbf{\Psi}\left(z\right)=\left(\left(1+\delta\left(\xi\right)\right)E/K\mathbf{HH}^H-z\left(1+\tilde{\delta}\left(\xi\right)\right)-z\mathbf{I}_M\right)^{-1},$$

$$\tilde{\mathbf{\Psi}}\left(z\right)=\left(\left(1+\tilde{\delta}\left(z\right)\right)\mathbf{B}^H\mathbf{H}^H\mathbf{HB}-z\left(1+\delta\left(z\right)\right)-z\mathbf{I}_N\right)^{-1},$$

and $\delta\left(z\right)$ and $\tilde{\delta}\left(z\right)$ are the unique solution to the following system of equations:

$$\begin{cases}\delta\left(z\right)=\frac{1}{M}\operatorname{Tr}\left[\mathbf{\Psi}\left(z\right)\right]\\ \tilde{\delta}\left(z\right)=\frac{1}{M}\operatorname{Tr}\left[\tilde{\mathbf{\Psi}}\left(z\right)\right].\end{cases}$$

[3]Given two random variables a,b, $a\asymp b$ denotes both quantities are asymptotic equivalents, i.e., $|a-b|\to 0$, almost surely.

Finally, based on the previous corollary, we have the following asymptotic deterministic equivalent for the approximation in (10) of the SNR in (7), namely,

Proposition 1: Under the previous statistical assumptions, for SNR_k uniformly bounded for all k, as $M,N\to\infty$, $M/N\to c<+\infty$,

$$\widetilde{\mathsf{SNR}}\left(\xi_r,\xi_c\right)\asymp\left|\sqrt{\mathsf{SNR}_1}\;\frac{\bar{U}_{1,1}\left(\xi_r\right)}{\bar{U}_{2,1}^{1/2}\left(\xi_r\right)}\frac{\bar{V}_{1,1}\left(\xi_c\right)}{\bar{V}_{2,1}^{1/2}\left(\xi_c\right)}\right|^2, \quad (11)$$

where $\bar{V}_{1,1}$ and $\bar{V}_{2,1}$ are defined equivalently to $\bar{U}_{1,1}$ and $\bar{U}_{2,1}$, respectively, for the covariance matrix $\hat{\mathbf{C}}$.

Proof: It is enough to show that the following quantity vanishes almost surely, as $M,N\to\infty$, $M/N\to c<+\infty$, namely,

$$\frac{\hat{U}_{1,1}\left(\xi_r\right)}{\hat{U}_{2,1}^{1/2}\left(\xi_r\right)}\frac{\hat{V}_{1,1}\left(\xi_c\right)}{\hat{V}_{2,1}^{1/2}\left(\xi_c\right)}-\frac{\bar{U}_{1,1}\left(\xi_r\right)}{\bar{U}_{2,1}^{1/2}\left(\xi_r\right)}\frac{\bar{V}_{1,1}\left(\xi_c\right)}{\bar{V}_{2,1}^{1/2}\left(\xi_c\right)},$$

or, equivalently, the almost surely convergence to zero of

$$\frac{\bar{V}_{1,1}\left(\xi_c\right)}{\bar{V}_{2,1}^{1/2}\left(\xi_c\right)}\left(\frac{\hat{U}_{1,1}\left(\xi_r\right)}{\hat{U}_{2,1}^{1/2}\left(\xi_r\right)}-\frac{\bar{U}_{1,1}\left(\xi_r\right)}{\bar{U}_{2,1}^{1/2}\left(\xi_r\right)}\right), \quad (12)$$

and

$$\frac{\hat{U}_{1,1}\left(\xi_r\right)}{\hat{U}_{2,1}^{1/2}\left(\xi_r\right)}\left(\frac{\hat{V}_{1,1}\left(\xi_c\right)}{\hat{V}_{2,1}^{1/2}\left(\xi_c\right)}-\frac{\bar{V}_{1,1}\left(\xi_c\right)}{\bar{V}_{2,1}^{1/2}\left(\xi_c\right)}\right). \quad (13)$$

Indeed, since all quantities in (12) and (13) are uniformly bounded for all K,M,N, this can be readily proved using the Cauchy-Schwarz inequality to show that the moments of the random variables (12) and (13) are $O\left(N^{-p}\right)$, for $p\geq 2$, and then applying the Borel-Cantelli lemma. ■

In the following section, we numerically evaluate the accuracy of the deterministic asymptotic equivalent (11) in describing the transient regime of pilot-aided MIMO transceivers with limited training.

V. Numerical results

We numerically validate the analytical characterization of the transient SNR performance of a training-based MIMO system. Specifically, we assume that both optimal transmit beamformer and receive combiner are empirically estimated from a finite collection of channel observations during a training phase of given length as described in Section II. In particular, the empirical performance in terms of averaged received SNR conditioned to the available training samples, i.e., $\mathbb{E}\left[\mathsf{SNR}\left(N+1\right)|\,\mathbf{y}\left(1\right),\ldots,\mathbf{y}\left(N\right)\right]$, is compared with the large-system performance approximation provided in Section IV. Figure 1 shows both numerically simulated and theoretically approximated SNR performance for a MIMO system with $K=10$ transmit antennas and $M=8$ receive antennas, versus the length of the training phase (normalized by K). The noise variance is assumed to be one. A fairly good prediction of the SNR achieved by a training-based MIMO system can be appreciated.

The previous analytical characterization can be used in practice for the evaluation of the loss in SNR performance

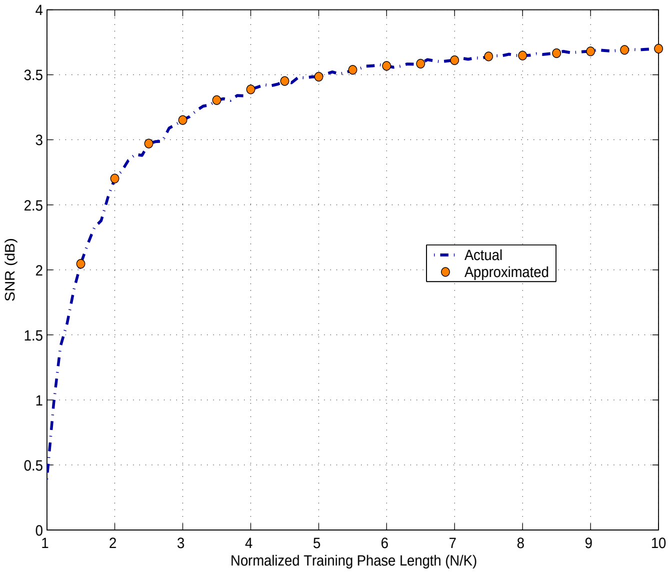

Fig. 1. Simulated and theoretically predicted SNR performance of MIMO system with empirically estimated optimal transceiver versus training phase length ($K = 10, M = 8$).

due to a finite number of training samples comparable to the system size, for a given type of channel. In particular, for instance, given the system dimensions (number of transmit and receive antennas) as well as the channel characterizing a certain MIMO wireless scenario, an estimate of the minimum number of training samples can be obtained such that a required SNR level is guaranteed.

VI. Conclusions

We have presented an analytical characterization of the transient regime of a training-based MIMO system exploiting the full diversity order of an arbitrarily correlated MIMO fading channel via optimal beamforming and combining. Since no channel state information is in practice available at either the transmitter or the receiver side, the design of the optimal transmit beamformer and receive combiner is most often based on a finite collection of samples observed during a training phase. If the length of the training sequence is comparable in magnitude to the system size, the MIMO system performance can be expected to suffer from a considerable degradation. While the finite-size statistical analysis of the problem is rather involved, a characterization based on the limiting behavior in the large-sample asymptotic regime does not provide any insight into the transient performance. In order to shed some light on the actual performance under practical conditions, we have proposed a large-system SNR performance analysis that builds upon new asymptotic convergence results concerning the eigenvectors of large-dimensional information-plus-noise covariance matrices. A power-iteration-based approach allows for a limiting description of the projection of the sample principal eigenvector onto the true principal eigenspaces. The proposed method is numerically validated in the context of a typical application of training-based MIMO systems.

References

[1] D. J. Love, R. W. H. Jr., W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Communications Magazine*, vol. 42, no. 10, pp. 54–59, October 2004.

[2] A. Goldsmith, S. A. 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.

[3] B. Hassibi and B. M. 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 *Proceedings of the 37th Annual Allerton Conference on Communications, Control, and Computing*, September 1999.

[5] T. L. Marzetta and B. M. 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. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[7] F. Rubio, D. Guo, M. L. Honig, and X. Mestre, "On optimal training and beamforming in uncorrelated MIMO systems with feedback," in *Proceedings of the 42nd Annual Conference on Information Sciences and Systems*, March 2008.

[8] W. Xiao and M. L. Honig, "Large system transient analysis of adaptive least squares filtering," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2447–2474, July 2005.

[9] L. R. Welch, "Lower bounds on the maximum cross correlation of signals," *IEEE Transactions on Information Theory*, vol. 20, pp. 397–399, May 1974.

[10] S. Verdú, *Multiuser Detection*. NY: Cambridge University Press, 1998.

[11] J. Massey and T. Mittelholzer, "Welch's bound and sequence sets for code-division multiple access systems," in *Sequences II, Methods in Communications, Security and Computer Science*, R. Capocelli, A. D. Santis, and U. Vaccaro, Eds. New York: Springer-Verlag, 1993, pp. 200–266.

[12] M. Rupf and J. L. Massey, "Optimum sequence multisets for synchronous code-division multiple-access channels," *IEEE Transactions on Information Theory*, vol. 40, pp. 1261–1266, July 1994.

[13] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. Wiley, 2003.

[14] A. Tulino and S. Verdú, "Random matrix theory and wireless communications," in *Foundations and Trends in Communications and Information Theory*. NOW Publishers, June 2004, vol. 1.

[15] J. W. Silverstein, "Weak convergence of random functions defined by the eigenvectors of sample covariance matrices," *Annals of Probability*, vol. 18, no. 3, pp. 1174–1194, August 1990.

[16] D. N. C. Tse and O. Zeitouni, "Linear multiuser receivers in random environment," *IEEE Transactions on Information Theory*, vol. 46, no. 1, pp. 171–188, January 2000.

[17] Z. Bai, G. Pan, and B. Miao, "On asymptotics of eigenvectors of large sample covariance matrix," *Annals of Probability*, vol. 35, no. 4, pp. 1532–1572, 2007.

[18] F. Rubio and X. Mestre, "Spectral convergence for a general class of random matrices," *Submitted to Statistics and Probability Letters*, 2009.

[19] F. Rubio, "Generalized consistent estimation in arbitrarily high dimensional signal processing," Ph.D. dissertation, Universitat Politècnica de Catalunya, Barcelona, April 2008.

[20] R. B. Dozier and J. W. Silverstein, "On the empirical distribution of eigenvalues of large dimensional information-plus-noise type matrices," *Journal of Multivariate Analysis*, vol. 98, no. 4, pp. 678–694, April 2007.

[21] W. Hachem, P. Loubaton, and J. Najim, "Deterministic equivalents for certain functionals of large random matrices," *Annals of Applied Probability*, vol. 17, no. 3, pp. 875–930, 2007.

[22] F. Rubio, D. Guo, M. L. Honig, and X. Mestre, "MIMO beamforming with limited training: A large system analysis," *In preparation*, 2009.

[23] E. K. S. Au, S. Jin, M. R. McKay, W. H. Mow, X. Gao, and I. B. Collings, "Analytical performance of MIMO-SVD systems in Ricean fading channels with channel estimation error and feedback delay," *IEEE Transactions on Wireless Communications*, vol. 7, no. 4, pp. 1315–1325, April 2008.