ABSTRACT

Title of dissertation:	OPTIMUM TRANSMIT STRATEGIES FOR GAUSSIAN MULTI-USER MIMO SYSTEMS WITH PARTIAL CSI AND NOISY CHANNEL ESTIMATION
	Alkan Soysal, Doctor of Philosophy, 2008
Dissertation directed by:	Professor Şennur Ulukuş Department of Electrical and Computer Engineering

Multiple antenna wireless communications systems are known to provide very large data rates, when perfect channel state information (CSI) is available at the receiver and the transmitter. Availability of perfect CSI at the receiver requires the receiver to perform a noise-free, multi-dimensional channel estimation, without using communication resources. Similarly, availability of perfect and instantaneous CSI at the transmitter requires a feedback scheme that sends the estimated CSI to the transmitter in its entirety and error-free. However, in practice, any channel estimation is noisy and uses system resources, and any feedback scheme is limited.

This thesis is devoted to the study of the effects of noisy channel estimation at the receiver and partial CSI at the transmitters on the optimum transmit strategies for Gaussian multi-input multi-output (MIMO) systems. The main focus of the thesis is on achievable rate maximization problems, solutions of which give the optimum resource allocation and channel estimation schemes for single-user and multi-user MIMO systems.

In the first part of the thesis, we focus on the effects of having non-perfect CSI at the transmitter side when the receiver is assumed to estimate the channel perfectly. We consider the capacity of a point-to-point channel and the sum-capacity of a MIMO multiple access channel (MAC). We analyze both the single-user and the multi-user MIMO systems from three different viewpoints. First, we consider a finite-sized system, and find the optimum transmit directions, and optimum power allocations along these directions, as well as beamforming optimality conditions. Second, we analyze the effects of increasing the number of users in the system, and show that the region where beamforming is optimal gets larger with the increasing number of users. Third, we consider the asymptotic case where the number of users is large, and show that beamforming is always optimal.

In the second part of the thesis, we consider the effects of channel estimation error at the receiver when partial CSI, in the form of covariance feedback, is available at the transmitter side. We solve the trade-off between estimating the channel better and increasing the achievable data rate. We consider a block fading MIMO channel, 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. In both single-user and multi-user cases, 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, we 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.

Optimum Transmit Strategies for Gaussian Multi-user MIMO Systems with Partial CSI and Noisy Channel Estimation

by

Alkan Soysal

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:

Professor Şennur Ulukuş, Chair/Advisor
Professor Prakash Narayan
Professor Armand Makowski
Professor T. Charles Clancy
Professor Carlos Berenstein

DEDICATION

To my sunshine Sanem, and princess Kayla.

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Professor Şennur Ulukuş for giving me the opportunity to work with her, and generously offering guidance throughout my Ph.D. studies. Without her knowledge and perfectionism, this dissertation could not have been written.

I would like to thank Professors Prakash Narayan, Armand Makowski, T. Charles Clancy, and Carlos Berenstein for reading my thesis, taking part in my PhD dissertation defense committee, and providing me with valuable comments and suggestions.

I have been very fortunate to meet many friends at the University of Maryland. Starting from the first day I set foot in the U.S., Tuna Güven, Tolga Girici and Onur Kaya have helped me in so many ways, and I am grateful for that. Talking with Anna Pantelidou, when we both got bored in the office was a great way to relax. I would also like to thank all of my current and past friends in Ulukuş group for the lengthy discussions we had during weekly meetings. I apologize to all of the friends that I have inadvertently left out who made my graduate study memorable.

Very special thanks go to my little princess Kayla. She has been so nice, cute and patient, which encouraged me to finish up this thesis. I apologize to you for the time I have stolen from you. You were my motivation.

During my studies, a force pushed me forward in every occasion, a hand offered itself whenever I was down, a shoulder whenever I cried, a friend whenever I needed,

a genuine smile greeted me after solving every problem, a partner joined me during dances, a companion to all my struggles, you my love are the bright sunshine giving me life. This is to you, Sanem. I owe you not only this thesis but everything I am happy for. Without your endless support, unconditional love, continuous encouragement, and tremendous patience, this life would be miserable.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

Over the last decade, the popularity of wireless applications has risen tremendously, and there is an ever increasing demand for higher data transmission rates. This demand stimulated a significant amount of research on wireless communications. Wireless communications is particularly challenging due to its unique characteristics such as random fluctuations in the channel and multi-user interference. In addition, in most of the future wireless systems, for example in the next generation cellular networks and wireless local area networks, the use of multiple antennas at both the transmitters and the receivers is proposed in order to achieve higher data rates. This adds another dimension to the already challenging problem of designing wireless systems with high data rates.

Achievable rates in a wireless communication system depend on how random fluctuations in the channel, which is called *fading*, and multi-user interference are handled. When fading is considered, achievable rates depend crucially on how well the channel state is estimated at the receiver and how much of the channel state knowledge is available at the transmitters. The channel state information (CSI) is observed only by the receiver, which can estimate it and feed the estimated CSI back to the transmitter. Theoretically, by using the perfect channel knowledge for signal detection at the receiver and for channel adaptive transmission at the transmitter, one can obtain the highest possible data rates.

Single-antenna systems, with perfect CSI available at both the receiver and the transmitter, have been very well studied. In a single-user system, the optimum channel adaptive transmission scheme that achieves the infromation theoretic capacity is found to be water-filling in time by Goldsmith and Varaiya [7]. In a multiple access channel (MAC), Knopp and Humblet [21] found the sum-capacity achieving scheme, and the entire capacity region was found by Tse and Hanly [43]. For single-user multi-input multi-output (MIMO) systems with perfect CSI available at both the receiver and the transmitter, Telatar reported the first capacity results [42], which can be identified as spatial water-filling, i.e., allocating power over the spatial channel dimensions that are created by the use of multiple antennas. In a MIMO-MAC, sum-capacity achieving iterative water-filling algorithm is proposed by Yu *et. al.* [49].

In additive white Gaussian noise channels, when perfect CSI is available at the receiver, the aforementioned capacity results are all achieved with Gaussian input signaling. In single-antenna systems, when the channel is fading, the variance of the Gaussian input signal is adapted to the realization of the channel. In Gaussian MIMO channels, the optimum signaling that achieves the capacity is Gaussian as well, but this time, the optimum covariance matrix of the transmit vector needs to be chosen.

Finding the transmit covariance matrix, in turn, involves two components: finding the optimum transmit directions, which are the eigenvectors of the transmit covariance matrix and the optimum power allocation policies, which are the eigenvalues of the transmit covariance matrix.

In [7], Goldsmith and Varaiya showed that for single-antenna systems, the resulting capacity does not decrease significantly when perfect CSI is not available at the transmitter. However, the solution to the capacity maximization problem in a MIMO system differs depending on the amount of information available at the transmitter side. Therefore, a significant amount of research has been conducted for MIMO systems with different CSI models. There are four basic CSI models for the transmitter side: *i)* the transmitter side does not know the state of the fading channel, i.e., no CSI [5], [10], [42], *ii)* the transmitter side perfectly knows the state of the fading channel, i.e., perfect CSI [42], [49], *iii)* the transmitter side knows the statistics of the fading channel, i.e., partial CSI [3], [14], [46], and *iv)* the transmitter side knows the quantized version of the realization of the channel, i.e., limited CSI [16], [29].

Although with perfect CSI at the receiver, one can obtain very high rates, when the channel knowledge is not perfect, achievable rates decrease significantly. This decrease is especially pronounced when there are multiple channels to estimate and feedback, as in the case with multiple antennas. Moreover, measuring the CSI and feeding it back to the transmitter uses communication resources, which could otherwise be used for useful information transmission. One way of measuring the CSI is that the transmitters send known training sequences, from which the receivers measure the channels. The receivers, then, extract the information (according to the feedback model) from the estimated channel, and feed the extracted information back to the transmitters. This overall process of estimating and feeding back CSI uses up time, bandwidth and power.

Recently, motivated mostly by practical issues, systems with non-perfect CSI at the receiver received more attention, but the research in this area mostly focused on single-user communication systems. For systems with no CSI at the receiver, [1] considered a single-antenna scenario, and [27], [51] considered a multi-antenna scenario. When the CSI is estimated but noisy, the capacity and the corresponding optimum signaling scheme are not known. However, lower and upper bounds for the capacity are obtained in [20, 28, 47].

In spite of recent progress, multi-antenna systems with partial CSI at the transmitter side and noisy CSI at the receiver side are not yet well-understood. In this thesis, we focus on such problems in both single-user and multi-user MIMO fading wireless communication systems. In particular, we analyze the effects of partial CSI at the transmitter side, and noisy channel estimation at the receiver side on the optimum transmit strategies that maximize the achievable data rates in wireless MIMO communications.

In Chapter 2, we focus our attention to the effects of partial and no CSI at the transmitters by assuming that the receiver has perfect CSI. In the partial CSI model, the receiver collects the long term statistics of the channel, and feeds this information back to the transmitter. We assume that the statistics of the channel do not change. When the fading in the channel is assumed to be a Gaussian process, statistics of the channel reduce to the mean and covariance information of the channel. Therefore, in Chapter 2, we consider three different CSI models, namely, no CSI model, partial CSI with covariance feedback model, and partial CSI with mean feedback model. Since it is already known that the capacity achieving input distribution is Gaussian, our goal here is to find the optimum transmit covariance matrices that achieve the capacity in a single-user system, and the sum-capacity in a MAC system.

For the no CSI model, [42] showed that the optimum transmit covariance matrix is proportional to the identity matrix, which is full-rank. However, for the partial CSI model, the rank of the transmit covariance matrix is determined by the structure of the channel feedback matrix. For a single-user case, when the partial CSI is in the form of either the covariance or the mean matrix of the channel, [3], [14], [46] first found the eigenvectors of the optimum transmit covariance matrix, and then the conditions on the covariance or mean matrix eigenvalues that guarantee that the transmit covariance matrix is unit-rank, and therefore beamforming is optimal. In the first part of Chapter 2, we extend these results to a MAC system. We first find the eigenvectors of the optimum transmit covariance matrices of all users. Then, we identify the necessary and sufficient conditions for the optimality of beamforming for all users.

In the second part of Chapter 2, we consider the effects of increasing the number of users on the region of channel parameters where beamforming is optimal. Here, in the covariance feedback case, we prove that this region gets larger as new users are added to the system. In the mean feedback case, this result does not necessarily hold. Nevertheless, we see through simulations that as the number of users gets large enough, the region where beamforming is optimal grows with the addition of new users, in the mean feedback case as well. Motivated by these results and the result of [31] which says that beamforming is optimal asymptotically (with respect to the number of users) in a deterministic MIMO-MAC, we ask the question whether beamforming is unconditionally optimal asymptotically in our case as well, where the receiver has perfect CSI, but the transmitters have no or partial CSI. In the remaining part of Chapter 2, we show that, in an asymptotically large system, unit-rank transmit covariance matrices are optimal for all users for no CSI and partial CSI models.

In Chapter 2, we mainly focus on the optimality of beamforming in a MIMO-MAC system with partial CSI at the transmitters. When beamforming is optimal, i.e., the transmit covariance matrix is unit-rank, the optimum power allocation problem is automatically solved. Since, when there is only one non-zero eigenvalue of the transmit covariance matrix, it is optimum to allocate all of the available power to the eigen-direction corresponding to that sole non-zero eigenvalue. However, for some channel realizations, in a single-user MIMO or in a MIMO-MAC with finite number of users, the channel statistics might be such that beamforming may never be optimal. For such scenarios, the optimum power allocation policies, i.e., the eigenvalues of the transmit covariance matrix, need to be solved.

In a single-user MIMO system, when both the receiver and the transmitter have perfect CSI and the channel is fixed, [42] showed that the optimum power allocation policy is to water-fill over the singular values of the deterministic channel matrix. In a multi-user MIMO-MAC system, when both the receiver and the transmitters have perfect CSI and the channel is fixed, [49] showed that the the power allocation policy can be found using an iterative algorithm that updates the power allocation policy of one user at a time. When the channel is changing over time due to fading, and perfect and instantaneous CSI is known both at the receiver and at the transmitter side, these solutions extend to water-filling over both the antennas and the channel states in single-user [42], and multi-user [50] MIMO systems.

However, for the covariance feedback case, there is no closed form solution for the power allocation problem, and therefore efficient and globally convergent algorithms are needed in order to solve for the optimum eigenvalues of the transmit covariance matrices. References [17], [44, 45] proposed algorithms that solve this problem for a single-user MISO system, and for a single-user MIMO system, respectively. However, in both cases, the convergence proofs for these algorithms were not provided. In a MIMO-MAC scenario with partial CSI available at the transmitters, no algorithm was available to find the optimum eigenvalues in a multi-user setting.

In Chapter 3, first, we give an alternative derivation for the algorithm proposed in [44, 45] for a single-user MIMO system by enforcing the Karush-Kuhn-Tucker (KKT) optimality conditions at each iteration. We prove that the convergence point of this algorithm is unique and is equal to the optimum eigenvalue allocation. The proposed algorithm converges to this unique point starting from any point on the space of feasible eigenvalues. Next, we consider the multi-user version of the problem. In this case, the problem is to find the optimum eigenvalues of the transmit covariance matrices of all users that maximize the sum-rate of the MIMO-MAC system. We apply the single-user algorithm iteratively to reach the global optimum point. At any given iteration, the multi-user algorithm updates the eigenvalues of one user, using the algorithm proposed for the single-user case, assuming that the eigenvalues of the remaining users are fixed. The algorithm iterates over all users in a round-robin fashion. We prove that this algorithm converges to the unique global optimum power allocation for all users.

For the case where the transmitters have partial CSI and the receiver has perfect and instantaneous CSI, Chapters 2 and 3 provide a complete extension from single-user to multi-user systems with finite and infinite numbers of users, including the transient behavior of the system with increasing number of users. Although having completely analyzed the effects of partial CSI at the transmitter side, Chapters 2 and 3 do not consider the problem of having non-perfect CSI at the receiver side; this will be the focus of Chapter 4.

When we consider the effects of having noisy CSI at the receiver, how we obtain the noisy CSI becomes part of the problem. One way of obtaining the channel estimate is to use a training based channel estimation mechanism. In this case, the transmitter sends a known training signal to the receiver, and the receiver estimates the CSI using the output of the channel and the known training signal. The variance of the channel estimation error inversely affects the average signal-to-noise ratio (SNR), and therefore, decreases the achievable rate.

In a training based estimation process, a block fading scenario is generally assumed, where the channel remains constant for a block (T symbols), and changes to an independent and identically distributed (i.i.d.) realization at the end of the block. In order to estimate the channel, the receiver performs a linear minimum mean square error (MMSE) estimation using training symbols over T_t symbols. During the remaining $T_d = T - T_t$ symbols, data transmission occurs. Intuitively, a longer training phase will result in a better channel estimate and therefore a larger achievable rate during the data transmission phase, since the channel estimation error contributes to the effective noise. 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. In Chapter 4, we solve these trade-offs.

When the CSI at the receiver is not perfect, most of the research focuses on single-user systems. The capacity and the corresponding optimum signaling scheme for this case is not known. However, lower and upper bounds for the capacity can be obtained [20, 28, 47]. It is important to note that [20, 28, 47] 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 single-user multiple-antenna system with no CSI available at the transmitter, [9] considers optimizing the achievable rate as a function of both the training and the data transmission phases.

In Chapter 4, we first consider a single-user, block-fading, correlated MIMO channel with noisy channel estimation at the receiver, and partial CSI available at the transmitter. The partial CSI feedback that we consider is covariance feedback which we also considered in Chapters 2 and 3. 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 [9] 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.

The training phase is characterized by three parameters, namely, the training signal, the training sequence length and the training sequence power. Similarly, the data transmission phase is characterized by the data carrying input signal, data transmission length, and the data transmission power. Assuming that the receiver uses linear MMSE detection to estimate the channel during the training phase, we first choose the training signal that minimizes the MMSE. 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 the data transmission phase parameters.

In a multi-user setting, the amount of resources required to measure the channel and to feed the estimated channel back to the transmitter increases substantially. Therefore, it is especially important to find the optimum transmit strategies in a MIMO-MAC with channel estimation error. In the second part of Chapter 4, we extend our results for the single-user MIMO case to the MIMO-MAC case. Interestingly, we find that the training signals of the users should be orthogonal in time. At the end of Chapter 4, we also provide detailed simulation results that investigate the effects of the power constraint, coherence interval (block length), and channel covariance matrix on our results.

1.1 Contributions of the Thesis

In Chapter 2, for MIMO systems with partial CSI at the transmitters in the form of covariance and mean information, our contribution is to provide a complete extension from single-user to multi-user systems with finite and infinite numbers of users, including the transient behavior of the system with increasing number of users. In particular, we first find that the optimum transmit directions of each user are the eigenvectors of its own channel covariance or mean matrix. Then, we find the conditions under which beamforming is optimal for all users for both the covariance and mean feedback models. We show in the covariance feedback model that the region that is formed by these conditions gets larger when a new user is added to the system.

At the end of Chapter 2, we show that beamforming is always optimal asymptotically in the number of users for all three feedback models we consider. The results in this chapter are published in [35], [36], [40].

Beamforming can be regarded as a special case of a power allocation policy. In Chapter 3, we focus on the general case, and propose provably convergent iterative algorithms that find the optimum power allocation policies, i.e., the eigenvalues of the transmit covariance matrices, for both point-to-point and multiple access channels. These algorithms are based on enforcing the KKT conditions at each iteration. Our main contribution in Chapter 3 is the convergence proof of the proposed single-user algorithm. Convergence is shown using the monotonicity property of the update function and the instability of the solution points that do not satisfy the KKT conditions. The results in this chapter are published in [37], [38].

In Chapter 4, we investigate the effects of channel estimation error on the achievable rate of a single-user and the achievable sum-rate of a multi-user MIMO channel, when the transmitter side has partial CSI in the form of covariance feedback. In this chapter, we consider a block fading channel, where a transmission block is divided into training and data transmission phases. Our contributions provide a solution to the data-rate optimization problem jointly over the training and data transmission phases. In a single-user case, we first find the optimum training signal that minimizes the mean square error of the channel estimation. Then, we develop an algorithm that maximizes the achievable rate of the data transmission phase jointly in terms the training and data transmission parameters. In the second part of Chapter 4, we extend our contributions to a multi-user scenario and study the effects of the power constraint, coherence interval (block length), and channel covariance matrices, numerically. The results in this chapter are submitted for publication in [33], [34], [39], [41].

Chapter 2

Transmit Directions and the Optimality of Beamforming

The use of multiple antennas at both the transmitters and the receivers in wireless communications promises very large information rates. In Gaussian MIMO systems, when the receiver side has perfect CSI, the calculation of the information theoretic capacity boils down to finding the transmit covariance matrices. In this chapter, for a system with perfect CSI at the receiver and partial or no CSI at the transmitters, we analyze the optimum transmit covariance matrix structures both in point-to-point and multiple access channels.

In [42], Telatar showed that in a single-user system, when the transmitter does not know the state of the fading channel, the optimum transmit covariance matrix is proportional to the identity matrix, which is full-rank. In order to achieve the capacity, either vector coding or parallel processing of scalar codes is needed. As stated in [42], vector coding will result in lower probability of error but higher complexity as compared to parallel scalar coding, which already is very complex [5].

Beamforming is a scalar coding strategy in which the transmit covariance matrix is unit-rank. In beamforming, the symbol stream is coded and multiplied by different coefficients at each antenna before transmission. Since the available mature scalar codec technology can be used, beamforming is highly desirable. However, in the setting of [42], where there is no CSI at the transmitters and the aim is to achieve the ergodic capacity, the optimum transmit covariance matrix is full-rank, and therefore beamforming is not optimal.

Although beamforming is not optimal for the no CSI case, it is shown by [3], [14], [46] for single-sided correlation structure, and by [18] for double-sided correlation structure that beamforming is conditionally optimal, in a single-user setting, when the transmitter has the partial knowledge of the channel. For the covariance feedback case, the fact that the optimal transmit covariance matrix and the channel covariance matrix have the same eigenvectors was shown in [46] for a multi-input single-output (MISO) system, and in [14] for a MIMO system. The conditions on the channel covariance matrix that guarantee that the transmit covariance matrix is unit-rank, and therefore beamforming is optimal, are identified in [3], [14]. This result is analogous to identifying the conditions on the channel state space and the average power in classical water-filling that guarantee that only one channel is filled as a result of having either a low power constraint or one very strong channel. In [18], these conditions are generalized to the case where the receive antennas are also correlated. For the mean feedback case, the eigenvectors of the optimal transmit covariance matrix were shown to be the same as the eigenvectors of the channel mean matrix for a MISO system in [46] and for a MIMO system in [14]. Using this, the conditions on the channel mean matrix that guarantee that the transmit covariance matrix is unit-rank, and therefore beamforming is optimal, are identified in [14].

In this chapter, we consider the sum-capacity point of a *multi-user* MIMO multiple access capacity region with various assumptions on the CSI. In the first part of the chapter, we concentrate on a finite-sized system. We show that, if there is covariance or mean feedback information at the transmitters, all users should transmit in the direction of the eigenvectors of their *own* covariance or mean feedback matrices. Consequently, we show that the transmit directions of the users are independent of the presence of other users, and therefore that the users maintain their single-user transmit direction strategies even in a multi-user scenario. Then, we identify the necessary and sufficient conditions for the optimality of beamforming for all users. This result generalizes the single-user conditions of [3], [14] to a multi-user setting. In the case of covariance feedback, these conditions depend only on the first and second largest eigenvalues of the channel covariance matrix of each user, and they form a region in a space whose dimension is twice the number of users. If these conditions are satisfied, beamforming is optimal for all users. In the case of mean feedback, these conditions depend only on the sole non-zero eigenvalue of the unit-rank channel mean matrix of each user, and they form a region in a space whose dimension is equal to the number of users. Similarly, if these conditions are satisfied, beamforming is optimal for all users.

We, then, consider the effects of increasing the number of users on the region of channel parameters where beamforming is optimal. In the covariance feedback case, we prove that this region gets larger as new users are added to the system. Although adding users increases the overall complexity of the system, being able to beamform for a greater range of channel values decreases the complexity. In the mean feedback case, this result does not necessarily hold. Nevertheless, we see through simulations that as the number of users gets large enough, the region where beamforming is optimal grows with the addition of new users, in the mean feedback case as well. These results raise the question of whether the region where beamforming is optimal spans the entire parameter space as the number of users grows to infinity. Therefore, next, we analyze our problem from an asymptotically large system viewpoint.

The optimality of beamforming in a MIMO-MAC system where the channel is deterministic and fully known to the transmitters is investigated in [31], where it was shown that if the number of users is much larger than the number of receive antennas, then unit-rank transmission is optimal for almost all users. Motivated by our result described above and the result of [31] that beamforming is optimal asymptotically (with respect to the number of users) in a deterministic multi-user MIMO-MAC, we ask the question whether beamforming is unconditionally optimal asymptotically in our case as well, where the receiver has perfect CSI, but the transmitters have no or partial CSI. When there is no CSI at the transmitters, it is counter-intuitive to think that beamforming would be optimal. Confirming this intuition, [42] already showed that in a finite-sized multi-user system with no CSI at the transmitters, the optimum transmit covariance matrices are full-rank for all users. However, we show that, in an asymptotically large system, unit-rank transmit covariance matrices are optimal for all users. The beamforming scheme we use in this case is simpler than usual; it may be characterized as an arbitrary antenna selection scheme, where for each user, only one antenna is used for transmission and that antenna is chosen arbitrarily.

When the transmitters have partial CSI in the form of either covariance or mean feedback, we show that the asymptotic optimality of beamforming still holds. In these cases however, arbitrary antenna selection scheme is no longer optimal. In the covariance feedback setting, each user beamforms in the direction of the strongest eigenvector of its channel feedback covariance matrix. As opposed to a finite-sized system, where beamforming may or may not be optimal depending on the eigenvalues of the channel covariance matrices, we show here that for an asymptotically large system, beamforming is always optimal. In the mean feedback setting, each user beamforms in the direction of the eigenvector corresponding to the sole non-zero eigenvalue of its channel feedback mean matrix. Similar to the covariance feedback case, beamforming is optimal asymptotically irrespective of the values of the mean feedback information. Asymptotic analysis has been used in the literature before, e.g., by [8], [10], [25], where it yielded simple characterizations to complex systems. In our model, with multiple users, with multiple transmit and receive antennas, and with fading in the channel, the optimal transmit strategy turns out to be simple beamforming, when only the number of users goes to infinity.

In this chapter, our contributions are three-fold: the analysis of a system with a finite number of users, determining the effects of increasing the number of users, and the analysis of a large system. Considering all three points of view, this chapter provides a complete extension from single-user to multi-user systems with finite and infinite numbers of users, including the transient behavior of the system with increasing number of users, for MIMO systems with partial CSI at the transmitters in the form of covariance and mean information.

2.1 System Model

We consider a multiple access channel with multiple transmit antennas at every user and multiple receive antennas at the receiver. The channel between user k and the receiver is represented by a random matrix $\mathbf{H}_k$ with dimensions of $n_R \times n_T$, where n_R and n_T are the number of antennas at the receiver and at the transmitter, respectively[1]. The receiver has the perfect knowledge of the channel, while the transmitters have only the statistical model of the channel. Each transmitter sends a vector $\mathbf{x}_k$, and the received vector is

$$\mathbf{r} = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{x}_k + \mathbf{n} \tag{2.1}$$

where K is the number of users, $\mathbf{n}$ is a zero-mean, identity-covariance complex Gaussian vector, and the entries of $\mathbf{H}_k$ are complex Gaussian random variables. Let $\mathbf{Q}_k = E[\mathbf{x}_k \mathbf{x}_k^\dagger]$ be the transmit covariance matrix of user k, which has an average power constraint of P_k, $\text{tr}(\mathbf{Q}_k) \leq P_k$.

We investigate three different statistical models. The first one is the "no CSI"

[1] Although we consider the case where all transmitters have the same number of antennas, our results immediately extend to the cases where the transmitters have different number of antennas.

model in which the transmitters only know the distribution of the channel state while the parameters of the distribution are not known. In this case, the entries of $\mathbf{H}_k$ are i.i.d., zero-mean, unit-variance complex Gaussian random variables. This model is used in [5], [10], [42].

The second model is the "partial CSI with covariance feedback" model where each transmitter knows the channel covariance information of all transmitters, in addition to the distribution of the channel. In this model, there exists correlation between the signals transmitted by or received at different antenna elements. For each user, the channel is modeled as [4],

$$\mathbf{H}_k = \mathbf{\Phi}_k^{1/2} \mathbf{Z}_k \mathbf{\Sigma}_k^{1/2} \tag{2.2}$$

where the receive antenna correlation matrix, $\mathbf{\Phi}_k$, is the correlation between the signals transmitted by user k, and received at the n_R receive antennas of the receiver, and the transmit antenna correlation matrix, $\mathbf{\Sigma}_k$, is the correlation between the signals transmitted from the n_T transmit antennas of user k. While writing (2.2), we separately apply the single-user model in [4] to every single transmitter-receiver link. In this main part of this chapter, we will assume that the receiver (e.g., base station) does not have physical restrictions and therefore, there is sufficient spacing between the antenna elements on the receiver. If the minimum antenna spacing is sufficiently large, the correlation introduced by antenna element spacing is low enough that the fades associated with two different antenna elements can be considered independent[2] [15].

[2] We refer the reader to the Appendix, Section 2.6.2, for the extension of our results to the case

As a result, the receive antenna correlation matrix becomes the identity matrix, i.e., $\mathbf{\Phi}_k = \mathbf{I}$. We also assume that the signals transmitted by different antenna elements are correlated, because of the lack of scatterers around the transmitters. Now, the channel of user k is written as

$$\mathbf{H}_k = \mathbf{Z}_k \mathbf{\Sigma}_k^{1/2} \tag{2.3}$$

where the entries of $\mathbf{Z}_k$ are i.i.d., zero-mean, unit-variance complex Gaussian random variables. Similar covariance feedback models are used in [3], [14], [17], [46] in the single-user setting. From this point on, we will refer to matrix $\mathbf{\Sigma}_k$ as the channel covariance feedback matrix of user k.

The third model we investigate is the "partial CSI with mean feedback" model where each transmitter knows the channel mean information of all transmitters, in addition to the distribution of the channel. This model is used in [11], [13], [14], [23], [46]. In this model, the transmitters have line-of-sight component with the receiver. As a result, the entries of the channel matrix are independent with a non-zero mean. In this case, the channel of user k can be written as

$$\mathbf{H}_k = \mathbf{H}_{\mu_k} + \mathbf{Z}_k \tag{2.4}$$

where the entries of $\mathbf{Z}_k$ are i.i.d., zero-mean, unit-variance complex Gaussian random variables, and $\mathbf{H}_{\mu_k}$ is the mean information representing the line-of-sight component

where the channel has double-sided correlation structure, i.e., to the case where the signals arriving at the receiver are correlated as well.

of the channel. This Ricean channel is modeled to be of unit-rank [23], and therefore, the mean matrix takes the form

$$\mathbf{H}_{\mu_k} = \mathbf{a}_{R_k}\mathbf{a}_{T_k}^{\dagger} \tag{2.5}$$

where $\mathbf{a}_{R_k}$ and $\mathbf{a}_{T_k}$ are the array response vectors at the receiver and the transmitter, respectively. In this most general case of the mean feedback model, the optimization problem that arises in the sum-capacity calculation seems intractable. In order to simplify the mathematics and obtain a tractable optimization problem, we assume that the user signals arrive at the base station in-phase, i.e., $\mathbf{a}_{R_k} = \mathbf{a}_R$, for all k. This mathematical simplification models a physical system where the transmitters are far away from the receiver and are close to each other. This can occur if a set of closely located transmitters have a line-of-sight "openning" with the receiver. From this point on, we will refer to matrix $\mathbf{H}_{\mu_k}$ as the channel mean matrix of user k.

2.2 Finite System Analysis

The sum-capacity of a multi-user MIMO-MAC is given as,

$$C_{sum} = \max_{\substack{\mathrm{tr}(\mathbf{Q}_k)\le P_k, \mathbf{Q}_k \succeq \mathbf{0} \\ k=1\ldots K}} E\left[\log\left|\mathbf{I}_{n_R} + \sum_{k=1}^{K}\mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^{\dagger}\right|\right] \tag{2.6}$$

where $E[\cdot]$ is the expectation operator with respect to the channel matrices of all users conditioned on the covariance or mean feedback, $|\cdot|$ is the determinant operator, $\mathrm{tr}(\cdot)$

denotes the trace of a matrix, and $\mathbf{Q}_k \succeq \mathbf{0}$ denotes positive semi-definite $\mathbf{Q}_k$. In this section, we will find the optimum transmit directions of the users, and the region where beamforming is optimal for all users, under various assumptions on the CSI available at the transmitters, for a multi-user MIMO-MAC with a finite number of users.

For a single-user system with no CSI at the transmitter and identity channel covariance matrix, i.e., $\mathbf{\Sigma} = \mathbf{I}$, Telatar [42] showed that the capacity is achieved when the transmitter divides its power equally over its antennas, i.e., the optimal transmitter covariance matrix, $\mathbf{Q}$, is equal to $(P/n_T)\mathbf{I}$. Clearly, in this setting, beamforming is not optimal, as the transmit covariance matrix is full-rank. For the multi-user case, [42] defines a stacked channel matrix as $\hat{\mathbf{H}} = [\mathbf{H}_1, \cdots, \mathbf{H}_K]$ and writes the sum-capacity as

$$\begin{aligned} C_{sum} &= E\left[\log\left|\mathbf{I}_{n_R} + \frac{KP}{Kn_T}\hat{\mathbf{H}}\hat{\mathbf{H}}^\dagger\right|\right] \\ &= E\left[\log\left|\mathbf{I}_{n_R} + \frac{P}{n_T}\sum_{k=1}^{K}\mathbf{H}_k\mathbf{H}_k^\dagger\right|\right] \end{aligned} \tag{2.7}$$

This means that in the multi-user setting as well, the sum-capacity maximizing transmit covariance matrix for each user is proportional to identity, i.e., $\mathbf{Q}_k = (P/n_T)\mathbf{I}$, for all k. Therefore, it is clear that beamforming is not optimal for any user in a finite-sized multi-user system when the transmitters do not have any CSI.

2.2.1 Covariance Feedback at the Transmitters

Transmit Directions

In a single-user system with partial CSI in the form of channel covariance matrix at the transmitter, the capacity is no longer achieved by an identity transmit covariance matrix. In this case, the problem becomes that of choosing a transmit covariance matrix $\mathbf{Q}$, which is subject to a trace constraint representing the average transmit power constraint,

$$C = \max_{\text{tr}(\mathbf{Q}) \leq P, \mathbf{Q} \succeq \mathbf{0}} E\left[\log | \mathbf{I}_{n_R} + \mathbf{H}\mathbf{Q}\mathbf{H}^{\dagger}|\right] \tag{2.8}$$

The channel covariance matrix $\mathbf{\Sigma}$, which is known at the transmitter, and the transmit covariance matrix $\mathbf{Q}$ have the eigenvalue decompositions $\mathbf{\Sigma} = \mathbf{U}_{\Sigma}\mathbf{\Lambda}_{\Sigma}\mathbf{U}_{\Sigma}^{\dagger}$, and $\mathbf{Q} = \mathbf{U}_{Q}\mathbf{\Lambda}_{Q}\mathbf{U}_{Q}^{\dagger}$, respectively. Here, $\mathbf{\Lambda}_{\Sigma}$ and $\mathbf{\Lambda}_{Q}$ are the diagonal matrices of ordered eigenvalues of $\mathbf{\Sigma}$, and $\mathbf{Q}$, and $\mathbf{U}_{\Sigma}$, and $\mathbf{U}_{Q}$ are unitary matrices.

References [14] and [30] showed that the eigenvectors of the transmit covariance matrix must be equal to the eigenvectors of the channel covariance matrix, i.e., $\mathbf{U}_{Q} = \mathbf{U}_{\Sigma}$. References [3] and [14] showed that under certain conditions on the covariance feedback matrix $\mathbf{\Sigma}$, the power matrix $\mathbf{\Lambda}_{Q}$ has only one non-zero diagonal element, i.e., the optimal transmit covariance matrix is unit-rank, and therefore beamforming in the direction of the eigenvector corresponding to this non-zero eigenvalue, is optimal.

In this chapter, for a multi-user setting with a finite number of users, where there is covariance feedback at the transmitters, we prove that all users should transmit along the eigenvectors of their own channel covariance matrices, regardless of the power allocation scheme. This is stated in the following theorem.

Theorem 1 *Let $\mathbf{\Sigma}_k = \mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{\Sigma_k}\mathbf{U}_{\Sigma_k}^{\dagger}$ be the spectral decomposition of the channel covariance matrix of user k. Then, the optimum transmit covariance matrix $\mathbf{Q}_k$ of user k has the form $\mathbf{Q}_k = \mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{Q_k}\mathbf{U}_{\Sigma_k}^{\dagger}$, for all users.*

Proof: From (2.3), we have the following zero-mean, identity-covariance random channel matrix representation $\mathbf{Z}_k$ for user k,

$$\mathbf{Z}_k\mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{\Sigma_k}^{1/2}\mathbf{U}_{\Sigma_k}^{\dagger} = \mathbf{H}_k \tag{2.9}$$

Then, inserting (2.9) into (2.6), we obtain

$$\begin{aligned} C_{sum} &= \max_{\substack{\mathrm{tr}(\mathbf{Q}_k)\leq P_k \\ k=1,\ldots,K}} E\left[\log\left|\mathbf{I}_{n_R} + \sum_{k=1}^{K}\mathbf{Z}_k\mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{\Sigma_k}^{1/2}\mathbf{U}_{\Sigma_k}^{\dagger}\mathbf{Q}_k\mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{\Sigma_k}^{1/2}\mathbf{U}_{\Sigma_k}^{\dagger}\mathbf{Z}_k^{\dagger}\right|\right] && (2.10)\\ &= \max_{\substack{\mathrm{tr}(\mathbf{Q}_k)\leq P_k \\ k=1,\ldots,K}} E\left[\log\left|\mathbf{I}_{n_R} + \sum_{k=1}^{K}\mathbf{Z}_k\mathbf{\Lambda}_{\Sigma_k}^{1/2}\mathbf{U}_{\Sigma_k}^{\dagger}\mathbf{Q}_k\mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{\Sigma_k}^{1/2}\mathbf{Z}_k^{\dagger}\right|\right] && (2.11)\end{aligned}$$

where we used the fact that the random matrices $\{\mathbf{Z}_k\mathbf{U}_{\Sigma_k}\}_{k=1}^{K}$ and $\{\mathbf{Z}_k\}_{k=1}^{K}$ have the same joint distribution for zero-mean identity-covariance Gaussian $\{\mathbf{Z}_k\}_{k=1}^{K}$ and unitary $\{\mathbf{U}_{\Sigma_k}\}_{k=1}^{K}$. This is true, since we can write the joint distribution of $\{\mathbf{Z}_k\mathbf{U}_{\Sigma_k}\}_{k=1}^{K}$ as a multiplication of their marginal distributions due to their independence, and the marginal distribution of $\mathbf{Z}_k\mathbf{U}_{\Sigma_k}$ is the same as the marginal distribution of $\mathbf{Z}_k$ [42]. We may spectrally decompose the expression sandwiched between $\mathbf{Z}_k$ and its conjugate transpose in (2.11) as

$$\mathbf{\Lambda}_{\Sigma_k}^{1/2}\mathbf{U}_{\Sigma_k}^{\dagger}\mathbf{Q}_k\mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{\Sigma_k}^{1/2} = \mathbf{U}_k\mathbf{\Lambda}_k\mathbf{U}_k^{\dagger} \tag{2.12}$$

where $\mathbf{\Lambda}_k$ is a diagonal matrix with ordered components such that $\lambda_{k1} \geq \lambda_{k2} \geq \cdots \geq \lambda_{kn_T}$. The optimization problem in (2.11) may now be written as

$$\begin{aligned} C_{sum} &= \max_{\substack{\mathrm{tr}(\mathbf{Q}_k)\leq P_k \\ k=1,\ldots,K}} E\left[\log\left|\mathbf{I}_{n_R} + \sum_{k=1}^{K}\mathbf{Z}_k\mathbf{U}_k\mathbf{\Lambda}_k(\mathbf{Z}_k\mathbf{U}_k)^{\dagger}\right|\right] \qquad (2.13)\\ &= \max_{\substack{\mathrm{tr}(\mathbf{Q}_k)\leq P_k \\ k=1,\ldots,K}} E\left[\log\left|\mathbf{I}_{n_R} + \sum_{k=1}^{K}\mathbf{Z}_k\mathbf{\Lambda}_k\mathbf{Z}_k^{\dagger}\right|\right] \qquad (2.14)\end{aligned}$$

where we again used the fact that the random matrices $\{\mathbf{Z}_k\mathbf{U}_k\}_{k=1}^{K}$ and $\{\mathbf{Z}_k\}_{k=1}^{K}$ have the same joint distribution. Using (2.12), the trace constraint on $\mathbf{Q}_k$ can be expressed as

$$\begin{aligned} \mathrm{tr}(\mathbf{Q}_k) &= \mathrm{tr}(\mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{\Sigma_k}^{-1/2}\mathbf{U}_k\mathbf{\Lambda}_k\mathbf{U}_k^{\dagger}\mathbf{\Lambda}_{\Sigma_k}^{-1/2}\mathbf{U}_{\Sigma_k}^{\dagger}) \qquad (2.15)\\ &= \mathrm{tr}(\mathbf{U}_k^{\dagger}\mathbf{\Lambda}_{\Sigma_k}^{-1}\mathbf{U}_k\mathbf{\Lambda}_k) \qquad (2.16)\end{aligned}$$

where the second equality follows from the identity $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$. Note that the optimization in (2.14), (2.16) is over $\mathbf{U}_k$ and $\mathbf{\Lambda}_k$, and the objective function does not involve $\mathbf{U}_k$. Therefore, we can insert any feasible $\mathbf{U}_k$ from the constraint set, and perform the optimization only over $\mathbf{\Lambda}_k$. In order to find a feasible $\mathbf{U}_k$, we examine the trace constraint in (2.16). From [26, Theorem 9.H.1.h, page 249], $\mathrm{tr}(\mathbf{\Lambda}_{\Sigma_k}^{-1}\mathbf{\Lambda}_k) \leq \mathrm{tr}(\mathbf{U}_k^{\dagger}\mathbf{\Lambda}_{\Sigma_k}^{-1}\mathbf{U}_k\mathbf{\Lambda}_k) \leq P_k$, for all unitary $\mathbf{U}_k$. This means that, $\mathbf{U}_k = \mathbf{I}$

choice is feasible. Then, using $\mathbf{U}_k = \mathbf{I}$, from (2.12), we have the desired result:

$$\mathbf{Q}_k = \mathbf{U}_{\Sigma_k} \mathbf{\Lambda}_{\Sigma_k}^{-1} \mathbf{\Lambda}_k \mathbf{U}_{\Sigma_k}^{\dagger} \tag{2.17}$$

with $\mathbf{\Lambda}_{Q_k} = \mathbf{\Lambda}_{\Sigma_k}^{-1} \mathbf{\Lambda}_k$. $\square$

Using Theorem 1, we can write the optimization problem in (2.10) as,

$$\begin{aligned} C_{sum} &= \max_{\substack{\mathrm{tr}(\mathbf{\Lambda}_{Q_k}) \le P_k, \mathbf{\Lambda}_{Q_k} \succeq \mathbf{0} \\ k=1,\ldots,K}} E\left[\log \left| \mathbf{I}_{n_R} + \sum_{k=1}^{K} \mathbf{Z}_k \mathbf{\Lambda}_{Q_k} \mathbf{\Lambda}_{\Sigma_k} \mathbf{Z}_k^{\dagger} \right| \right] && (2.18) \\ &= \max_{\substack{\sum_{i=1}^{n_T} \lambda_{ki}^{Q} \le P_k, \lambda_{ki}^{Q} \ge 0 \\ k=1,\ldots,K}} E\left[\log \left| \mathbf{I}_{n_R} + \sum_{k=1}^{K} \sum_{i=1}^{n_T} \lambda_{ki}^{Q} \lambda_{ki}^{\Sigma} \mathbf{z}_{ki} \mathbf{z}_{ki}^{\dagger} \right| \right] && (2.19) \end{aligned}$$

where $\mathbf{z}_{ki}$ is the i^{th} column of $\mathbf{Z}_k$, i.e., $\{\mathbf{z}_{ki}, k = 1, \ldots, K, i = 1, \ldots, n_T\}$ is a set of $n_R \times 1$ dimensional i.i.d., zero-mean, identity-covariance Gaussian random vectors.

In a MIMO system, a transmit strategy is a combination of a transmit direction strategy and a transmit power allocation strategy. A result of Theorem 1 is that the optimal multi-user transmit direction strategies are decoupled into a set of single-user transmit direction strategies. However, in general, this is not true for the optimal transmit power allocation strategies. The amount of power each user allocates in each direction depends on both the transmit directions and the power allocations of other users, which we show in Chapter 3. Because of this, finding the conditions under which beamforming is optimal becomes even more critical in the multi-user case. When beamforming is optimal, the optimal transmit power allocation strategy for each user reduces to allocating all of its power to its strongest eigen-direction, and this strategy does not require the user to know the channel covariance matrices of the other users.

Conditions for the Optimality of Beamforming

In this section, we identify the conditions for the optimality of beamforming in a multi-user system with a finite number of users. References [3] and [14] found these conditions in a single-user system. For a single-user system, let λ_1^{Σ} and λ_2^{Σ} denote the largest and second largest eigenvalues of the channel covariance matrix $\mathbf{\Sigma}$, respectively. Then, the necessary and sufficient condition for the optimality of beamforming is [14]:

$$P\lambda_2^{\Sigma} < \frac{1 - E\left[\frac{1}{1+P\lambda_1^{\Sigma}\mathbf{z}^T\mathbf{z}}\right]}{n_R - 1 + E\left[\frac{1}{1+P\lambda_1^{\Sigma}\mathbf{z}^T\mathbf{z}}\right]} \tag{2.20}$$

where $\mathbf{z}$ is an $n_R \times 1$ dimensional Gaussian random vector with zero-mean and identity-covariance. In this chapter, we find the necessary and sufficient conditions for the optimality of beamforming for all users in a multi-user setting. Inserting $K = 1$ in our results would reduce them to (2.20). In our results, the number of conditions equals the number of users. The condition corresponding to user k depends on the two largest eigenvalues of the channel covariance matrix of that user, and the largest eigenvalues of the channel covariance matrices of all other users. Before stating our theorem in this section, we need the following lemma.

Lemma 1 *When $\mathbf{A}$ and $\mathbf{A}_k$ are defined as in Theorem 2, the following identities hold* *for* $\quad i \neq 1$

$$E_{k1} = \lambda_{k1}^{\Sigma} E\left[\mathbf{z}_{k1}^{\dagger}\mathbf{A}^{-1}\mathbf{z}_{k1}\right] = \frac{1}{P_k}\left(1 - E\left[\frac{1}{1 + P_k\lambda_{k1}^{\Sigma}\mathbf{z}_{k1}^{T}\mathbf{A}_k^{-1}\mathbf{z}_{k1}}\right]\right) \tag{2.21}$$

$$E_{ki} = \lambda_{ki}^{\Sigma} E\left[\mathbf{z}_{ki}^{\dagger}\mathbf{A}^{-1}\mathbf{z}_{ki}\right] = \lambda_{ki}^{\Sigma}\left(n_R - K + \sum_{l=1}^{K} E\left[\frac{1}{1 + P_l\lambda_{l1}^{\Sigma}\mathbf{z}_{l1}^{T}\mathbf{A}_l^{-1}\mathbf{z}_{l1}}\right]\right) \tag{2.22}$$

A proof of Lemma 1 is given in Section 2.6.1 in the Appendix.

Theorem 2 *In a MIMO-MAC system where the transmitters have partial CSI in the form of covariance feedback, the transmit covariance matrices of all users that maximize (2.19) have unit-rank (i.e., beamforming is optimal for all users) if and only if*

$$P_k\lambda_{k2}^{\Sigma} < \frac{1 - E\left[\frac{1}{1+P_k\lambda_{k1}^{\Sigma}\mathbf{z}_{k1}\mathbf{A}_k^{-1}\mathbf{z}_{k1}}\right]}{n_R - K + \sum_{l=1}^{K} E\left[\frac{1}{1+P_l\lambda_{l1}^{\Sigma}\mathbf{z}_{l1}\mathbf{A}_l^{-1}\mathbf{z}_{l1}}\right]}, \quad k = 1, \ldots, K \tag{2.23}$$

where $\mathbf{A} = \mathbf{I}_{n_R} + \sum_{l=1}^{K} P_l\lambda_{l1}^{\Sigma}\mathbf{z}_{l1}\mathbf{z}_{l1}^{\dagger}$, $\mathbf{A}_k = \mathbf{A} - P_k\lambda_{k1}^{\Sigma}\mathbf{z}_{k1}\mathbf{z}_{k1}^{\dagger}$, λ_{ki}^{Σ} *is the* i^{th} *largest eigenvalue of the channel covariance matrix of user* k, *and* $\mathbf{z}_{l1}$ *are* $n_R \times 1$ *dimensional i.i.d., Gaussian random vectors with zero-mean and identity-covariance.*

Proof: The Lagrangian for the optimization problem in (2.19), with μ_k as the Lagrange multiplier of user k corresponding to its power constraint, is

$$L = E\left[\log\left|\mathbf{I}_{n_R} + \sum_{k=1}^{K}\sum_{i=1}^{n_T}\lambda_{ki}^{Q}\lambda_{ki}^{\Sigma}\mathbf{z}_{ki}\mathbf{z}_{ki}^{\dagger}\right|\right] - \sum_{k=1}^{K}\mu_k\left(\sum_{i=1}^{n_T}\lambda_{ki}^{Q} - P_k\right) \tag{2.24}$$

In order to derive the Karush-Kuhn-Tucker (KKT) conditions, we need the following identity which is proved in [14],

$$\frac{\partial}{\partial x} \log |\mathbf{A} + x\mathbf{B}| = \operatorname{tr}\left[(\mathbf{A} + x\mathbf{B})^{-1}\mathbf{B}\right] \tag{2.25}$$

Using this identity, the KKT conditions for user k are

$$\lambda_{ki}^{\Sigma} E\left[\mathbf{z}_{ki}^{\dagger}\left(\mathbf{I} + \sum_{l=1}^{K}\sum_{i=1}^{n_T} \lambda_{li}^{Q}\lambda_{li}^{\Sigma}\mathbf{z}_{li}\mathbf{z}_{li}^{\dagger}\right)^{-1}\mathbf{z}_{ki}\right] \leq \mu_k, \quad i = 1, \ldots, n_T \tag{2.26}$$

where the conditions are satisfied with equality if the corresponding eigenvalue of the transmit covariance matrix is non-zero. Beamforming is optimal for all users, if the inequalities corresponding to $i = 1$ for $k = 1, \ldots, K$ are satisfied with equality, and the rest of the inequalities remain as strict inequalities. In this case, $\lambda_{k1}^{Q} = P_k$, for $k = 1, \ldots, K$, and all other eigenvalues of the transmit covariance matrices are zero. We have the following for user k,

$$E_{k1} = \lambda_{k1}^{\Sigma} E\left[\mathbf{z}_{k1}^{\dagger}\mathbf{A}^{-1}\mathbf{z}_{k1}\right] = \mu_k \tag{2.27}$$

$$E_{ki} = \lambda_{ki}^{\Sigma} E\left[\mathbf{z}_{ki}^{\dagger}\mathbf{A}^{-1}\mathbf{z}_{ki}\right] < \mu_k, \quad \forall i \neq 1 \tag{2.28}$$

Equivalently, the conditions for the optimality of beamforming for all users are

$$\frac{E_{k1}}{E_{ki}} > 1, \quad \forall i \neq 1, \quad k = 1, \ldots, K \tag{2.29}$$

Due to the symmetry in these conditions, we will derive the condition for user k only. Using Lemma 1 and (2.29) for user k, we have

$$P_k \lambda_{ki}^{\Sigma} < \frac{1 - E\left[\frac{1}{1+P_k \lambda_{k1}^{\Sigma} \mathbf{z}_{k1}^T \mathbf{A}_k^{-1} \mathbf{z}_{k1}}\right]}{n_R - K + \sum_{l=1}^{K} E\left[\frac{1}{1+P_l \lambda_{l1}^{\Sigma} \mathbf{z}_{l1}^T \mathbf{A}_l^{-1} \mathbf{z}_{l1}}\right]}, \quad i = 2, \ldots, n_T \tag{2.30}$$

Note that the left hand side is maximized for $i = 2$, that is, if the condition for $i = 2$ holds, then it holds for all other i, as well. Therefore, inserting $i = 2$ in (2.30) gives the condition in (2.23) for user k. $\square$

Note that inserting $K = 1$ in (2.23), we obtain the condition in (2.20), which is derived in [14]. In our case, the right hand side of (2.23) depends only on the largest eigenvalues of all users. Therefore, in order to have the optimality of beamforming, a combination of the largest eigenvalues of all users induce an upper bound on the second largest eigenvalues of all users. If the second largest eigenvalues of all users satisfy (2.23), then beamforming is optimal for all users.

One important issue in the analysis of the region where beamforming is optimal, is the change in the region with varying numbers of users. In the next theorem, we show that the region where beamforming is optimal grows with the addition of new users into the system.

Theorem 3 *In a MIMO-MAC system where the transmitters have partial CSI in the form of covariance feedback, the region where beamforming is optimal gets larger by the addition of new users.*

Proof: From (2.29), beamforming is optimal for all users if and only if

$$\lambda_{k2}^{\Sigma} < \frac{\lambda_{k1}^{\Sigma} E\left[\mathbf{z}_{k1}^{\dagger}\mathbf{A}^{-1}\mathbf{z}_{k1}\right]}{E\left[\mathbf{z}_{k2}^{\dagger}\mathbf{A}^{-1}\mathbf{z}_{k2}\right]}, \quad k = 1, \ldots, K \tag{2.31}$$

Note that $\mathbf{z}_{k2}$ is independent of $\mathbf{A}$, and has identity covariance. Therefore, the denominator of the right hand side of (2.31) becomes $E[\mathrm{tr}(\mathbf{A}^{-1})]$. Let us define the "boundary function" $f_k(\boldsymbol{\lambda})$ as

$$f_k(\boldsymbol{\lambda}) = \frac{\lambda_{k1}^{\Sigma} E[\mathbf{z}_{k1}^{\dagger}\mathbf{A}^{-1}\mathbf{z}_{k1}]}{E[\mathrm{tr}(\mathbf{A}^{-1})]}, \quad k = 1, \ldots, K \tag{2.32}$$

where $\boldsymbol{\lambda} = [\lambda_{11}^{\Sigma}, \lambda_{21}^{\Sigma}, \ldots, \lambda_{K1}^{\Sigma}]^T$ contains the largest eigenvalues of the covariance feedback matrices of all users. Then, beamforming is optimal for all users if and only if

$$\lambda_{k2}^{\Sigma} < f_k(\boldsymbol{\lambda}), \qquad k = 1, \ldots, K \tag{2.33}$$

We will show that, $f_k(\boldsymbol{\lambda})$ increases in every component of the vector $\boldsymbol{\lambda}$, for all k. This will prove that, when a user is added to the system, i.e., the eigenvalue of the corresponding user is increased to a positive number from zero, the region in which beamforming is optimal for all users increases as long as the the condition for the new user is also satisfied. In order to prove that each $f_k(\boldsymbol{\lambda})$ increases in $\boldsymbol{\lambda}$, we will prove that every component of the vector of boundary functions, $\mathbf{f}(\boldsymbol{\lambda}) = [f_1(\boldsymbol{\lambda}), \ldots, f_K(\boldsymbol{\lambda})]^T$, increases in $\boldsymbol{\lambda}$. Let us define $\bar{\mathbf{Z}} = [\mathbf{z}_{11}, \mathbf{z}_{21}, \ldots, \mathbf{z}_{K1}]$, and $\bar{\boldsymbol{\Lambda}}$ and $\bar{\mathbf{P}}$ as diagonal matrices having $\{\lambda_{11}^{\Sigma}, \lambda_{21}^{\Sigma}, \ldots, \lambda_{K1}^{\Sigma}\}$ and $\{P_1, P_2, \ldots, P_K\}$ along their diagonals, respectively. Then,

$$\mathbf{f}(\boldsymbol{\lambda}) = \frac{\text{diag}\left\{E\left[\bar{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{Z}}^{\dagger}\mathbf{A}^{-1}\bar{\mathbf{Z}}\bar{\mathbf{\Lambda}}^{1/2}\right]\right\}}{E\left[\text{tr}(\mathbf{A}^{-1})\right]} \tag{2.34}$$

where diag$\{\cdot\}$ is the vector composed of the diagonal elements of its argument, and $\mathbf{A}$ can be expressed in terms of $\bar{\mathbf{Z}}$, $\bar{\mathbf{P}}$, and $\bar{\mathbf{\Lambda}}$ as $\mathbf{A} = \mathbf{I} + \bar{\mathbf{Z}}\bar{\mathbf{P}}\bar{\mathbf{\Lambda}}\bar{\mathbf{Z}}^{\dagger}$. Note that the expectation of the $(k,l)^{th}$ off-diagonal element of the random matrix $\bar{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{Z}}^{\dagger}\mathbf{A}^{-1}\bar{\mathbf{Z}}\bar{\mathbf{\Lambda}}^{1/2}$ is zero. The reason for this is that when the expectation is expressed as an integral, the contribution to the integral at $\mathbf{z}_{l1}$ is cancelled by the contribution at $-\mathbf{z}_{l1}$, due to the odd function property of $\lambda_{k1}^{\Sigma}\mathbf{z}_{k1}^{\dagger}\mathbf{A}^{-1}\mathbf{z}_{l1}$. Note also that, since $\lambda_{l1}^{\Sigma}\mathbf{z}_{l1}\mathbf{z}_{l1}^{\dagger} = \lambda_{l1}^{\Sigma}(-\mathbf{z}_{l1})(-\mathbf{z}_{l1})^{\dagger}$, the matrix $\mathbf{A}$ and the value of the probability density function are the same for $\mathbf{z}_{l1}$ and $-\mathbf{z}_{l1}$. Hence, we conclude that $E[\bar{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{Z}}^{\dagger}\mathbf{A}^{-1}\bar{\mathbf{Z}}\bar{\mathbf{\Lambda}}^{1/2}]$ is diagonal, and therefore its diagonal elements are the same as its eigenvalues.

Now, we will show that the eigenvalues of $E[\bar{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{Z}}^{\dagger}\mathbf{A}^{-1}\bar{\mathbf{Z}}\bar{\mathbf{\Lambda}}^{1/2}]$ increase in $\boldsymbol{\lambda}$, in two steps. First, we will show that the eigenvalues of $\bar{\mathbf{P}}\bar{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{Z}}^{\dagger}\mathbf{A}^{-1}\bar{\mathbf{Z}}\bar{\mathbf{\Lambda}}^{1/2}$, for any given realization of $\bar{\mathbf{Z}}$, increase in $\boldsymbol{\lambda}$, and then we will show that the eigenvalues of $E[\bar{\mathbf{P}}\bar{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{Z}}^{\dagger}\mathbf{A}^{-1}\bar{\mathbf{Z}}\bar{\mathbf{\Lambda}}^{1/2}]$ increase in $\boldsymbol{\lambda}$. This immediately implies that the eigenvalues of $E[\bar{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{Z}}^{\dagger}\mathbf{A}^{-1}\bar{\mathbf{Z}}\bar{\mathbf{\Lambda}}^{1/2}]$ increase in $\boldsymbol{\lambda}$. First consider a fixed realization of the random matrix $\bar{\mathbf{Z}}$. Note that,

$$\bar{\mathbf{Z}}\bar{\mathbf{P}}\bar{\mathbf{\Lambda}}\bar{\mathbf{Z}}^{\dagger} = \sum_{k=1}^{K} P_k \lambda_{k1}^{\Sigma}\mathbf{z}_{k1}\mathbf{z}_{k1}^{\dagger} \tag{2.35}$$

If we increase any one of λ_{k1}^{Σ}, $k = 1, \ldots, K$, to $(\lambda_{k1}^{\Sigma})'$, this can be seen as an addition of the positive semidefinite matrix $P_k \left((\lambda_{k1}^{\Sigma})' - \lambda_{k1}^{\Sigma} \right) \mathbf{z}_{k1} \mathbf{z}_{k1}^{\dagger}$ to the summation in (2.35). Using the corollary to Weyl's monotonicity theorem [12, page 181-182] which states that all eigenvalues of a Hermitian matrix increase if a positive semidefinite matrix is added to it, we can conclude that the eigenvalues of $\bar{\mathbf{Z}}\bar{\mathbf{P}}\bar{\mathbf{\Lambda}}\bar{\mathbf{Z}}^{\dagger}$ increase in $\boldsymbol{\lambda}$ for any fixed $\bar{\mathbf{Z}}$. Now, note that, if we denote the eigenvalues of $\bar{\mathbf{Z}}\bar{\mathbf{P}}\bar{\mathbf{\Lambda}}\bar{\mathbf{Z}}^{\dagger}$ as α_i, then the eigenvalues of $\mathbf{A}^{-1}\bar{\mathbf{Z}}\bar{\mathbf{P}}\bar{\mathbf{\Lambda}}\bar{\mathbf{Z}}^{\dagger}$ are given by $\frac{\alpha_i}{1+\alpha_i}$. Further, the eigenvalues of $\bar{\mathbf{P}}\bar{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{Z}}^{\dagger}\mathbf{A}^{-1}\bar{\mathbf{Z}}\bar{\mathbf{\Lambda}}^{1/2}$ are either $\frac{\alpha_i}{1+\alpha_i}$ or 0, depending on the dimensions of $\bar{\mathbf{Z}}$. Therefore, we conclude that when $\boldsymbol{\lambda}$ increases, all α_i increase as shown above, and therefore all $\frac{\alpha_i}{1+\alpha_i}$ increase as well.

Until now, we have shown that the eigenvalues of the random $\bar{\mathbf{P}}\bar{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{Z}}^{\dagger}\mathbf{A}^{-1}\bar{\mathbf{Z}}\bar{\mathbf{\Lambda}}^{1/2}$ increase in $\boldsymbol{\lambda}$. Next, we will show that the eigenvalues of $E[\bar{\mathbf{P}}\bar{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{Z}}^{\dagger}\mathbf{A}^{-1}\bar{\mathbf{Z}}\bar{\mathbf{\Lambda}}^{1/2}]$ increase in $\boldsymbol{\lambda}$ as well. We note that this expectation can be written as a positive weighted sum of positive semidefinite Hermitian matrices $\bar{\mathbf{P}}\bar{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{Z}}^{\dagger}\mathbf{A}^{-1}\bar{\mathbf{Z}}\bar{\mathbf{\Lambda}}^{1/2}$ for all realizations of the random matrix $\bar{\mathbf{Z}}$. An increase in $\boldsymbol{\lambda}$, can again be seen as an addition of a positive semidefinite matrix to the expectation. Therefore, invoking the corollary to Weyl's monotonicity theorem [12, page 181-182] once again, we conclude that the eigenvalues of $E[\bar{\mathbf{P}}\bar{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{Z}}^{\dagger}\mathbf{A}^{-1}\bar{\mathbf{Z}}\bar{\mathbf{\Lambda}}^{1/2}]$, and consequently the eigenvalues of $E[\bar{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{Z}}^{\dagger}\mathbf{A}^{-1}\bar{\mathbf{Z}}\bar{\mathbf{\Lambda}}^{1/2}]$ increase in $\boldsymbol{\lambda}$. We also note that $E\left[\mathrm{tr}(\mathbf{A}^{-1})\right]$ decreases in $\boldsymbol{\lambda}$, since the eigenvalues of $\mathbf{A}^{-1}$, i.e., $\frac{1}{1+\alpha_i}$, decrease as $\boldsymbol{\lambda}$ increases. Therefore, the ratios on the right hand side of (2.34), and therefore, $\mathbf{f}(\boldsymbol{\lambda})$, increase in $\boldsymbol{\lambda}$. $\square$

Theorem 3 shows that with the addition of more and more users into the system, beamforming becomes optimal for more and more channel covariance matrices.

Whether the growth in the region where beamforming is optimal is bounded, or whether beamforming is unconditionally optimal for very large numbers of users in a fading environment will be addressed in Section 2.3.

2.2.2 Mean Feedback at the Transmitters

Transmit Directions

As in the case of covariance feedback, for a single-user system with partial CSI in the form of the channel mean matrix at the transmitter, the capacity is no longer achieved by an identity transmit covariance matrix. The optimization problem in this case is the same as (2.8), with the difference that, in this setting, the channel covariance matrix is identity, i.e., $\mathbf{\Sigma} = \mathbf{I}$, and the channel mean matrix $\mathbf{H}_\mu$ is fedback to the transmitter. With the assumption that $\mathbf{H}_\mu$ is unit-rank, [14], [46] showed that the optimal transmit covariance matrix $\mathbf{Q}$ that solves (2.8) can be written as

$$\mathbf{Q} = \mathbf{U}_\mu \mathbf{\Lambda}_Q \mathbf{U}_\mu^\dagger \tag{2.36}$$

where the first column of the unitary matrix $\mathbf{U}_\mu$ is the eigenvector corresponding to the non-zero eigenvalue of $\mathbf{H}_\mu$, and the remaining columns are arbitrary, with the restriction that the columns of $\mathbf{U}_\mu$ are orthonormal.

In this section, we show that, in a multi-user setting, every user should transmit along the eigenvectors of its own channel mean matrix. In the multi-user setting, let the singular value decomposition of the channel mean matrix of user k be

$$\mathbf{H}_{\mu_k} = \mathbf{U}_{\mu_k}\mathbf{\Lambda}_{\mu_k}\mathbf{V}_{\mu_k}^{\dagger} \tag{2.37}$$

Since $\mathbf{H}_{\mu_k}$ is a unit-rank matrix as in (2.5), the first column of $\mathbf{U}_{\mu_k}$ can be chosen as $\frac{\mathbf{a}_R}{|\mathbf{a}_R|}$; and the rest of the columns can be chosen arbitrarily as long as $\mathbf{U}_{\mu_k}$ has orthonormal columns. Also, note that $\mathbf{U}_{\mu_k} = \mathbf{U}_{\mu}$, for $k = 1, \ldots, K$. Similarly, the first column of $\mathbf{V}_{\mu_k}$ can be chosen as $\frac{\mathbf{a}_{T_k}}{|\mathbf{a}_{T_k}|}$ and the rest of the columns can be chosen arbitrarily as long as $\mathbf{V}_{\mu_k}$ has orthonormal columns. Unlike $\mathbf{U}_{\mu_k}$, $\mathbf{V}_{\mu_k}$ is different for different users. The diagonal matrix $\mathbf{\Lambda}_{\mu_k}$ has only one non-zero element, which is $|\mathbf{a}_R||\mathbf{a}_{T_k}|$.

The following theorem identifies the optimum transmit directions for all users. The single-user version of this theorem was proved in [11], [13].

Theorem 4 *Let $\mathbf{H}_{\mu_k} = \mathbf{U}_{\mu}\mathbf{\Lambda}_{\mu_k}\mathbf{V}_{\mu_k}^{\dagger}$ be the singular value decomposition of the channel mean matrix of user k. Then, the optimum transmit covariance matrix $\mathbf{Q}_k$ of user k may be expressed as $\mathbf{Q}_k = \mathbf{V}_{\mu_k}\mathbf{\Lambda}_k\mathbf{V}_{\mu_k}^{\dagger}$, for all users.*

Proof: We prove the theorem in two steps. In the first step, we show that the sum-capacity resulting from $\{\mathbf{H}_{\mu_k}\}_{k=1}^{K}$ as the channel mean matrices and the sum-capacity resulting from $\{\mathbf{\Lambda}_{\mu_k}\}_{k=1}^{K}$ as the channel mean matrices are the same.

The optimization problem in (2.6) with channel mean matrices $\{\mathbf{H}_{\mu_k}\}_{k=1}^{K}$ can be written as

$$C_{sum}(\{\mathbf{H}_{\mu_k}\}_{k=1}^K) = \max_{\substack{\text{tr}(\mathbf{Q}_k)\le P_k \\ k=1,\dots,K}} E\left[\log\left|\mathbf{I}_{n_R} + \sum_{k=1}^{K}(\mathbf{H}_{\mu_k}+\mathbf{Z}_k)\mathbf{Q}_k(\mathbf{H}_{\mu_k}+\mathbf{Z}_k)^\dagger\right|\right] \tag{2.38}$$

using (2.4). Using the singular value decomposition of the channel mean matrix of user k and the invariance of the joint distribution of zero-mean, identity-covariance matrices $\mathbf{Z}_k$ under unitary transformations, i.e., that $\{\mathbf{Z}_k\}_{k=1}^K$ and $\{\mathbf{U}_k\mathbf{Z}_k\mathbf{V}_k\}_{k=1}^K$ have the same joint distribution, we have

$$C_{sum}(\{\mathbf{H}_{\mu_k}\}_{k=1}^K) = \max_{\substack{\text{tr}(\mathbf{Q}_k)\le P_k \\ k=1,\dots,K}} E\left[\log\left|\mathbf{I} + \sum_{k=1}^{K}(\mathbf{U}_\mu\mathbf{\Lambda}_{\mu_k}\mathbf{V}_{\mu_k}^\dagger+\mathbf{Z}_k)\mathbf{Q}_k(\mathbf{U}_\mu\mathbf{\Lambda}_{\mu_k}\mathbf{V}_{\mu_k}^\dagger+\mathbf{Z}_k)^\dagger\right|\right] \tag{2.39}$$

$$= \max_{\substack{\text{tr}(\mathbf{Q}_k)\le P \\ k=1\dots K}} E\left[\log\left|\mathbf{I} + \sum_{k=1}^{K}\mathbf{U}_\mu(\mathbf{\Lambda}_{\mu_k}+\mathbf{Z}_k)\mathbf{V}_{\mu_k}^\dagger\mathbf{Q}_k\mathbf{V}_{\mu_k}(\mathbf{\Lambda}_{\mu_k}+\mathbf{Z}_k)^\dagger\mathbf{U}_\mu^\dagger\right|\right] \tag{2.40}$$

$$= \max_{\substack{\text{tr}(\tilde{\mathbf{Q}}_k)\le P_k \\ k=1,\dots,K}} E\left[\log\left|\mathbf{I} + \sum_{k=1}^{K}(\mathbf{\Lambda}_{\mu_k}+\mathbf{Z}_k)\tilde{\mathbf{Q}}_k(\mathbf{\Lambda}_{\mu_k}+\mathbf{Z}_k)^\dagger\right|\right] \tag{2.41}$$

$$= C_{sum}(\{\mathbf{\Lambda}_{\mu_k}\}_{k=1}^K) \tag{2.42}$$

where we used $|\mathbf{I}+\mathbf{AB}| = |\mathbf{I}+\mathbf{BA}|$ to cancel $\mathbf{U}_\mu$. Note that $\text{tr}(\mathbf{Q}_k) = \text{tr}(\tilde{\mathbf{Q}}_k)$, since $\tilde{\mathbf{Q}}_k = \mathbf{V}_{\mu_k}^\dagger\mathbf{Q}_k\mathbf{V}_{\mu_k}$. By comparing (2.38) and (2.41), we see that the diagonal eigenvalue matrices of the channel mean matrices result in the same sum-capacity as the channel mean matrices themselves except that we changed the transmit covariance matrices accordingly.

In the second step, our goal is to show that the optimal $\tilde{\mathbf{Q}}_k$ in (2.41) is diagonal. In order to prove this, we use the technique presented in [13]. Let $\mathbf{\Xi}$ be an $n_T \times n_T$ diagonal matrix, whose i^{th} diagonal entry is -1, and all other diagonal entries are 1. Let $\tilde{\mathbf{\Xi}}$ be an $n_R \times n_R$ diagonal matrix such that if $n_R < n_T$, then $\tilde{\mathbf{\Xi}} = \mathbf{I}_{n_R}$ and if $n_R > n_T$, then the i^{th} diagonal entry of $\tilde{\mathbf{\Xi}}$ is -1, and all other diagonal entries are 1. Then, we have

$$\tilde{\mathbf{\Xi}}\mathbf{\Lambda}_{\mu_k}\mathbf{\Xi} = \mathbf{\Lambda}_{\mu_k}, \quad k = 1, \ldots, K \tag{2.43}$$

Let us consider now a set of arbitrary transmit covariance matrices $\{\tilde{\mathbf{Q}}_k\}_{k=1}^K$, and define another set of transmit covariance matrices as $\hat{\mathbf{Q}}_k = \mathbf{\Xi}^\dagger \tilde{\mathbf{Q}}_k \mathbf{\Xi}$, for $k = 1, \ldots, K$. Note that the entries of $\hat{\mathbf{Q}}_k$ are equal to the entries of $\tilde{\mathbf{Q}}_k$ except that the off-diagonal entries in the i^{th} row and column are negated. We can rewrite the optimization problem in (2.41) as

$$\begin{aligned}
C_{sum}(\{\tilde{\mathbf{Q}}_k\}_{k=1}^K) &= E\left[\log\left|\mathbf{I}_{n_R} + \sum_{k=1}^K (\mathbf{\Lambda}_{\mu_k} + \mathbf{Z}_k)\mathbf{\Xi}\hat{\mathbf{Q}}_k\mathbf{\Xi}^\dagger(\mathbf{\Lambda}_{\mu_k} + \mathbf{Z}_k)^\dagger\right|\right] && (2.44)\\
&= E\left[\log\left|\mathbf{I}_{n_R} + \sum_{k=1}^K (\tilde{\mathbf{\Xi}}\mathbf{\Lambda}_{\mu_k}\mathbf{\Xi} + \mathbf{Z}_k)\hat{\mathbf{Q}}_k(\tilde{\mathbf{\Xi}}\mathbf{\Lambda}_{\mu_k}\mathbf{\Xi} + \mathbf{Z}_k)^\dagger\right|\right] && (2.45)\\
&= E\left[\log\left|\mathbf{I}_{n_R} + \sum_{k=1}^K (\mathbf{\Lambda}_{\mu_k} + \mathbf{Z}_k)\hat{\mathbf{Q}}_k(\mathbf{\Lambda}_{\mu_k} + \mathbf{Z}_k)^\dagger\right|\right] && (2.46)\\
&= C_{sum}(\{\hat{\mathbf{Q}}_k\}_{k=1}^K) && (2.47)
\end{aligned}$$

where we again used the fact that $\{\mathbf{Z}_k\}_{k=1}^K$ and $\{\tilde{\mathbf{\Xi}}\mathbf{Z}_k\mathbf{\Xi}\}_{k=1}^K$ have the same joint distribution, and inserted (2.43) into (2.44) to obtain (2.46).

Now, let us define the set of transmit covariance matrices as $\mathbf{Q}_k^* = \frac{1}{2}\tilde{\mathbf{Q}}_k + \frac{1}{2}\hat{\mathbf{Q}}_k$, for $k = 1, \ldots, K$. The entries of $\mathbf{Q}_k^*$ are equal to the entries of $\tilde{\mathbf{Q}}_k$ except that the off-diagonal entries in the i^{th} row and column are zero. By the concavity of the mutual information, it follows that the mutual information achieved by $\{\mathbf{Q}_k^*\}_{k=1}^K$ is greater than or equal to the mutual information achieved by $\{\tilde{\mathbf{Q}}_k\}_{k=1}^K$,

$$\begin{aligned} C_{sum}(\{\mathbf{Q}_k^*\}_{k=1}^K) &\geq \frac{1}{2}\left(C_{sum}(\{\tilde{\mathbf{Q}}_k\}_{k=1}^K) + C_{sum}(\{\hat{\mathbf{Q}}_k\}_{k=1}^K)\right) && (2.48) \\ &= C_{sum}(\{\tilde{\mathbf{Q}}_k\}_{k=1}^K) && (2.49) \end{aligned}$$

Applying this procedure to every i for $1 \leq i \leq n_T$, we have shown that nulling the off-diagonal elements of the transmit covariance matrices increases the capacity. This proves that the optimal $\tilde{\mathbf{Q}}_k$ is diagonal, and is equal to $\mathbf{\Lambda}_k$, for all k. This also proves the theorem since we have $\mathbf{Q}_k = \mathbf{V}_{\mu_k}\tilde{\mathbf{Q}}_k\mathbf{V}_{\mu_k}^\dagger = \mathbf{V}_{\mu_k}\mathbf{\Lambda}_k\mathbf{V}_{\mu_k}^\dagger$ $\square$

Using Theorem 4, we can write the optimization problem in (2.38) as

$$\begin{aligned} C_{sum} &= \max_{\substack{\mathrm{tr}(\mathbf{\Lambda}_k)\leq P_k, \mathbf{\Lambda}_k \succeq \mathbf{0} \\ k=1,\ldots,K}} E\left[\log\left|\mathbf{I}_{n_R} + \sum_{k=1}^K \hat{\mathbf{Z}}_k\mathbf{\Lambda}_k\hat{\mathbf{Z}}_k^\dagger\right|\right] && (2.50) \\ &= \max_{\substack{\sum_{i=1}^{n_T}\lambda_{ki}^Q\leq P_k, \lambda_{ki}^Q\geq 0 \\ k=1,\ldots,K}} E\left[\log\left|\mathbf{I}_{n_R} + \sum_{k=1}^K\sum_{i=1}^{n_T} \lambda_{ki}^Q\hat{\mathbf{z}}_{ki}\hat{\mathbf{z}}_{ki}^\dagger\right|\right] && (2.51) \end{aligned}$$

where $\hat{\mathbf{Z}}_k = \mathbf{\Lambda}_{\mu_k} + \mathbf{Z}_k$. Note that while the first column of this matrix is a non-zero mean Gaussian vector, all of the remaining columns are zero-mean Gaussian vectors.

Similar to the covariance feedback case, in a MIMO system, a transmit strategy is a combination of a transmit direction strategy and a transmit power allocation strategy. A result of Theorem 4 is that the optimal multi-user transmit direction strategies are decoupled into single-user transmit direction strategies. However, in general, this is not true for the optimal transmit power allocation strategies. On the other hand, we know that when beamforming is optimal, the optimal transmit power allocation strategy for each user is to allocate all of its power to its strongest eigen-direction. Therefore, for the range of parameters where beamforming is optimal, both the optimal transmit direction and the optimal transmit power allocation strategies are decoupled among users.

Conditions for the Optimality of Beamforming

In this section, we determine the conditions for the optimality of beamforming in a multi-user system with a finite number of users, when partial CSI available at the transmitters is in the form of mean feedback. Reference [14] identified these conditions for a single-user system. For a single-user system, let λ^{μ} denote the non-zero eigenvalue of the channel mean matrix $\mathbf{H}_{\mu}$. Then, the necessary and sufficient condition for the optimality of beamforming is [14]:

$$P < \frac{1 - E\left[\frac{1}{1+P\hat{\mathbf{z}}^{\dagger}\hat{\mathbf{z}}}\right]}{n_R - 1 + E\left[\frac{1}{1+P\hat{\mathbf{z}}^{\dagger}\hat{\mathbf{z}}}\right]} \tag{2.52}$$

where $\hat{\mathbf{z}}$ is an $n_R \times 1$ dimensional Gaussian random vector with identity-covariance. The first entry of $\hat{\mathbf{z}}$ has a mean of λ_{μ}, while all other entries have zero-mean.

Similar to the covariance feedback case, we find the conditions for the optimality of beamforming for all users in a multi-user setting. Inserting $K = 1$ in our results would reduce them to (2.52). In our results, the number of conditions equals the number of users. The condition corresponding to user k depends on the non-zero eigenvalues of the channel mean matrices of all users. We have the following theorem.

Theorem 5 *In a MIMO-MAC system where the transmitters have partial CSI in the form of mean feedback, the transmit covariance matrices of all users that maximize (2.51) have unit-rank (i.e., beamforming is optimal for all users) if and only if*

$$P_k < \frac{1 - E\left[\frac{1}{1+P_k \hat{\mathbf{z}}_{k1}^\dagger \mathbf{B}_k^{-1} \hat{\mathbf{z}}_{k1}}\right]}{n_R - K + \sum_{l=1}^{K} E\left[\frac{1}{1+P_l \hat{\mathbf{z}}_{l1}^\dagger \mathbf{B}_l^{-1} \hat{\mathbf{z}}_{l1}}\right]}, \quad k = 1, \dots, K \tag{2.53}$$

where $\mathbf{B} = \mathbf{I}_{n_R} + \sum_{l=1}^{K} P_l \hat{\mathbf{z}}_{l1} \hat{\mathbf{z}}_{l1}^\dagger$, $\mathbf{B}_k = \mathbf{B} - P_k \hat{\mathbf{z}}_{k1} \hat{\mathbf{z}}_{k1}^\dagger$, *and* $\hat{\mathbf{z}}_{k1} = \lambda_k^\mu \mathbf{e}_1 + \mathbf{z}_{k1}$ *is the first column of the matrix* $\hat{\mathbf{Z}}_k$.

Proof: The Lagrangian for the optimization problem in (2.51), with ν_k as the Lagrange multiplier of user k corresponding to its power constraint, is

$$L = E\left[\log\left|\mathbf{I}_{n_R} + \sum_{k=1}^{K}\sum_{i=1}^{n_T} \lambda_{ki}^Q \hat{\mathbf{z}}_{ki} \hat{\mathbf{z}}_{ki}^\dagger\right|\right] - \sum_{k=1}^{K} \nu_k \left(\sum_{i=1}^{n_T} \lambda_{ki}^Q - P_k\right) \tag{2.54}$$

This Lagrangian for the mean feedback case is similar to the Lagrangian for the covariance feedback case in (2.24) with the difference that there are no second largest eigenvalues of the channel mean matrices. The following KKT conditions for user k

can be derived using (2.25),

$$E\left[\hat{\mathbf{z}}_{ki}^{\dagger}\left(\mathbf{I}+\sum_{k=1}^{K}\sum_{i=1}^{n_T}\lambda_{ki}^{Q}\hat{\mathbf{z}}_{ki}\hat{\mathbf{z}}_{ki}^{\dagger}\right)^{-1}\hat{\mathbf{z}}_{ki}\right]\leq\nu_k,\quad i=1,\ldots,n_T \tag{2.55}$$

Similar to the covariance feedback case, in order for beamforming to be optimal, we should have $\lambda_{k1}^{Q}=P_k$, and all other eigenvalues of the transmit covariance matrices to be zero. We have the following for user k,

$$E_{k1} \;=\; E\left[\hat{\mathbf{z}}_{k1}^{\dagger}\mathbf{B}^{-1}\hat{\mathbf{z}}_{k1}\right]=\nu_k \tag{2.56}$$

$$E_{ki} \;=\; E\left[\hat{\mathbf{z}}_{ki}^{\dagger}\mathbf{B}^{-1}\hat{\mathbf{z}}_{ki}\right]<\nu_k,\quad \forall i\neq 1 \tag{2.57}$$

Equivalently, the conditions for the optimality of beamforming for all users are

$$\frac{E_{k1}}{E_{ki}}>1,\quad \forall i\neq 1,\quad k=1,\ldots,K \tag{2.58}$$

Finally, using Lemma 1, we have (2.53). $\square$

Note that inserting $K=1$ in (2.53), we obtain the condition in (2.52), which is derived in [14]. In our case, the condition in (2.53) depends only on the sole non-zero channel mean eigenvalues of all users. Therefore, if the powers and the eigenvalues of the feedback mean matrices of all users are such that they satisfy the inequalities in (2.53), then beamforming is optimal for all users.

Contrary to the covariance feedback case, in the mean feedback case, the region where beamforming is optimal does not necessarily grow with the addition of new users into the system. The reason that the proof of Theorem 3 does not follow in the mean feedback case is the following. Note that, in the covariance feedback case, the off-diagonal entries of the matrix in the numerator of (2.34) were zero. However, in the mean feedback case, the off-diagonal entries of the corresponding matrix are not zero. Therefore, proving that the eigenvalues of that matrix increase, does not prove that the diagonal entries of the same matrix increase as well. However, for relatively large numbers of users, we see through simulations that it is harder to violate the beamforming condition. We discuss this issue in more detail in Section 2.4.

We have proved for the covariance feedback case and observed through simulations for the mean feedback case with relatively large numbers of users that the region where beamforming is optimal for all users grows, as new users are added to the system. These results and the asymptotic results of [31] with deterministic channel assumption motivate us to investigate whether the growth of the region where beamforming is optimal is bounded, or whether beamforming is unconditionally optimal for very large numbers of users in a fading environment. We address this issue in the next section.

2.3 Asymptotic Analysis

It is not immediate from the previous section that the region where beamforming is optimal covers the entire channel parameter space for all users when the number of users grows to infinity. In this section, we show that for very large numbers of users, even with the assumption that the transmitters have no knowledge of the channel, beamforming achieves a sum-rate which approaches the sum-capacity. For asymptotic analysis, we need the following lemma.

Lemma 2 *Let* $\mathbf{x}_i$, $i = 1, 2, \ldots$ *be a sequence of i.i.d. random vectors of length* M, *which have zero-mean and identity-covariance matrix, and let* α_i, $i = 1, 2, \ldots$ *be a sequence of bounded real numbers. Then,*

$$E\left[\log\left|\mathbf{I}_M + \sum_{i=1}^{N} \alpha_i \mathbf{x}_i \mathbf{x}_i^\dagger\right|\right] \doteq M \log\left(1 + \sum_{i=1}^{N} \alpha_i\right) \tag{2.59}$$

where the symbol $\doteq$ *denotes "equal for asymptotically large* N*".*

This is a version of the Strong Law of Large Numbers (SLLN), which states that the sum of independent, non-identically distributed random variables, converges to the sum of the means of the random variables. In particular, this version of SLLN is applied to independent random vectors $\sqrt{\alpha_i}\mathbf{x}_i$ in (2.59) which are non-identically distributed. A formal proof of Lemma 2 is given in Section 2.6.1 in the Appendix.

Lemma 2 will be used to state a form of channel "hardening" in the next three sub-sections. We will use Lemma 2 to say that, when the number of users grows to infinity, a form of channel hardening will occur, i.e., roughly speaking $\sum_{k=1}^{K} \sum_{i=1}^{n_T} \mathbf{h}_{ki} \mathbf{h}_{ki}^\dagger$ in (2.7), $\sum_{k=1}^{K} \sum_{i=1}^{n_T} \lambda_{ki}^{Q} \lambda_{ki}^{\Sigma} \mathbf{z}_{ki} \mathbf{z}_{ki}^\dagger$ in (2.19), and $\sum_{k=1}^{K} \sum_{i=1}^{n_T} \lambda_{ki}^{Q} \hat{\mathbf{z}}_{ki} \hat{\mathbf{z}}_{ki}^\dagger$ in (2.51) will converge to deterministic quantities almost surely, and that those deterministic quantities can be approached if simple beamforming is used. When beamforming is used, the sum $\sum_{i=1}^{n_T}$ drops in all three sums, however the sum over k, i.e., $\sum_{k=1}^{K}$ suffices to create the effect of SLLN.

The concept of channel hardening has been observed in [10] also, where instead of a SLLN approach, a Central Limit Theorem (CLT) approach is used to conclude that the mutual information converges to a Gaussian random variable whose variance vanishes. In [10], the number(s) of transmit and receive antennas grow(s) large for a single-user system, while here, the number of transmit and receive antennas are fixed, but the number of users goes to infinity in a MAC. Nevertheless, we observe similar mathematical phenomena as in [10].

2.3.1 No CSI at the Transmitters

When there is no CSI at the transmitters, the optimal transmit strategy is to use an identity transmit covariance matrix for all users [42]. In this section, we show that when there is no CSI at the transmitters, for an asymptotically large system, an arbitrary antenna selection scheme is sufficient to achieve the sum-capacity. This is stated in the following theorem.

Theorem 6 *In a system where there is no CSI at the transmitters, if the number of users grows to infinity, then the sum-rate achieved by unit-rank transmit covariance matrices approaches the sum capacity. In particular, this unit-rank transmission scheme takes the form of a simple antenna selection.*

Proof: The sum-capacity in this case is given in (2.6) with $\mathbf{Q}_k = \frac{P_k}{n_T}\mathbf{I}$, for all k. We define C_{sum}^{aas} as the achievable sum-rate by performing arbitrary antenna selection (aas) at all transmitters:

$$C_{sum}^{aas} \quad = \quad E\left[\log\left|\mathbf{I}_{n_R} + \sum_{k=1}^{K} P_k \mathbf{h}_{ka_k}\mathbf{h}_{ka_k}^{\dagger}\right|\right] \tag{2.60}$$

where $\mathbf{h}_{ka_k}$ is the a_k^{th} column of the channel matrix of user k, and a_k is the antenna chosen by user k, $1 \leq a_k \leq n_T$. The choice of the columns does not affect our result. All users may select their first antenna, i.e., $a_k = 1$, for all k, or they may select an antenna arbitrarily. Since SLLN averages out the randomness in the channel regardless of the realizations, so long as the columns of the channel matrices are independent, the transmit antenna each user selects is immaterial.

By noting that P_k are a series of bounded numbers, we apply Lemma 2 to (2.6) by inserting $\mathbf{Q}_k = \frac{P_k}{n_T}\mathbf{I}$, for all k, and to (2.60). We have

$$C_{sum} \doteq C_{sum}^{aas} \doteq n_R \log\left(1 + \sum_{k=1}^{K} P_k\right) \tag{2.61}$$

Therefore, we see that the sum-rates achievable by the optimal power allocation and the arbitrary antenna selection scheme converge to the same quantity asymptotically.

□

We note that this result does not contradict with the result of [42] which is stated in Section 2.2. For a multi-user system, full-rank transmit covariance matrices are optimum in the sense of maximizing the sum-rate [42]. Theorem 6 states that arbitrary antenna selection scheme is also sufficient to achieve the optimum when the number of users grows to infinity. In other words, the performance of the arbitrary antenna selection scheme converges to the optimum when the number of users goes to infinity.

2.3.2 Covariance Feedback at the Transmitters

When the transmitters have partial CSI in the form of covariance feedback, Theorem 1 shows that for any number of users, the transmit directions of a user are the eigenvectors of its own channel covariance feedback matrix. For sufficiently large numbers of users, the asymptotic optimality of beamforming in achieving the sum-capacity is proved in the following theorem.

Theorem 7 *In a system where there is partial CSI at the transmitters in the form of covariance feedback, if the number of users grows to infinity, then the sum-rate achieved by unit-rank transmit covariance matrices (i.e., beamforming) approaches the sum-capacity. In particular, this beamforming, for each user, is in the direction of the strongest eigenvector of the channel covariance matrix of that user.*

Proof: Note that λ_{ki}^{Q} is bounded for all (k, i), since power constraints for all users are finite, and λ_{ki}^{Σ} is bounded for all (k, i), since the covariance matrix, $\mathbf{\Sigma}_k$, of the channel has finite trace. Now, we can apply Lemma 2 to (2.19) with $\alpha_{ki} = \lambda_{ki}^{Q}\lambda_{ki}^{\Sigma}$. We have

$$C_{sum} \doteq \max_{\substack{\sum_{i=1}^{n_T} \lambda_{ki}^{Q} \leq P_k \\ k=1\dots K}} n_R \log\left(1 + \sum_{k=1}^{K}\sum_{i=1}^{n_T} \lambda_{ki}^{Q}\lambda_{ki}^{\Sigma}\right) \tag{2.62}$$

In order to solve the above optimization problem, we form the Lagrangian with μ_k's as the Lagrange multipliers,

$$L = n_R \log\left(1 + \sum_{k=1}^{K}\sum_{i=1}^{n_T} \lambda_{ki}^{Q}\lambda_{ki}^{\Sigma}\right) - \sum_{k=1}^{K} \mu_k \left(\sum_{i=1}^{n_T} \lambda_{ki}^{Q} - P_k\right) \tag{2.63}$$

The KKT optimality conditions are,

$$\frac{n_R \lambda_{ki}^{\Sigma}}{1 + \sum_{k=1}^{K}\sum_{i=1}^{n_T} \lambda_{ki}^{Q}\lambda_{ki}^{\Sigma}} \leq \mu_k, \qquad i = 1, \dots, n_T, \quad k = 1, \dots, K \tag{2.64}$$

where (2.64) is satisfied with equality if $\lambda_{ki}^{Q} > 0$. Note that the denominators on the left hand side of all the KKT conditions are the same. Without loss of generality, let $\lambda_{kn_T}^{\Sigma} < \cdots < \lambda_{k1}^{\Sigma}$ for user k. Assume that $\lambda_{kj}^{Q} > 0$ and $\lambda_{ki}^{Q} > 0$. Then, we must have $\lambda_{kj}^{\Sigma} = \lambda_{ki}^{\Sigma}$, which is a contradiction. Therefore, for user k, only one λ_{kj}^{Q}, $j = 1, \cdots, n_T$ can be non-zero. From the objective function in (2.62), we observe that the non-zero λ_{kj}^{Q} must correspond to the largest eigenvalue of the channel covariance feedback matrix. Hence, the only non-zero power component in $\mathbf{\Lambda}_{Q_k}$ is the first diagonal element. Finally, from the trace constraint, we have $\lambda_{k1}^{Q} = P_k$, for all k. The asymptotic sum capacity becomes

$$C_{sum} \doteq n_R \log \left(1 + \sum_{k=1}^{K} P_k \lambda_{k1}^{\Sigma} \right) \tag{2.65}$$

□

2.3.3 Mean Feedback at the Transmitters

When the transmitters have partial CSI in the form of mean feedback, Theorem 4 shows that for any number of users, the transmit directions of a user are the eigenvectors of its own channel mean feedback matrix. For sufficiently large numbers of users, the asymptotic optimality of beamforming in achieving the sum-capacity is proved in the following theorem.

Theorem 8 *In a system where there is partial CSI at the transmitters in the form of mean feedback, if the number of users grows to infinity, then the sum-rate achieved by unit-rank transmit covariance matrices (i.e., beamforming) approaches the sum-capacity. In particular, this beamforming, for each user, is in the direction of the eigenvector corresponding to the sole non-zero eigenvalue of the channel mean matrix of that user.*

Proof: First, note that λ_{ki}^{μ} for all (k, i) is bounded, since the channel has finite mean information. Applying Lemma 2 to (2.51), while noting that λ_{ki}^{μ} is non-zero for only

$i = 1$, and therefore using different α_{ki} (all of which are bounded) for diagonal and off-diagonal entries, we get

$$C_{sum} \doteq \max_{\substack{\sum_{i=1}^{n_T} \lambda_{ki}^Q \leq P_k \\ k=1\ldots K}} (n_R - 1) \log \left(1 + \sum_{k=1}^{K} P_k\right) + \log \left(1 + \sum_{k=1}^{K} \left((\lambda_{k1}^{\mu})^2 \lambda_{k1}^{Q} + P_k\right)\right) \tag{2.66}$$

Since only one eigenvalue from the transmit covariance matrix of each user appears in (2.66), the optimum choice is to allocate all of the power of each user to the eigenvector of its own channel mean matrix corresponding to the only non-zero eigenvalue, i.e., $\lambda_{k1}^{Q} = P_k$, for all k. The resulting sum-capacity becomes,

$$C_{sum} \doteq (n_R - 1) \log \left(1 + \sum_{k=1}^{K} P_k\right) + \log \left(1 + \sum_{k=1}^{K} \left((\lambda_{k1}^{\mu})^2 P_k + P_k\right)\right) \tag{2.67}$$

$\square$

2.4 Numerical Results

The region where beamforming is optimal is multi-dimensional. In order to illustrate the effects of having more than two users, we plot two dimensional slices from the region where beamforming is optimal for all users. We first consider the covariance feedback case, and plot these slices for $K = 1, 2, 3, 5, 10$ users in Figure 2.1. These slices give the maximum possible λ_{12}^{Σ} for a range of λ_{11}^{Σ}. The largest eigenvalues of the remaining users are kept constant. The number of transmit and receive antennas is $n_T = n_R = 2$. We see that the region where beamforming is optimal gets

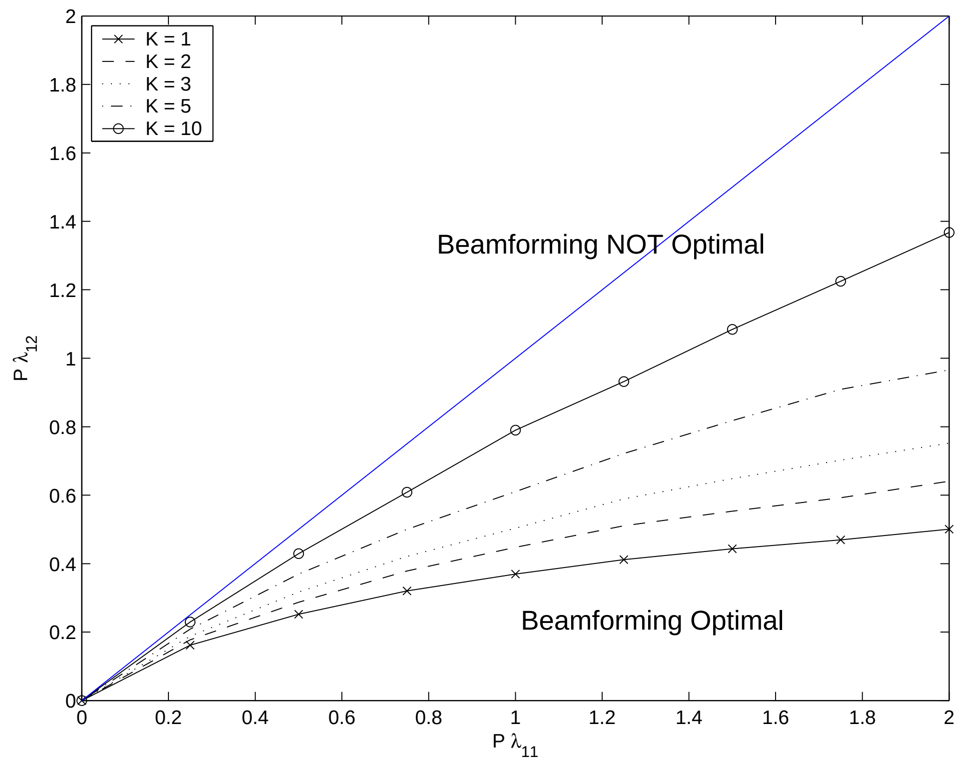


Figure 2.1: The region where beamforming is optimal for various numbers of users in the covariance feedback model.

larger with increasing number of users. Note that these curves have to lie below the $\lambda_{12}^{\Sigma} = \lambda_{11}^{\Sigma}$ line, because λ_{11}^{Σ} is the largest eigenvalue. The top-most line in Figure 2.1 is the $\lambda_{12}^{\Sigma} = \lambda_{11}^{\Sigma}$ line. We observe that the curves get closer to the $\lambda_{12}^{\Sigma} = \lambda_{11}^{\Sigma}$ line as K increases. This figure shows that with the addition of more and more users into the system, a larger range of $(\lambda_{11}^{\Sigma}, \lambda_{12}^{\Sigma})$ pairs becomes optimal.

For the mean feedback model, we will demonstrate two different cases. In the first case, the region where beamforming is optimal gets larger by addition of new users into the system. In Figure 2.2, we plot one dimensional slices from the region corresponding to $K = 1, 2, 3, 5, 10$. These lines give λ_1^{μ} values for beamforming to be optimal for a given power constraint, $P_k = 1$ for all k. The largest eigenvalues of

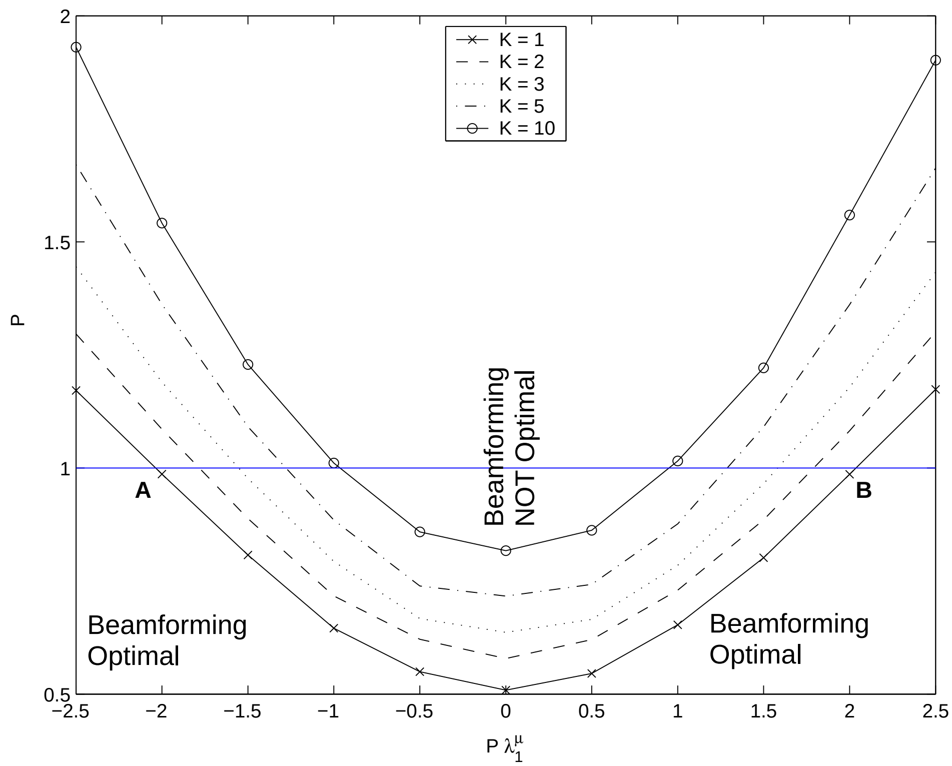


Figure 2.2: The region where beamforming is optimal for various numbers of users in the mean feedback model. This is an example where the region gets larger as more users are added into the system.

all other users, which are kept constant, are comparable in value to each other. The number of transmit and receive antennas is $n_T = n_R = 2$. The curves in Figure 2.2 correspond to the left hand side of (2.58) for $k = 1$. Beamforming is optimal for the range of λ_1^μ, where the curves stay above the horizontal line at $P_k = 1$. For example, in the single user case, beamforming is optimal for λ_1^μ values to the left of point A and to the right of point B, while beamforming is not optimal for all λ_1^μ values between points A and B. In the second case, the region where beamforming is optimal first gets smaller by the addition of new users into the system, however it then starts to get larger as the number of users is increased further. In Figure 2.3, we plot one dimensional slices from the region corresponding to $K = 1, 2, 5, 10, 20, 30$. These

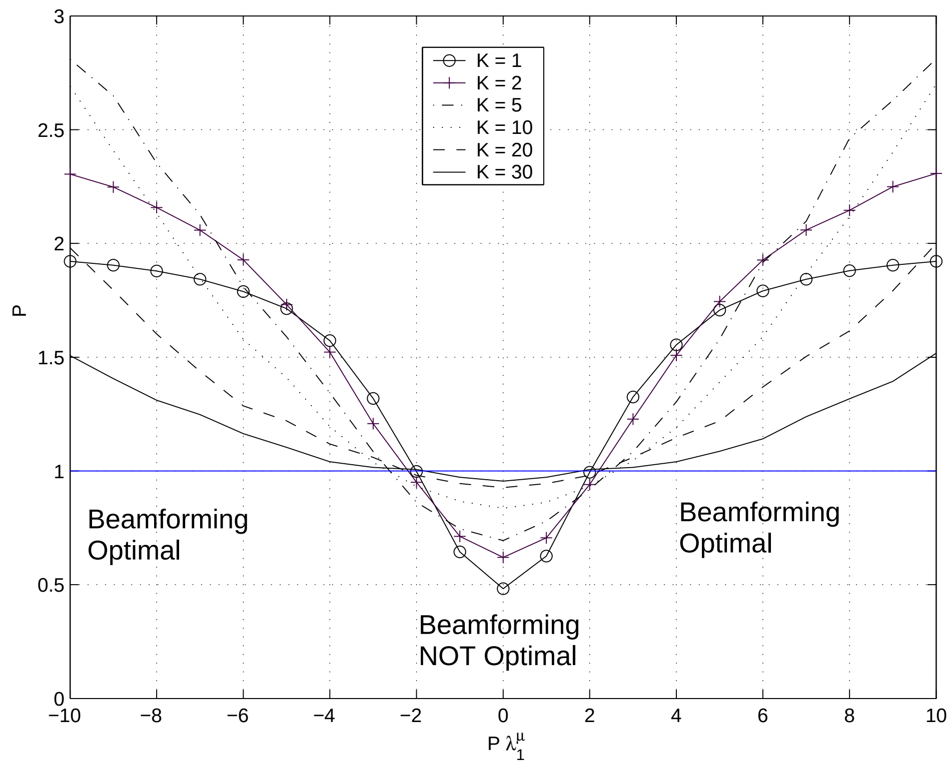


Figure 2.3: The region where beamforming is optimal for various numbers of users in the mean feedback model. This is an example where the region does not get larger as more users are added into the system.

curves give λ_1^μ values for beamforming to be optimal for a given power constraint, $P_k = 1$ for all k. The largest eigenvalues of all other users are kept constant, and each new user that is added to the system has a larger mean channel value than those of the users that are already in the system. The number of transmit and receive antennas is $n_T = n_R = 2$. In Figure 2.4, we zoom into the center of Figure 2.3 in order to show the details.

In light of these two examples, we conclude that, for the mean feedback model, the region where beamforming is optimal does not necessarily get larger with increasing number of users. However, we observe that as the number of users grows, the region starts to get larger regardless of the mean channel values of the users. The situation

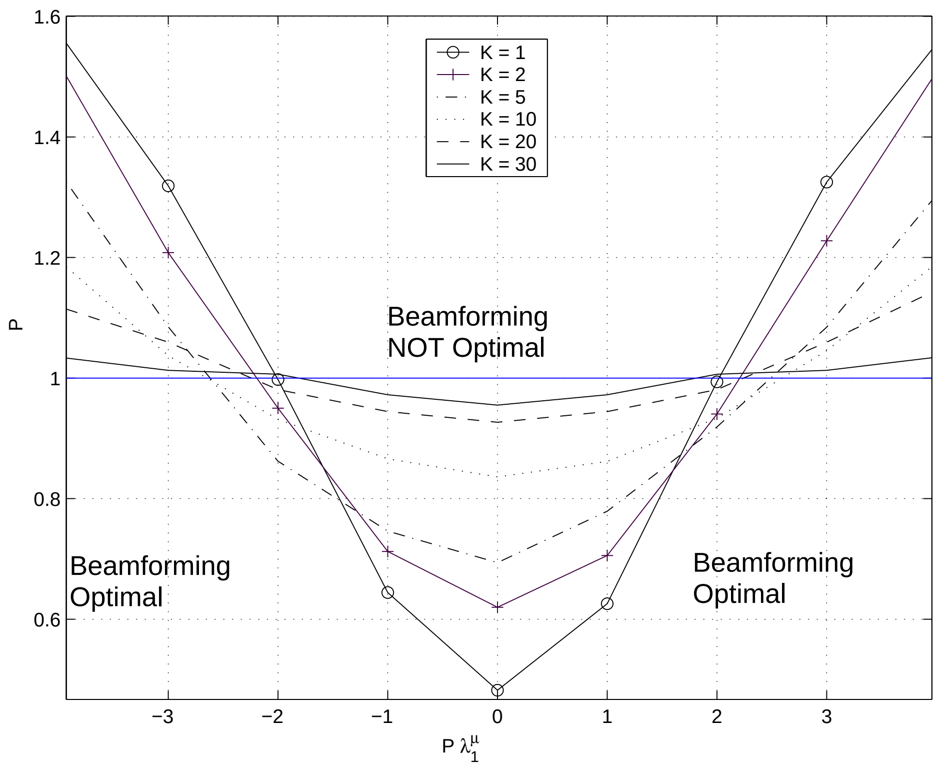


Figure 2.4: This is the same as Figure 2.3 where we zoom into the center of the figure to show details.

in Figures 2.3 and 2.4 is a worst case scenario. Even in this worst case, only the region corresponding to the first user does not get larger, while the corresponding regions for all other users in the system get larger. This possibly follows from the fact that the first user has the lowest mean channel value.

In Figure 2.5, we illustrate the change in the region where beamforming is optimal with the number of receive antennas for the covariance feedback model, while the number of transmit antennas is kept at $n_T = 2$. We observe that the region gets smaller as the number of receive antennas is increased. However, for a fixed number of receive antennas, the region grows with the number of users, and eventually equals the entire parameter region asymptotically as the number of users goes to infinity.

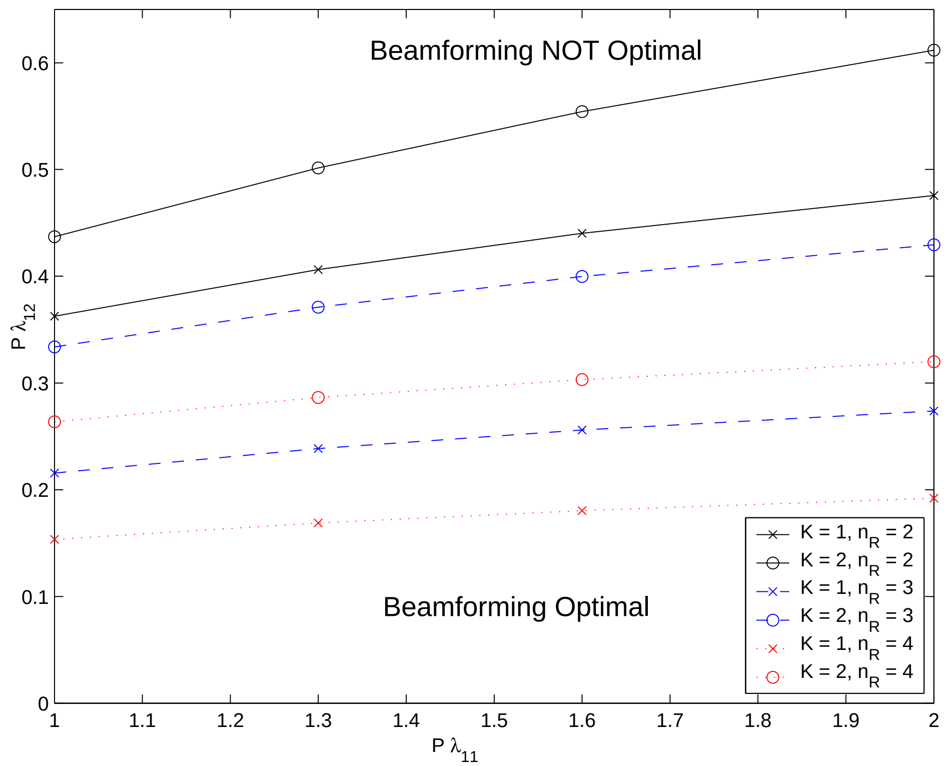


Figure 2.5: The region where beamforming is optimal for various numbers of receive antennas in the covariance feedback model.

For the asymptotic analysis, in Figure 2.6, we show three simple examples for different numbers of receive and transmit antennas. We plot the sum-rates resulting from optimal power allocation and arbitrary antenna selection schemes for the no CSI model. We observe that, for this instance, even for a small number of users, arbitrary antenna selection performs very close to the optimum power allocation scheme.

2.5 Conclusions

We determined the optimal transmit directions and the region where beamforming is optimal for all users under covariance and mean feedback CSI models for a multi-

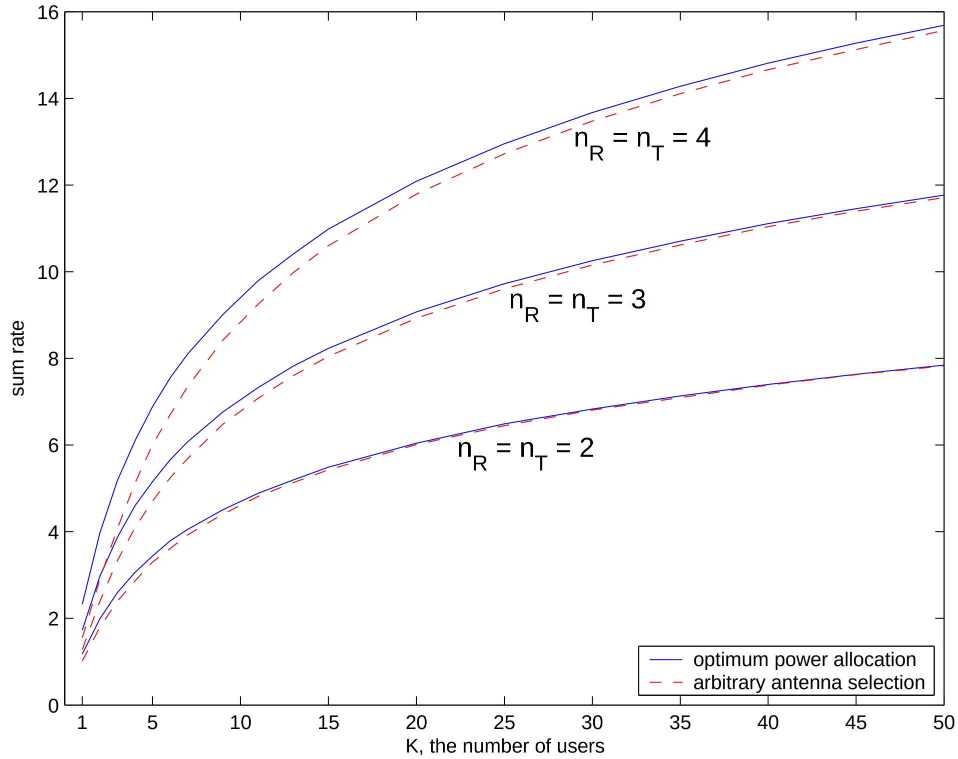


Figure 2.6: Sum-rates for the optimal and arbitrary antenna selection schemes, as a function of the number of users in the no CSI model.

user MIMO-MAC. We proved that the region where beamforming is optimal gets larger by the addition of new users into the system in the covariance feedback case. In the mean feedback case, we observed through simulations that the region where beamforming is optimal gets larger for relatively large numbers of users. We showed that in an asymptotically large system, beamforming is always optimal for all users not only for the covariance and the mean feedback cases, but also for the no CSI case as well. Combining our results with those of [31], we conclude that in a large multi-user MIMO-MAC system, beamforming is optimal under full, partial (covariance and mean), and no CSI cases.

The results in this chapter are published in [35], [36], [40].

2.6 Appendix

2.6.1 Proofs of Lemmas 1 and 2

Proof (Lemma 1): Using the matrix inversion lemma [12, page 19]

$$\mathbf{A}^{-1} = \mathbf{A}_k^{-1} - \mathbf{A}_k^{-1}\mathbf{z}_{k1}\left(\frac{1}{P_k\lambda_{k1}^{\Sigma}} + \mathbf{z}_{k1}^{\dagger}\mathbf{A}_k^{-1}\mathbf{z}_{k1}\right)^{-1}\mathbf{z}_{k1}^{\dagger}\mathbf{A}_k^{-1} \tag{2.68}$$

Multiplying this with $\mathbf{z}_{k1}^{\dagger}$ from left, and $\mathbf{z}_{k1}$ from right yields

$$\mathbf{z}_{k1}^{\dagger}\mathbf{A}^{-1}\mathbf{z}_{k1} = \frac{1}{P_k\lambda_{k1}^{\Sigma}}\left(1 - \frac{1}{1 + P_k\lambda_{k1}^{\Sigma}\mathbf{z}_{k1}^{\dagger}\mathbf{A}_k^{-1}\mathbf{z}_{k1}}\right) \tag{2.69}$$

By taking the expectation of both sides, (2.21) follows. In order to derive the identity in (2.22), note that

$$\lambda_{ki}^{\Sigma}E\left[\mathbf{z}_{ki}^{\dagger}\mathbf{A}^{-1}\mathbf{z}_{ki}\right] = \lambda_{ki}^{\Sigma}E\left[\mathrm{tr}(\mathbf{A}^{-1})\right] \tag{2.70}$$

since $\mathbf{z}_{k2}$ is independent of $\mathbf{A}$, and has identity-covariance. Applying the matrix inversion lemma [12, page 19] to $\mathbf{A} = \mathbf{I}_{n_R} + \bar{\mathbf{Z}}\tilde{\mathbf{\Lambda}}\bar{\mathbf{Z}}^{\dagger}$, with $\tilde{\mathbf{\Lambda}} = \bar{\mathbf{P}}\bar{\mathbf{\Lambda}}$

$$\mathbf{A}^{-1} = \mathbf{I}_{n_R} - \bar{\mathbf{Z}}\tilde{\mathbf{\Lambda}}^{1/2}(\mathbf{I}_K + \tilde{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{Z}}^{\dagger}\bar{\mathbf{Z}}\tilde{\mathbf{\Lambda}}^{1/2})^{-1}\tilde{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{Z}}^{\dagger} \tag{2.71}$$

Calculating the traces of both sides, we obtain

$$\begin{aligned}\operatorname{tr}(\mathbf{A}^{-1}) &= n_R - \operatorname{tr}\left((\mathbf{I}_K + \tilde{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{Z}}^{\dagger}\bar{\mathbf{Z}}\tilde{\mathbf{\Lambda}}^{1/2})^{-1}(\mathbf{I}_K + \tilde{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{Z}}^{\dagger}\bar{\mathbf{Z}}\tilde{\mathbf{\Lambda}}^{1/2} - \mathbf{I}_K)\right) && (2.72)\\ &= n_R - K + \operatorname{tr}\left((\mathbf{I}_K + \tilde{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{Z}}^{\dagger}\bar{\mathbf{Z}}\tilde{\mathbf{\Lambda}}^{1/2})^{-1}\right) && (2.73)\\ &= n_R - K + \frac{\sum_{k=1}^{K}|\mathbf{A}_k|}{|\mathbf{A}|} && (2.74)\end{aligned}$$

where in the last equation, we used the definition of an inverse of a matrix [12, page 21] and $|\mathbf{I}+\mathbf{AB}| = |\mathbf{I}+\mathbf{BA}|$. Noting that $\frac{|\mathbf{A}_k|}{|\mathbf{A}|} = \frac{1}{1+P_k\lambda_{k1}^{\Sigma}\mathbf{z}_{k1}^{\dagger}\mathbf{A}_k^{-1}\mathbf{z}_{k1}}$, and taking the expectations of both sides, we have (2.22). $\square$

Proof (Lemma 2): We will apply a version of the SLLN from [32, page 27, Theorem D]. In this version of the SLLN, the sum of a sequence of independent random variables with different means and variances converges to the sum of the sequence of means of the random variables, subject to the condition that $\sum_i^N \frac{\sigma_i^2}{i^2}$ converges, where σ_i^2 are the variances of the random variables. We will apply this theorem to every element of the matrix at hand, that is, to $\sum_{i=1}^N \alpha_i x_{ik}x_{ij}^*$, for all (k,j). In order to invoke the theorem, we let $\alpha_i x_{ik}x_{ij}^*$, for all i, be the sequence of independent but not identically distributed random variables. Note that the expectations of the diagonal elements are α_i, and the expectations of the off-diagonal elements are zero. Since α_i are assumed to be bounded and $\mathbf{x}_i$ have zero-mean and identity covariance, $\sum_i^N \frac{\alpha_i^2 E[x_{ik}^2 x_{ij}^2]}{i^2}$ converges, for all (k,j). As a result, we have $\sum_{i=1}^N \alpha_i\mathbf{x}_i\mathbf{x}_i^{\dagger} \to \sum_{i=1}^N \alpha_i\mathbf{I}_M$.

Due to [32, page 24], if a random variable converges to a deterministic number, a, then a function, f, of that random variable converges to $f(a)$. Therefore,

$$\log \left| \mathbf{I}_M + \sum_{i=1}^{N} \alpha_i \mathbf{x}_i \mathbf{x}_i^\dagger \right| \rightarrow M \log \left(1 + \sum_{i=1}^{N} \alpha_i \right), \qquad \text{a.s.} \tag{2.75}$$

If a random variable converges to a number almost surely, then the expectation of that random variable will be equal to the same number (for large N). That is,

$$E \left[\log \left| \mathbf{I}_M + \sum_{i=1}^{N} \alpha_i \mathbf{x}_i \mathbf{x}_i^\dagger \right| \right] \doteq M \log \left(1 + \sum_{i=1}^{N} \alpha_i \right). \tag{2.76}$$

$\square$

2.6.2 General Receive Antenna Correlation Matrix

In the model that we considered in the main part of this chapter, the receiver side correlation matrix is the identity matrix as a result of the assumption that the receiver (e.g., a base station) is not physically limited and one can place the antenna elements sufficiently far away from each other. In a different physical model with receiver side correlation present in the system, similar results can be found. For the single-user scenario, it is already known that the transmit directions are still the eigenvectors of the transmitter side channel correlation matrix, even when there is receiver side channel correlation in the system [18]. Beamforming optimality condition for this case is also found previously [18]. For the multi-user scenario, our approach generalizes to the case where there is receiver side channel correlation in the system, when the receiver side channel correlation matrices of all users are the same. In this case, the channel is modeled as,

$$\mathbf{H}_k = \mathbf{\Phi}^{1/2}\mathbf{Z}_k\mathbf{\Sigma}_k^{1/2} \tag{2.77}$$

where the receive antenna correlation matrix, $\mathbf{\Phi}$, is the correlation between the signals received at the n_R receive antennas of the receiver. This correlation matrix does not depend on the specific transmit antenna from which the signal is transmitted [4]. In a MAC, since we have a single receiver and the correlation matrix does not depend on the transmitters, we have the same $\mathbf{\Phi}$ for all users. The transmit antenna correlation matrix, $\mathbf{\Sigma}_k$, is the correlation between the signals transmitted from the n_T transmit antennas of user k.

The result of Theorem 1, which states that the optimal transmit directions are the eigenvectors of the transmit antenna correlation matrix, remains exactly the same. The result of Theorem 2, which states the conditions under which beamforming is optimal for all users, changes slightly. The refined conditions involve the eigenvalues of the receive correlation matrix. The region formed by these refined conditions still grows with the addition of a new user, and therefore, Theorem 3 also remains exactly the same. Finally, the asymptotic sum-capacity expression in Theorem 7 changes slightly and involves the eigenvalues of the receive correlation matrix. Below, we outline the reasons that Theorems 1 and 3 remain the same, and state the refined conditions for the optimality of beamforming and the new asymptotic sum-capacity expression.

Proof (Theorem 1): Let $\mathbf{\Phi} = \mathbf{U}_\Phi \mathbf{\Lambda}_\Phi^{1/2} \mathbf{U}_\Phi^\dagger$ be the spectral decomposition of the receive antenna correlation matrix. Inserting this into (2.2), we have

$$\mathbf{H}_k = \mathbf{U}_\Phi \mathbf{\Lambda}_\Phi^{1/2} \mathbf{U}_\Phi^\dagger \mathbf{Z}_k \mathbf{U}_{\Sigma_k} \mathbf{\Lambda}_{\Sigma_k}^{1/2} \mathbf{U}_{\Sigma_k}^\dagger \tag{2.78}$$

Then, inserting (2.78) into (2.6) and following similar lines to [18], we obtain

$$C_{sum} = \max_{\substack{\mathrm{tr}(\mathbf{Q}_k) \le P_k, \mathbf{Q}_k \succeq \mathbf{0} \\ k=1,\dots,K}} E\left[\log\left|\mathbf{I}_{n_R} + \mathbf{\Lambda}_\Phi \sum_{k=1}^{K} \mathbf{Z}_k \mathbf{\Lambda}_{Q_k} \mathbf{\Lambda}_{\Sigma_k} \mathbf{Z}_k^\dagger\right|\right] \tag{2.79}$$

The only difference between the proofs of Theorem 1 in uncorrelated and correlated receiver structures is that, here, we have the matrix $\mathbf{\Lambda}_\Phi$ in front of the summation term inside the logarithm compared to (2.18), which does not affect the derivations. Therefore, we observe that, even when the receive antenna correlation matrix is not equal to identity, the transmit directions of all users continue to depend only on their own transmit antenna correlation matrices. However, the resulting sum-capacity is different, and the optimal power allocation will depend on the eigenvalues of the receive antenna correlation matrix. $\square$

The sum-capacity expression in this case can be written, similar to (2.19), as

$$C_{sum} = \max_{\substack{\sum_{i=1}^{n_T} \lambda_{ki}^Q \le P_k, \lambda_{ki}^Q \ge 0 \\ k=1,\dots,K}} E\left[\log\left|\mathbf{I}_{n_R} + \sum_{k=1}^{K} \sum_{i=1}^{n_T} \lambda_{ki}^Q \lambda_{ki}^\Sigma \tilde{\mathbf{z}}_{ki} \tilde{\mathbf{z}}_{ki}^\dagger\right|\right] \tag{2.80}$$

where $\{\tilde{\mathbf{z}}_{ki} = \mathbf{\Lambda}_{\Phi}^{1/2}\mathbf{z}_{ki}, k = 1, \ldots, K, i = 1, \ldots, n_T\}$ is a set of i.i.d. Gaussian random vectors with zero-mean and covariance matrix $\mathbf{\Lambda}_{\Phi}$.

Proof (Theorem 2): After taking the derivative of the Lagrangian for the optimization problem in (2.80), the conditions for the optimality of beamforming for all users become

$$\frac{\tilde{E}_{k1}}{\tilde{E}_{ki}} = \frac{\lambda_{k1}^{\Sigma} E\left[\tilde{\mathbf{z}}_{k1}^{\dagger}\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{z}}_{k1}\right]}{\lambda_{ki}^{\Sigma} E\left[\tilde{\mathbf{z}}_{ki}^{\dagger}\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{z}}_{ki}\right]} > 1, \quad \forall i \neq 1, \quad k = 1, \ldots, K \tag{2.81}$$

where $\tilde{\mathbf{A}} = \mathbf{I}_{n_R} + \sum_{l=1}^{K} P_l \lambda_{l1}^{\Sigma} \tilde{\mathbf{z}}_{l1}\tilde{\mathbf{z}}_{l1}^{\dagger}$. The identities in Lemma 1 change slightly for the general $\mathbf{\Phi}$ case. The details of the derivations only require matrix algebra and are omitted here due to space limitations. Inserting the new identities from Lemma 1 into (2.81), we have

$$P\lambda_{k2}^{\Sigma} < \frac{1 - E\left[\frac{1}{1+P_k\lambda_{k1}^{\Sigma}\tilde{\mathbf{z}}_{k1}^{T}\tilde{\mathbf{A}}_k^{-1}\tilde{\mathbf{z}}_{k1}}\right]}{\sum_{i=1}^{n_R}\lambda_i^{\Phi} - \sum_{l=1}^{K} E\left[\frac{P_l\lambda_{l1}^{\Sigma}\tilde{\mathbf{z}}_{l1}^{\dagger}\mathbf{\Lambda}_{\Phi}^{1/2}\tilde{\mathbf{A}}_l^{-1}\mathbf{\Lambda}_{\Phi}^{1/2}\tilde{\mathbf{z}}_{l1}}{1+P_l\lambda_{l1}^{\Sigma}\tilde{\mathbf{z}}_{l1}^{T}\tilde{\mathbf{A}}_l^{-1}\tilde{\mathbf{z}}_{l1}}\right]}, \quad k = 1, \ldots, K \tag{2.82}$$

where $\tilde{\mathbf{A}}_k = \tilde{\mathbf{A}} - P_k\lambda_{k1}^{\Sigma}\tilde{\mathbf{z}}_{k1}\tilde{\mathbf{z}}_{k1}^{\dagger}$, for all k. $\square$

Inserting $\mathbf{\Phi} = \mathbf{I}$, and adding and subtracting "1" from the numerator of the expectation term in the denominator of (2.82), we get (2.23). And, inserting $K = 1$, and $\tilde{\mathbf{A}}_k = \mathbf{I}$, for all k, in (2.82), we get the single-user condition derived in [18].

Proof (Theorem 3): The proof exactly follows the original proof. We only use the fact that $\{\tilde{\mathbf{z}}_{ki}\}$ are independent and zero-mean random vectors. $\square$

Proof (Theorem 7): Applying Lemma 2 to (2.80), the objective function in (2.62) changes to a summation of log functions, instead of $n_R \log(\cdot)$. Using the Lagrangian method, (2.65) becomes

$$C_{sum} \doteq \sum_{i=1}^{n_R} \log\left(1 + \lambda_i^{\Phi} \sum_{k=1}^{K} P_k \lambda_{k1}^{\Sigma}\right) \tag{2.83}$$

$\square$

Chapter 3

Optimum Power Allocation Policies

In Gaussian MIMO multiple access systems, when the receiver side has perfect CSI, and the transmitter side has partial CSI, our goal is to find the optimum transmit covariance matrices of the users, or equivalently to find the optimum transmit directions and the optimum power allocation policies. In Chapter 2, we found the optimum transmit directions, however, we only focused on a special case of the optimum power allocation problem, which was in the form of beamforming. In this chapter, we consider the general power allocation problem, and find the eigenvalues of the transmit covariance matrices, both in the single-user and multi-user cases, when the transmitters have partial CSI in the form of covariance feedback.

In a single-user MIMO system, when both the receiver and the transmitter have perfect CSI and the channel is fixed, [42] showed that the optimum transmit directions are the right singular vectors of the deterministic channel matrix, and the optimum power allocation policy is to water-fill over the singular values of the deterministic channel matrix. In a multi-user MIMO system, when both the receiver and the transmitters have perfect CSI and the channel is fixed, [49] showed that the optimum transmit directions and the power allocation policies can be found using an iterative algorithm that updates the transmit directions and the power allocation policy of one user at a time. When the channel is changing over time due to fading, and perfect and instantaneous CSI is known both at the receiver and at the transmitter side, these solutions extend to water-filling over both the antennas and the channel states in single-user [42], and multi-user [50] MIMO systems. However, in most of the wireless communication scenarios, especially in wireless MIMO communications, it is unrealistic to assume that the transmitter side has the perfect knowledge of the instantaneous CSI. In such scenarios, it might be more realistic to assume that only the receiver side can perfectly estimate the instantaneous CSI, while the transmitter side has only a statistical knowledge of the channel.

When the fading in the channel is assumed to be a Gaussian process, statistics of the channel reduce to the mean and covariance information of the channel. The problem in this setting as well is to find the optimum transmit covariance matrices, i.e., the optimum transmit directions and the optimum power allocation policies. However, in this case, the transmit directions and the power allocations are not functions of the channel states, but they are functions of the statistics of the channel states, that are fed by the receiver back to the transmitters. The optimization criteria that we consider are the maximum rate in a single-user system, and the maximum sum-rate in a multi-user system. For the covariance feedback case, it was shown in [46] for a MISO system, and in [3,14] for a MIMO system that the optimal transmit covariance matrix and the channel covariance matrix have the same eigenvectors, i.e., the optimal transmit directions are the eigenvectors of the channel covariance matrix. For the mean feedback case, the eigenvectors of the optimal transmit covariance matrix were shown to be the same as the right singular vectors of the channel mean matrix for a MISO system in [46] and for a MIMO system in [14]. In Chapter 2, we generalized these results, both in covariance and mean feedback cases, to MIMO-MAC systems. We showed that in a MIMO-MAC system with partial CSI at the transmitters, all users should transmit in the direction of the eigenvectors of their *own* channel parameter matrices. Consequently, we showed that, the transmit directions of the users in a MIMO-MAC with partial CSI at the transmitters are independent of the presence of other users, and therefore, that the users maintain their single-user transmit direction strategies even in a multi-user scenario.

On the other hand, in this aforementioned literature, the optimization of the eigenvalues of the transmit covariance matrices, i.e., the power allocation policies, are left as additional optimization problems. The optimum eigenvalues are known only for specific conditions, called *beamforming optimality conditions*. If the channel statistics satisfy these conditions, then unit-rank transmit covariance matrices are optimum for all users, i.e., users allocate all of their powers to the direction of their strongest eigenvectors.

Although having beamforming optimality conditions is extremely helpful, as we have shown in Chapter 2, beamforming is unconditionally optimal only when the number of users grows to infinity in a fading multi-user MIMO setting when partial

CSI is available at the transmitters. In a single-user MIMO or in a MIMO-MAC with finite number of users, the channel statistics might be such that beamforming may never be optimal. For such scenarios, efficient and globally convergent algorithms are needed in order to solve for the optimum eigenvalues of the transmit covariance matrices. References [17], [44, 45] proposed algorithms that solve this problem for a single-user MISO system, and for a single-user MIMO system, respectively. However, in both cases, the convergence proofs for these algorithms were not provided. In a MIMO-MAC scenario with partial CSI available at the transmitters, no algorithm was available to find the optimum eigenvalues in a multi-user setting.

In this chapter, first, we give an alternative derivation for the algorithm proposed in [44, 45] for a single-user MIMO system by enforcing the KKT optimality conditions at each iteration. Our main contribution in this chapter is to prove that the convergence point of this algorithm is unique and is equal to the optimum eigenvalue allocation. We showed that the proposed algorithm converges to this unique point starting from any point on the space of feasible eigenvalues. Next, we consider the multi-user version of the problem. In this case, our contribution is to develop an iterative algorithm that finds the optimum eigenvalues of the transmit covariance matrices of all users that maximize the sum-rate of the MIMO-MAC system. We apply the single-user algorithm iteratively to reach the global optimum point. At any given iteration, the multi-user algorithm updates the eigenvalues of one user, using the algorithm proposed for the single-user case, assuming that the eigenvalues of the remaining users are fixed. The algorithm iterates over all users in a round-robin fashion. We prove that, this algorithm converges to the unique global optimum power allocation for all users.

3.1 System Model

The system model that we consider in this chapter is the same as in Chapter 2. We summarize our model here for completeness purposes. The channel between user k and the receiver is represented by a random matrix $\mathbf{H}_k$ with dimensions of $n_R \times n_T$. The receiver has the perfect knowledge of the channel, while the transmitters have only the statistical model of the channel. Each transmitter sends a vector $\mathbf{x}_k$, and the received vector is

$$\mathbf{r} = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{x}_k + \mathbf{n} \tag{3.1}$$

where K is the number of users, $\mathbf{n}$ is a zero-mean, identity-covariance complex Gaussian vector, and the entries of $\mathbf{H}_k$ are complex Gaussian random variables. Let $\mathbf{Q}_k = E[\mathbf{x}_k \mathbf{x}_k^\dagger]$ be the transmit covariance matrix of user k, which has an average power constraint of P_k, $\mathrm{tr}(\mathbf{Q}_k) \leq P_k$.

The statistical model that we consider in this chapter is the "partial CSI with covariance feedback" model where each transmitter knows the channel covariance information of all transmitters, in addition to the distribution of the channel. In this model, there exists correlation between the signals transmitted by or received at different antenna elements. However, we 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[1]. As a result, the channel of user k is written as [4]

$$\mathbf{H}_k = \mathbf{Z}_k \mathbf{\Sigma}_k^{1/2} \tag{3.2}$$

where the entries of $\mathbf{Z}_k$ are i.i.d., zero-mean, unit-variance complex Gaussian random variables. From this point on, we will refer to matrix $\mathbf{\Sigma}_k$ as the channel covariance feedback matrix of user k. Similar covariance feedback models have been used in [3], [14], [17], [46].

3.2 Power Allocation for Single-User MIMO

In this section, we will assume that $K = 1$. In a single-user system with partial CSI in the form of the channel covariance matrix at the transmitter, the optimization problem is that of choosing a transmit covariance matrix $\mathbf{Q}$, which is subject to a trace constraint representing the average transmit power constraint,

$$C = \max_{\mathrm{tr}(\mathbf{Q}) \leq P} E\left[\log |\, \mathbf{I}_{n_R} + \mathbf{H}\mathbf{Q}\mathbf{H}^{\dagger}|\right] \tag{3.3}$$

where we note that the cost function of the optimization problem in (3.3) is concave in $\mathbf{Q}$ and the constraint set is convex.

[1]We refer the reader to Section 3.7, for a discussion on extending our results to the case where the channel has double-sided correlation structure, i.e., to the case where the signals arriving at the receiver are correlated as well.

The channel covariance matrix $\mathbf{\Sigma}$, which is known at the transmitter, has the eigenvalue decomposition $\mathbf{\Sigma} = \mathbf{U}_\Sigma \mathbf{\Lambda}_\Sigma \mathbf{U}_\Sigma^\dagger$ with unitary $\mathbf{U}_\Sigma$ and diagonal $\mathbf{\Lambda}_\Sigma$ of ordered eigenvalues. The transmit covariance matrix $\mathbf{Q}$ has the eigenvalue decomposition $\mathbf{Q} = \mathbf{U}_Q \mathbf{\Lambda}_Q \mathbf{U}_Q^\dagger$ with unitary $\mathbf{U}_Q$ and diagonal $\mathbf{\Lambda}_Q$. It has been shown that the eigenvectors of the optimum transmit covariance matrix must be equal to the eigenvectors of the channel covariance matrix, i.e., $\mathbf{U}_Q = \mathbf{U}_\Sigma$ [14]. By inserting this into (3.3), and using the fact that the random matrices $\mathbf{Z}\mathbf{U}_\Sigma$ and $\mathbf{Z}$ have the same probability distribution for zero-mean identity-covariance Gaussian $\mathbf{Z}$ and unitary $\mathbf{U}_\Sigma$ [42], we get

$$C = \max_{\mathrm{tr}(\mathbf{\Lambda}_Q) \le P} E\left[\log |\, \mathbf{I}_{n_R} + \mathbf{Z}\mathbf{U}_\Sigma \mathbf{\Lambda}_\Sigma^{1/2} \mathbf{U}_\Sigma^\dagger \mathbf{U}_Q \mathbf{\Lambda}_Q \mathbf{U}_Q^\dagger \mathbf{U}_\Sigma \mathbf{\Lambda}_\Sigma^{1/2} \mathbf{U}_\Sigma^\dagger \mathbf{Z}^\dagger |\right] \tag{3.4}$$

$$= \max_{\mathrm{tr}(\mathbf{\Lambda}_Q) \le P} E\left[\log |\, \mathbf{I}_{n_R} + \mathbf{Z}\mathbf{U}_\Sigma \mathbf{\Lambda}_Q \mathbf{\Lambda}_\Sigma \mathbf{U}_\Sigma^\dagger \mathbf{Z}^\dagger |\right] \tag{3.5}$$

$$= \max_{\mathrm{tr}(\mathbf{\Lambda}_Q) \le P} E\left[\log |\, \mathbf{I}_{n_R} + \mathbf{Z} \mathbf{\Lambda}_Q \mathbf{\Lambda}_\Sigma \mathbf{Z}^\dagger |\right] \tag{3.6}$$

$$= \max_{\sum_{i=1}^{n_T} \lambda_i^Q \le P} E\left[\log \left| \mathbf{I}_{n_R} + \sum_{i=1}^{n_T} \lambda_i^Q \lambda_i^\Sigma \mathbf{z}_i \mathbf{z}_i^\dagger \right|\right] \tag{3.7}$$

where $\mathbf{z}_i$ is the i^{th} column of $\mathbf{Z}$, i.e., $\{\mathbf{z}_i, i = 1, \ldots, n_T\}$ is a set of $n_R \times 1$ dimensional i.i.d., zero-mean, identity-covariance Gaussian random vectors. The Lagrangian for the above optimization problem is,

$$L = E\left[\log \left| \mathbf{I}_{n_R} + \sum_{i=1}^{n_T} \lambda_i^Q \lambda_i^\Sigma \mathbf{z}_i \mathbf{z}_i^\dagger \right|\right] - \mu \left(\sum_{i=1}^{n_T} \lambda_i^Q - P \right) \tag{3.8}$$

where μ is the Lagrange multiplier. Using the identity in (2.25), the KKT conditions can be written as

$$\lambda_i^{\Sigma} E\left[\mathbf{z}_i^{\dagger}\left(\mathbf{I}_{n_R}+\sum_{j=1}^{n_T}\lambda_j^Q\lambda_j^{\Sigma}\mathbf{z}_j\mathbf{z}_j^{\dagger}\right)^{-1}\mathbf{z}_i\right]\leq\mu,\quad i=1,\ldots,n_T \tag{3.9}$$

Defining $\mathbf{A}=\mathbf{I}_{n_R}+\sum_{j=1}^{n_T}\lambda_j^Q\lambda_j^{\Sigma}\mathbf{z}_j\mathbf{z}_j^{\dagger}$, and $\mathbf{A}_i=\mathbf{A}-\lambda_i^Q\lambda_i^{\Sigma}\mathbf{z}_i\mathbf{z}_i^{\dagger}$, and using the matrix inversion lemma [12, page 19], we get

$$E_i(\boldsymbol{\lambda}^Q)\triangleq E\left[\frac{\lambda_i^{\Sigma}\mathbf{z}_i^{\dagger}\mathbf{A}_i^{-1}\mathbf{z}_i}{1+\lambda_i^Q\lambda_i^{\Sigma}\mathbf{z}_i^{\dagger}\mathbf{A}_i^{-1}\mathbf{z}_i}\right]\leq\mu,\quad i=1,\ldots,n_T \tag{3.10}$$

where we defined the left hand side of (3.10) as $E_i(\boldsymbol{\lambda}^Q)$. The i^{th} inequality in (3.10) is satisfied with equality whenever the optimum λ_i^Q is non-zero, and with strict inequality whenever the optimum λ_i^Q is zero. We note that in classical water-filling solutions, since the channel is either fixed or known instantaneously at the transmitter, the corresponding KKT conditions do not involve an expectation, and therefore, non-zero λ_i^Q's can be solved for in terms of the Lagrange multiplier and the eigenvalues of the fixed/instantaneous channel matrix. However, in our case, we cannot directly solve for λ_i^Q in (3.10). Instead, we multiply both sides of (3.10) by λ_i^Q,

$$\lambda_i^Q E_i(\boldsymbol{\lambda}^Q)=\mu\lambda_i^Q,\quad i=1,\ldots,n_T \tag{3.11}$$

We note that when $\lambda_i^Q=0$, both sides of (3.11) are equal to zero. Therefore, unlike (3.10), (3.11) is always satisfied with equality for optimum eigenvalues. By summing both sides over all antennas, we find μ, and by substituting this μ into (3.11), we find the fixed point equations which have to be satisfied by the optimum eigenvalues,

$$\lambda_i^Q = \frac{\lambda_i^Q E_i(\boldsymbol{\lambda}^Q)}{\sum_{j=1}^{n_T} \lambda_j^Q E_j(\boldsymbol{\lambda}^Q)} P = \frac{P}{\sum_j \frac{\lambda_j^Q E_j(\boldsymbol{\lambda}^Q)}{\lambda_i^Q E_i(\boldsymbol{\lambda}^Q)}} \triangleq f_i(\boldsymbol{\lambda}^Q), \quad i = 1, \ldots, n_T \tag{3.12}$$

where $\boldsymbol{\lambda}^Q = [\lambda_1^Q, \ldots, \lambda_{n_T}^Q]$, and we defined the right hand side of (3.12) which depends on all of the eigenvalues as $f_i(\boldsymbol{\lambda}^Q)$. It is important to emphasize that the optimum solution of the KKT conditions always satisfies the fixed point in (3.12), even if the optimum solution has some zero components.

We propose to use the following fixed point algorithm

$$\boldsymbol{\lambda}^Q(n+1) = \mathbf{f}(\boldsymbol{\lambda}^Q(n)) \tag{3.13}$$

where $\mathbf{f} = [f_1, \ldots, f_{n_T}]$. In order to solve for the optimum eigenvalues, (3.13) updates the eigenvalues at step $n+1$ as a function of the eigenvalues at step n. We claim that this algorithm converges and that the unique stable fixed point of the algorithm is equal to the optimum eigenvalues. Although this algorithm is the same as the one proposed in [44, 45], here, we also provide a convergence proof.

3.3 Convergence Proof

As stated in (3.7), the constraint set of the optimization problem is $\sum_{i=1}^{n} \lambda_i^Q \leq P$. We know that the optimum value is obtained when the summation is equal to P. If the summation was strictly less than P, we could increase the value of the objective function by increasing any one of the λ_i^Q's, while keeping the rest fixed. Therefore, the constraint set becomes $\sum_{i=1}^{n} \lambda_i^Q = P$. This equality defines a simplex in the n_T-dimensional space (see Figure 3.1), and all feasible eigenvalue vectors are located on this simplex. Note that if the algorithm is initiated at an exact corner point of the simplex, then the updates stay at the same point indefinitely. The reason for this is that while we obtain (3.11) from (3.10), we create some artificial fixed points. That is, although some non-optimum $\lambda_i^Q = 0$ does not satisfy (3.10) with equality, the same non-optimum $\lambda_i^Q = 0$ always satisfies (3.11) with equality.

As a result, in addition to the point that is the solution of the KKT conditions, the solution set of the fixed point equation in (3.12) includes some artificial fixed points. Since our optimization problem is concave and the constraint set is convex, the solution of the KKT conditions is the unique optimum point of the optimization problem. On the other hand, artificial fixed points are the solutions to some reduced optimization problems, which are obtained by forcing some of the components of the power allocation vector to be zero. When we force a choice of $n_T - 1$ components to be zero, we can find one optimum solution to the corresponding reduced optimization problem for each choice. Since there are $\binom{n_T}{n_T-1}$ ways of choosing zero components, this adds n_T artificial fixed points, which are the corner points of the simplex, to the solution set of the fixed point equation. Similarly, when we force a choice of $n_T - 2$ components to be zero, we can find one optimum solution to the corresponding reduced optimization problem for each choice. This adds $\binom{n_T}{n_T-2}$ artificial fixed points to the solution set of the fixed point equation. By counting all possibilities, we find that there are a total of $2^{n_T} - 2$ artificial fixed points. However, it is important to note that one of these counted points might be the optimum solution of the KKT conditions, if there are some zero components in the optimum eigenvalue vector. If the optimum eigenvalues are all non-zero, then the solution of the KKT conditions is different than these artificial fixed points. Therefore, we call a point an artificial fixed point only if it is not the optimum solution.

In this section, we will first prove that our algorithm converges. Then, we will prove that the algorithm cannot converge to an artificial fixed point, and therefore, the only point that the algorithm can converge to is the unique solution of the KKT conditions. The main ingredient of our convergence proof is the following lemma.

Lemma 3 *Let us have two feasible vectors on the simplex, $\boldsymbol{\lambda}^Q$ and $\bar{\boldsymbol{\lambda}}^Q$, such that $\lambda_i^Q > \bar{\lambda}_i^Q$, then $f_i(\boldsymbol{\lambda}^Q) > f_i(\bar{\boldsymbol{\lambda}}^Q)$.*

Proof: Note that $\lambda_i^Q > \bar{\lambda}_i^Q$ implies $\sum_{j\neq i} \lambda_j^Q < \sum_{j\neq i} \bar{\lambda}_j^Q$, since all λ_i^Q sum up to P. Therefore, the lemma can be proved equivalently by proving that $f_i(\boldsymbol{\lambda}^Q)$ is increasing in λ_i^Q when the rest of the λ_j^Q, $j \neq i$ are fixed, and $f_i(\boldsymbol{\lambda}^Q)$ is decreasing in $\sum_{j\neq i} \lambda_j^Q$, when λ_i^Q is fixed. The first part of the claim is easy to show. Consider (3.12), it can be shown that the partial derivative of $\lambda_i^Q E_i(\boldsymbol{\lambda}^Q)$ with respect to λ_i^Q is positive, and the partial derivatives of $\lambda_j^Q E_j(\boldsymbol{\lambda}^Q)$, for $j \neq i$, with respect to λ_i^Q are all negative. Therefore, $\frac{\lambda_j^Q E_j(\boldsymbol{\lambda}^Q)}{\lambda_i^Q E_i(\boldsymbol{\lambda}^Q)}$ is decreasing (for all j), and $f_i(\boldsymbol{\lambda}^Q)$ is increasing, in λ_i^Q when the rest of the λ_j^Q, $j \neq i$ are fixed. The second part of the claim is a little bit involved. In order to show that $f_i(\boldsymbol{\lambda}^Q)$ is decreasing in $\sum_{j\neq i} \lambda_j^Q$, we need to show that $\sum_{j\neq i} \frac{\partial f_i(\boldsymbol{\lambda}^Q)}{\partial \lambda_j^Q} < 0$. It is sufficient to show $\frac{\partial f_i(\boldsymbol{\lambda}^Q)}{\partial \lambda_j^Q} < 0$ for all $j \neq i$. In order to show this, consider (3.12), it is easy to show that the partial derivative of $\lambda_i^Q E_i(\boldsymbol{\lambda}^Q)$ with respect to λ_j is negative. We, then, need to show that $\frac{\partial(\sum_{k=1}^{n_T} \lambda_k^Q E_k(\boldsymbol{\lambda}^Q))}{\partial \lambda_j^Q} > 0$. We will give the proof of this in the Appendix, Section 3.7.1. $\square$

In Lemma 3, we showed the monotonicity property of the algorithm. By using this property, in the next lemma, we will show that the algorithm converges.

Lemma 4 *The algorithm in (3.13) converges to one of the points in the solution set of the fixed point equation in (3.12) when it is initiated at any arbitrary feasible point,* $\boldsymbol{\lambda}^Q(0)$*, that is not on the boundary of the simplex.*

Proof: After the first iteration of the algorithm, we have one of the following three cases for each λ_i^Q. The first case is that $\lambda_i^Q(1) = f_i(\boldsymbol{\lambda}^Q(0)) = \lambda_i^Q(0)$. This means that we have started the algorithm at the optimum point that solves the KKT conditions. Since all of the artificial fixed points are on the boundary of the simplex, this point cannot be an artificial fixed point.

The second case is that $\lambda_i^Q(1) = f_i(\boldsymbol{\lambda}^Q(0)) > \lambda_i^Q(0)$. In this case, by applying Lemma 3 repeatedly, we get $\lambda_i^Q(n) > \lambda_i^Q(n-1) > \cdots > \lambda_i^Q(1) > \lambda_i^Q(0)$. Since $\lambda_i^Q(n)$ is a monotonically increasing sequence and it is upper bounded, it is guaranteed to converge.

The third case is that $\lambda_i^Q(1) = f_i(\boldsymbol{\lambda}^Q(0)) < \lambda_i^Q(0)$. In this case, by applying Lemma 3 repeatedly, we get $\lambda_i^Q(n) < \lambda_i^Q(n-1) < \cdots < \lambda_i^Q(1) < \lambda_i^Q(0)$. Since $\lambda_i^Q(n)$

is a monotonically decreasing sequence and it is lower bounded, it is guaranteed to converge.

Finally, since each component of $\boldsymbol{\lambda}^Q$ converges, the vector itself also converges to a point inside the solution set of the fixed point equation. $\square$

Although we proved that the algorithm converges when it is initiated at any arbitrary feasible point that is not on the boundary of the simplex, there is a possibility that it converges to an artificial fixed point instead of the optimum solution of the KKT conditions. In the following lemma, we will show that this is never the case.

Lemma 5 *The artificial fixed points are unstable. For a very small and fixed ϵ, if we are ϵ away from an artificial fixed point, with one iteration of the algorithm, we will move further away from that artificial fixed point.*

Proof: The main idea of the proof is the following. We will start from an artificial fixed point that is not the optimum solution of the KKT conditions of the original optimization problem, and show that by perturbing this artificial fixed point by an ϵ amount, we move further away from that artificial fixed point. We give the proof of the most general scenario with n_T antennas and starting from any arbitrary artificial fixed point in the Appendix, Section 3.7.2. Here, we give the outline and the basic methodology of the general proof by considering a simple case where $n_T = 3$. In this case, we have $2^{n_T} - 2 = 6$ artificial fixed points. Three of them are the corner points of the 3-dimensional simplex. The other three of them lie on the boundary of the simplex, each point corresponding to a solution of the reduced optimization problem where one of the components is forced to be zero. These points can be seen in Figure 3.1.

Here, in this outline of the general proof, we will also assume that the artificial fixed point we focus on has only one zero component. In particular, we assume that we are at the artificial fixed point $\mathbf{p}_4 = (a, b, 0)$, see Figure 3.1. Since this is a fixed point, the following equalities hold from (3.12),

$$a = \frac{aE_1(\mathbf{p}_4)}{aE_1(\mathbf{p}_4) + bE_2(\mathbf{p}_4)}P, \qquad b = \frac{bE_2(\mathbf{p}_4)}{aE_1(\mathbf{p}_4) + bE_2(\mathbf{p}_4)}P \tag{3.14}$$

From above, we find that $aE_1(\mathbf{p}_4)+bE_2(\mathbf{p}_4) = PE_1(\mathbf{p}_4) = PE_2(\mathbf{p}_4)$. This is equivalent to saying that the KKT conditions of the reduced optimization problem corresponding to the first and second components are satisfied with equality, that is, $E_1(\mathbf{p}_4) = E_2(\mathbf{p}_4) = \mu'$. We call this Lagrange multiplier μ', because this is possibly different than the Lagrange multiplier of the original optimization problem. For $E_3(\mathbf{p}_4)$, we have three possibilities. $E_3(\mathbf{p}_4) = \mu'$ cannot hold, because that would mean that the third KKT condition is also satisfied with equality and this can only happen when optimal λ_3^Q is non-zero. $E_3(\mathbf{p}_4) < \mu'$ cannot hold, because that would mean that we satisfy all three KKT conditions of the original optimization problem with $\mu' = \mu$, and this fixed point is optimum. This contradicts with our assumption that we are at an artificial fixed point that is not the optimum solution of the original optimization problem. Therefore, the only possibility at an artificial fixed point is that $E_3(\mathbf{p}_4) > \mu'$.

Now, we will show that by perturbing this artificial fixed point by an ϵ amount, we move further away from this fixed point. We run the algorithm for $\mathbf{p}_4' = (a-\epsilon, b, \epsilon)$. We first calculate $E_1(\mathbf{p}_4')$,

$$E_1(\mathbf{p}_4') = E\left[\frac{\lambda_1^\Sigma \mathbf{z}_1^\dagger (\mathbf{I}_{n_R} + (a-\epsilon)\lambda_1^\Sigma \mathbf{z}_1\mathbf{z}_1^\dagger + b\lambda_2^\Sigma \mathbf{z}_2\mathbf{z}_2^\dagger)^{-1}\mathbf{z}_1}{1+(a-\epsilon)\lambda_1^\Sigma \mathbf{z}_1^\dagger (\mathbf{I}_{n_R} + (a-\epsilon)\lambda_1^\Sigma \mathbf{z}_1\mathbf{z}_1^\dagger + b\lambda_2^\Sigma \mathbf{z}_2\mathbf{z}_2^\dagger)^{-1}\mathbf{z}_1}\right] \tag{3.15}$$

Let us first look at the expression in the numerator of (3.15). We consider this as a function, $h(x) = \lambda_1^\Sigma \mathbf{z}_1^\dagger \left(\mathbf{A}_{p_4} - x(\lambda_1^\Sigma \mathbf{z}_1\mathbf{z}_1^\dagger)\right)^{-1} \mathbf{z}_1$ evaluated at $x = \epsilon$, where $\mathbf{A}_{p_4} = \mathbf{I}_{n_R} + a\lambda_1^\Sigma \mathbf{z}_1\mathbf{z}_1^\dagger + b\lambda_2^\Sigma \mathbf{z}_2\mathbf{z}_2^\dagger$. Using the matrix inversion lemma [12, page 19], we get $h(x) = \frac{\lambda_1^\Sigma \mathbf{z}_1^\dagger \mathbf{A}_{p_4}\mathbf{z}_1}{1-x\lambda_1^\Sigma \mathbf{z}_1^\dagger \mathbf{A}_{p_4}\mathbf{z}_1}$, and using the Taylor series expansion formula around $x = 0$, we obtain

$$h(\epsilon) = \lambda_1^\Sigma \mathbf{z}_1^\dagger \mathbf{A}_{p_4}^{-1}\mathbf{z}_1 + \epsilon(\lambda_1^\Sigma \mathbf{z}_1^\dagger \mathbf{A}_{p_4}^{-1}\mathbf{z}_1)^2 + \epsilon^2(\lambda_1^\Sigma \mathbf{z}_1^\dagger \mathbf{A}_{p_4}^{-1}\mathbf{z}_1)^3 + \ldots \tag{3.16}$$

$$= \lambda_1^\Sigma \mathbf{z}_1^\dagger \mathbf{A}_{p_4}^{-1}\mathbf{z}_1 + O(\epsilon) \tag{3.17}$$

where $O(\epsilon)$ is used to describe an asymptotic upper bound for the magnitude of the residual in terms of ϵ. Mathematically, a function, $\bar{h}(\epsilon)$ is order $O(\epsilon)$ as $\epsilon \to 0$ if and only if $0 < \limsup_{\epsilon\to 0} \frac{\bar{h}(\epsilon)}{\epsilon} < \infty$ [24]. Now, when we insert this into (3.15), we obtain

$$E_1(\mathbf{p}_4') = E\left[\frac{\lambda_1^\Sigma \mathbf{z}_1^\dagger \mathbf{A}_{p_4}^{-1}\mathbf{z}_1 + O(\epsilon)}{1 + a\lambda_1^\Sigma \mathbf{z}_1^\dagger \mathbf{A}_{p_4}^{-1}\mathbf{z}_1 + O(\epsilon)}\right] \tag{3.18}$$

$$= \frac{1}{a}\left(1 - E\left[\frac{1}{1 + a\lambda_1^\Sigma \mathbf{z}_1^\dagger \mathbf{A}_{p_4}^{-1}\mathbf{z}_1 + O(\epsilon)}\right]\right) \tag{3.19}$$

We again use the Taylor series expansion formula, this time with $h(x) = 1/x$, around $x = 1 + a\lambda_1^{\Sigma}\mathbf{z}_1^{\dagger}\mathbf{A}_{p_4}^{-1}\mathbf{z}_1$,

$$h(1 + a\lambda_1^{\Sigma}\mathbf{z}_1^{\dagger}\mathbf{A}_{p_4}^{-1}\mathbf{z}_1 + O(\epsilon)) = \frac{1}{1 + a\lambda_1^{\Sigma}\mathbf{z}_1^{\dagger}\mathbf{A}_{p_4}^{-1}\mathbf{z}_1 + O(\epsilon)} \tag{3.20}$$

$$= \frac{1}{1 + a\lambda_1^{\Sigma}\mathbf{z}_1^{\dagger}\mathbf{A}_{p_4}^{-1}\mathbf{z}_1} - O(\epsilon)\left(\frac{1}{1 + a\lambda_1^{\Sigma}\mathbf{z}_1^{\dagger}\mathbf{A}_{p_4}^{-1}\mathbf{z}_1}\right)^2 + \ldots \tag{3.21}$$

$$= \frac{1}{1 + a\lambda_1^{\Sigma}\mathbf{z}_1^{\dagger}\mathbf{A}_{p_4}^{-1}\mathbf{z}_1} + O(\epsilon) \tag{3.22}$$

Finally, (3.15) becomes

$$E_1(\mathbf{p}_4') = E\left[\frac{\lambda_1^{\Sigma}\mathbf{z}_1^{\dagger}\mathbf{A}_{p_4}^{-1}\mathbf{z}_1}{1 + a\lambda_1^{\Sigma}\mathbf{z}_1^{\dagger}\mathbf{A}_{p_4}^{-1}\mathbf{z}_1}\right] + O(\epsilon) \tag{3.23}$$

$$= E_1(\mathbf{p}_4) + O(\epsilon) \tag{3.24}$$

By using similar arguments, we can conclude that $E_i(\mathbf{p}_4') = E_i(\mathbf{p}_4) + O(\epsilon)$, for $i = 1, 2, 3$. If we insert these into $f_3(\mathbf{p}_4')$, we obtain

$$f_3(\mathbf{p}_4') = \frac{\epsilon(E_3(\mathbf{p}_4) + O(\epsilon))}{(a-\epsilon)(E_1(\mathbf{p}_4) + O(\epsilon)) + b(E_2(\mathbf{p}_4) + O(\epsilon)) + \epsilon(E_3(\mathbf{p}_4) + O(\epsilon))}P \tag{3.25}$$

$$= \frac{\epsilon E_3(\mathbf{p}_4)}{aE_1(\mathbf{p}_4) + bE_2(\mathbf{p}_4) + O(\epsilon)}P + O(\epsilon^2) \tag{3.26}$$

where the last equation follows, because the summation of terms that are in the order of $O(\epsilon)$ and smaller will be in the order of $O(\epsilon)$. Finally, by applying Taylor series expansion one more time with $h(x) = 1/x$, we get

$$f_3(\mathbf{p}_4') = \frac{\epsilon E_3(\mathbf{p}_4)}{aE_1(\mathbf{p}_4) + bE_2(\mathbf{p}_4)} P + O(\epsilon^2) \tag{3.27}$$

We know from (3.14) that $aE_1(\mathbf{p}_4) + bE_2(\mathbf{p}_4) = PE_1(\mathbf{p}_4) = PE_2(\mathbf{p}_4)$. Inserting this into the above equation, we have

$$f_3(\mathbf{p}_4') = \epsilon \frac{E_3(\mathbf{p}_4)}{E_1(\mathbf{p}_4)} + O(\epsilon^2) \tag{3.28}$$

$$> \epsilon \tag{3.29}$$

where the last inequality follows from the fact that $E_3(\mathbf{p}_4) > \mu' = E_1(\mathbf{p}_4)$. This result tells us that starting from ϵ away from an artificial fixed point, the third component of the updated vector, and therefore the updated vector itself moves further away from that artificial fixed point. Finally, by using Lemma 3, we note that the algorithm will move away from the artificial fixed point at each iteration. Therefore, this artificial fixed point is unstable. $\square$

As a result of Lemma 5, the algorithm never converges to an artificial fixed point, if it is not initiated at the boundary of the simplex. Therefore, the point that the algorithm converges to, always satisfies the KKT conditions of the original optimization problem. Since this point is unique, when the algorithm converges, it does so to the unique optimum power allocation policy.

3.3.1 Comparison to Water-filling

In this section, we will compare our results to the classical water-filling solution, when the channel is known perfectly both at the receiver and at the transmitter. We note that in [42], the channel matrix $\mathbf{H}$ is known to both the receiver and the transmitter. The singular value decomposition of $\mathbf{H}$ can be written as $\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^\dagger$, where $n_R \times n_R$ dimensional $\mathbf{U}$, and $n_T \times n_T$ dimensional $\mathbf{V}$ are unitary, and $n_R \times n_T$ dimensional $\mathbf{D}$ is non-negative and diagonal. Let the diagonal elements of $\mathbf{D}$ be denoted by d_i, for $i = 1, \ldots, \min(n_R, n_T)$. The solution of the KKT conditions for this case yields,

$$\lambda_i^Q = \left(\frac{1}{\mu} - \frac{1}{d_i}\right)^+, \quad i = 1, \ldots, \min(n_R, n_T) \tag{3.30}$$

where $(x)^+ = \max\{0, x\}$. Although λ_i^Q is given explicitly, the Lagrange multiplier μ still has to be solved. On the other hand, note that the algorithm proposed in this chapter calculates the eigenvalues directly, without the need for calculating the Lagrange multiplier of the KKT conditions. Considering this fact, we can propose the following new algorithm for the water-filling solution in [42], using the idea in this chapter.

$$\lambda_i^Q(n+1) = \frac{\frac{\lambda_i^Q(n)d_i}{1+\lambda_i^Q(n)d_i}}{\sum_j \frac{\lambda_j^Q(n)d_j}{1+\lambda_j^Q(n)d_j}} P, \quad i = 1, \ldots, \min(n_R, n_T) \tag{3.31}$$

Note that this algorithm has the same properties as (3.13), and finds the optimum eigenvalues without the need for calculating the Lagrange multiplier μ.

3.4 Power Allocation for Multi-User MIMO

The sum-capacity of a MIMO-MAC is given as [6],

$$C_{sum} = \max_{\substack{\mathrm{tr}(\mathbf{Q}_k) \leq P_k \\ k=1,\ldots,K}} E\left[\log\left|\mathbf{I}_{n_R} + \sum_{k=1}^{K} \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^\dagger\right|\right] \tag{3.32}$$

Let $\mathbf{\Sigma}_k = \mathbf{U}_{\Sigma_k} \mathbf{\Lambda}_{\Sigma_k} \mathbf{U}_{\Sigma_k}^\dagger$ be the spectral decomposition of the channel covariance matrix of user k. Then, the optimum transmit covariance matrix $\mathbf{Q}_k$ of user k has the form $\mathbf{Q}_k = \mathbf{U}_{\Sigma_k} \mathbf{\Lambda}_{Q_k} \mathbf{U}_{\Sigma_k}^\dagger$, for all users from Chapter 2. This means that each user transmits along the directions of its own channel covariance matrix. While proving this in Chapter 2, we used the fact that the random matrices $\{\mathbf{Z}_k \mathbf{U}_{\Sigma_k}\}_{k=1}^K$ and $\{\mathbf{Z}_k\}_{k=1}^K$ have the same joint distribution for zero-mean identity-covariance Gaussian $\{\mathbf{Z}_k\}_{k=1}^K$ and unitary $\{\mathbf{U}_{\Sigma_k}\}_{k=1}^K$. Since the structure of the sum-capacity expression is similar to the single-user capacity expression except for the summation inside the determinant, single-user solution easily generalizes to the multi-user case. By inserting this into

(3.32), we get

$$C_{sum} = \max_{\substack{\mathrm{tr}\left(\boldsymbol{\Lambda}_{Q_k}\right) \leq P_k \\ k=1,\ldots,K}} E\left[\log\left|\mathbf{I}_{n_R} + \sum_{k=1}^{K} \mathbf{Z}_k \boldsymbol{\Lambda}_{Q_k} \boldsymbol{\Lambda}_{\Sigma_k} \mathbf{Z}_k^{\dagger}\right|\right] \tag{3.33}$$

$$= \max_{\substack{\sum_{i=1}^{n_T} \lambda_{ki}^{Q} \leq P_k \\ k=1,\ldots,K}} E\left[\log\left|\mathbf{I}_{n_R} + \sum_{k=1}^{K} \sum_{i=1}^{n_T} \lambda_{ki}^{Q} \lambda_{ki}^{\Sigma} \mathbf{z}_{ki} \mathbf{z}_{ki}^{\dagger}\right|\right] \tag{3.34}$$

where $\mathbf{z}_{ki}$ is the i^{th} column of $\mathbf{Z}_k$, i.e., $\{\mathbf{z}_{ki}, k = 1, \ldots, K, i = 1, \ldots, n_T\}$ is a set of $n_R \times 1$ dimensional i.i.d., zero-mean, identity-covariance Gaussian random vectors.

A result of Chapter 2 is that the optimal multi-user transmit direction strategies are decoupled into a set of single-user transmit direction strategies. However, in general, this is not true for the optimal transmit power allocation strategies. The amount of power each user allocates in each direction depends on both the transmit directions and the power allocations of other users. If the eigenvalues of the channel covariance matrices satisfy the conditions given in Chapter 2, then beamforming becomes optimal, and the optimal transmit power allocation strategy for each user reduces to allocating all of its power to its strongest eigen-direction, and this strategy does not require the user to know the channel covariance matrices of the other users. However, if the eigenvalues of the channel covariance matrices do not satisfy these conditions, finding the optimum eigenvalues becomes a harder task. In this section, we will give an iterative algorithm that finds the optimum eigenvalues for all users. We will follow a similar direction as in the single-user case. By writing the Lagrangian for (3.34) and using the identity in (2.25), we obtain the KKT conditions for user k

as

$$E_{ki}(\boldsymbol{\lambda}^Q) \triangleq E\left[\frac{\lambda_{ki}^{\Sigma}\mathbf{z}_{ki}^{\dagger}\mathbf{A}_{ki}^{-1}\mathbf{z}_{ki}}{1+\lambda_{ki}^{Q}\lambda_{ki}^{\Sigma}\mathbf{z}_{ki}^{\dagger}\mathbf{A}_{ki}^{-1}\mathbf{z}_{ki}}\right] \leq \mu_k, \quad i=1,\ldots,n_T \tag{3.35}$$

where $\boldsymbol{\lambda}^Q = [\boldsymbol{\lambda}_1^Q,\ldots,\boldsymbol{\lambda}_K^Q]$, $\boldsymbol{\lambda}_k^Q = [\lambda_{k1}^Q,\ldots,\lambda_{kn_T}^Q]$ is the eigenvalue vector of user k, and μ_k is the Lagrange multiplier corresponding to user k, $\mathbf{A}_{ki} = \mathbf{A} - \lambda_{ki}^Q\lambda_{ki}^{\Sigma}\mathbf{z}_{ki}\mathbf{z}_{ki}^{\dagger}$, and $\mathbf{A} = \mathbf{I}_{n_R} + \sum_{k=1}^{K}\sum_{j=1}^{n_T}\lambda_{kj}^Q\lambda_{kj}^{\Sigma}\mathbf{z}_{kj}\mathbf{z}_{kj}^{\dagger}$. The inequalities in (3.35) are satisfied with equality whenever the optimum λ_{ki}^Q is non-zero, and with strict inequality whenever the optimum λ_{ki}^Q is zero. Similar to the single-user case, λ_{ki}^Q cannot be solved directly from (3.35) because of the expectation operator. Again, we will multiply both sides of (3.35) by λ_{ki}^Q,

$$\lambda_{ki}^Q E_{ki}(\boldsymbol{\lambda}^Q) = \lambda_{ki}^Q\mu_k, \quad i=1,\ldots,n_T \tag{3.36}$$

Note that, similar to the single-user case, (3.36) is satisfied with equality for all λ_{ki}^Q, and we have created some artificial fixed points while obtaining (3.36) from (3.35). For any k, we can find μ_k by summing over all antennas, and by inserting this μ_k into (3.36), we can find the fixed point equations that have to be satisfied by the optimum power values of user k,

$$\lambda_{ki}^Q = \frac{\lambda_{ki}^Q E_{ki}(\boldsymbol{\lambda}^Q)}{\sum_j \lambda_{kj}^Q E_{kj}(\boldsymbol{\lambda}^Q)} P_k \triangleq g_{ki}(\boldsymbol{\lambda}^Q), \quad i=1,\ldots,n_T \tag{3.37}$$

where we defined the right hand side of (3.37) which depends on all of the eigenvalues as $g_{ki}(\boldsymbol{\lambda}^Q)$.

We propose the following algorithm, that enforces (3.37),

$$\boldsymbol{\lambda}_k^Q(n+1) = \mathbf{g}_k\left(\boldsymbol{\lambda}_1^Q, \ldots, \boldsymbol{\lambda}_{k-1}^Q, \boldsymbol{\lambda}_k^Q(n), \boldsymbol{\lambda}_{k+1}^Q, \ldots, \boldsymbol{\lambda}_K^Q\right) \tag{3.38}$$

where $\mathbf{g}_k = [g_{k1}, \ldots, g_{kn_T}]$ is the vector valued update function of user k. This algorithm finds the optimum eigenvalues of a given user by assuming that the eigenvalues of the rest of the users are fixed. The algorithm moves to another user, after (3.38) converges. A complete update corresponding to user k only, i.e., running the algorithm in (3.38) for user k until it converges while the eigenvalues of the other users are fixed, is equivalent to the single-user algorithm proposed in (3.13). Therefore, we know from the previous section that the algorithm in (3.38) converges to the unique optimum point, when the eigenvalues of the rest of the users are fixed. The optimization problem that is solved by (3.38) is,

$$C_k = \max_{\sum_{i=1}^{n_T} \lambda_{ki}^Q \leq P_k} E\left[\log\left|\mathbf{B}_k + \sum_{i=1}^{n_T} \lambda_{ki}^Q \lambda_{ki}^\Sigma \mathbf{z}_{ki}\mathbf{z}_{ki}^\dagger\right|\right] \tag{3.39}$$

where $\mathbf{B}_k = \mathbf{I}_{n_R} + \sum_{l\neq k}^K \sum_{i=1}^{n_T} \lambda_{li}^Q \lambda_{li}^\Sigma \mathbf{z}_{li}\mathbf{z}_{li}^\dagger$ depends on the fixed eigenvalues of all other users. Such an algorithm is guaranteed to converge to the global optimum [2, page 219], since C_{sum} is a concave function of λ_{ki} for all k and i, C_k is a strictly concave function of λ_{ki} for all i, and the constraint set is convex and has a Cartesian product structure among the users. Note that in [49], this kind of an algorithm is used in order to find the iterative water-filling solution. However, in that setting, where both the receiver and the transmitters know the perfect CSI, an iteration corresponding to user k does not include another algorithm, but it is just a single-user water-filling solution.

In order to improve the convergence rate, we also propose the following multi-user algorithm,

$$\boldsymbol{\lambda}_{k'}^{Q}(n+1) = \mathbf{g}_{k'}\left(\boldsymbol{\lambda}_{1}^{Q}(n+1), \ldots, \boldsymbol{\lambda}_{k'-1}^{Q}(n+1), \boldsymbol{\lambda}_{k'}^{Q}(n), \boldsymbol{\lambda}_{k'+1}^{Q}(n), \ldots, \boldsymbol{\lambda}_{K}^{Q}(n)\right) \quad (3.40)$$

where $k' = (n+1) \mod K$. At a given time $n+1$, this algorithm updates the eigenvalues of user k'. In the next iteration, it moves to another user. Since at a given iteration corresponding to user k, this algorithm does not solve (3.39) completely, we cannot conclude its convergence using [2, page 219]. However, we have observed the convergence of this algorithm experimentally through many simulations. One potential method to prove the convergence of this algorithm could be through proving that each iteration of the single-user algorithm in (3.13) increases the objective function of the optimization problem, i.e., the rate. Even though we proved that each iteration of this algorithm either monotonically increases or monotonically decreases each eigenvalue, and therefore, monotonically decreases the distance between the iterated eigenvalue vector and the optimum eigenvalue vector, we have not been able to prove mathematically that each iteration monotonically increases the objective function. Yet, we have observed this monotonicity through extensive simulations. Given that the objective function is a strictly concave function of the eigenvalue vector, we conjecture that the algorithm in (3.13) increases the objective function monotonically.

On the other hand, we have observed experimentally that the algorithm in (3.40) converges much faster than the algorithm in (3.38). This could be due to the fact that, while the algorithm in (3.38) runs many iterations of the same user before it moves to another user, the algorithm in (3.40) runs only one iteration for each user before it moves to the next user.

3.5 Numerical Results

In this section, we will provide numerical examples for the performances of the proposed algorithms. In Figure 3.1 and Figure 3.2, we plot the trajectories of the iterations of the proposed single-user algorithm for a MIMO system with $n_R = n_T = P = 3$. We run the algorithm three times for each figure with different initial points, which are ϵ away from the three corner points of the 3-dimensional simplex. In Figure 3.1, all of the optimum eigenvalues are non-zero, and in Figure 3.2, one of the optimum eigenvalues is zero. We observe, from the two figures, that the algorithm converges to the unique optimum point.

In Figure 3.3 and Figure 3.4, we plot the eigenvalues as a function of the iteration index. We observe that the eigenvalues converge to the same unique convergence point starting from various initial points. In addition to the points that are ϵ away from the corner points, the other initial points are: the all-one vector, and the point corresponding to the channel covariance matrix eigenvalues, which is normalized to satisfy the power constraint. In Figure 3.3, all of the optimum eigenvalues are non-

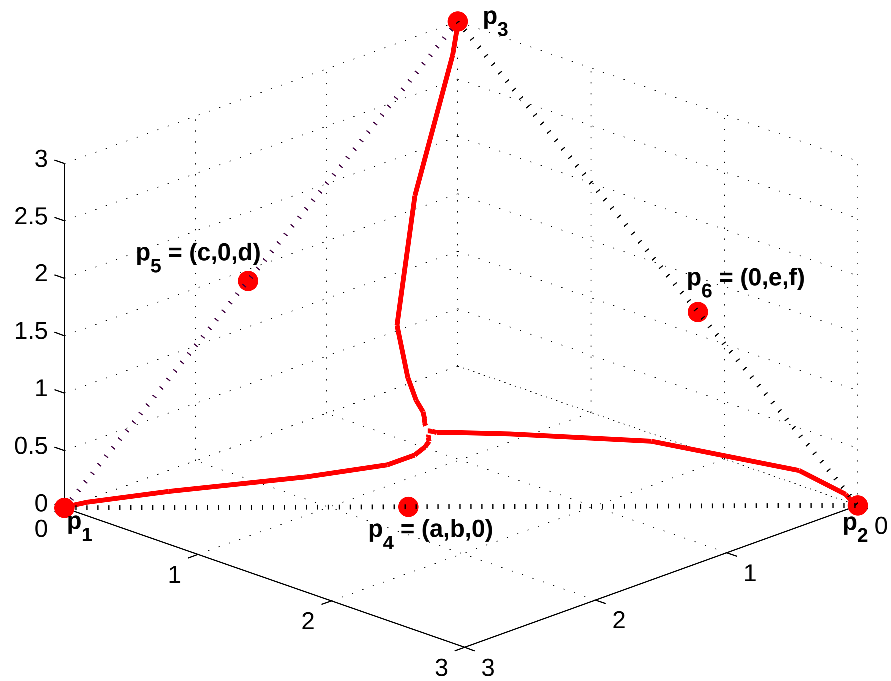


Figure 3.1: The trajectories of the single-user algorithm when it is started from the corner points of the simplex for the case where the optimal eigenvalues are all non-zero.

zero, and in Figure 3.4, one of the optimum eigenvalues is zero. As we see from Figure 3.3, the algorithm needs much less time to converge to the optimum point when it is started from the normalized channel covariance eigenvalue point compared to the cases when it is started from any other points on the simplex. This is true mainly because of an argument similar to the water-filling argument, where we allocate more power to the strongest channel. As a result, the unique optimum transmit covariance eigenvalue vector is located close to the normalized channel covariance eigenvalue vector. Since they are located close by, it takes less time for the algorithm to converge to the optimum. Therefore, we may prefer to start the algorithm from the normalized channel covariance matrix eigenvalues, in order to improve the convergence rate of

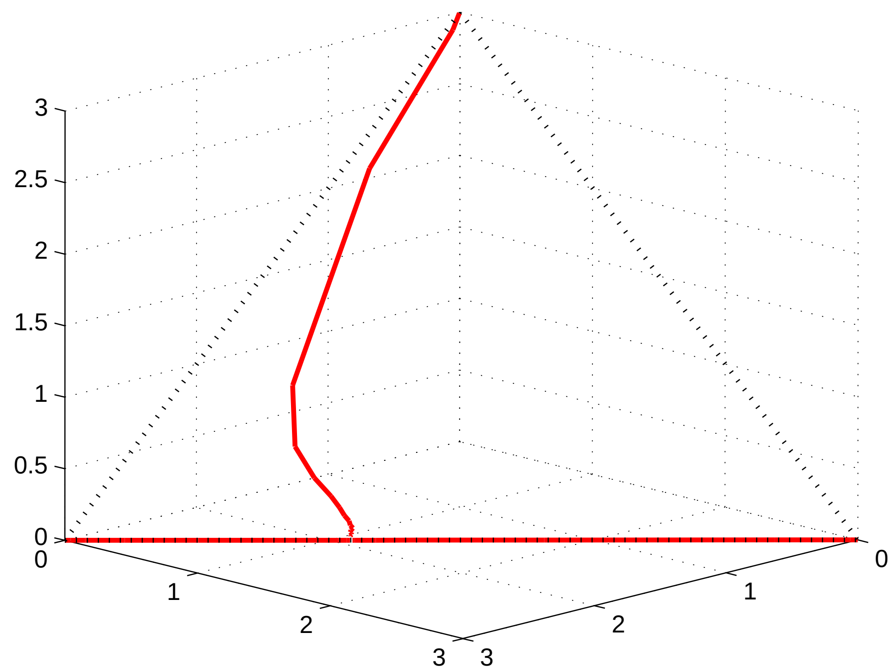


Figure 3.2: The trajectories of the single-user algorithm when it is started from the corner points of the simplex for the case where one of the optimal eigenvalues is zero.

the algorithm. We note however that the algorithm converges to the optimum point from any arbitrary initial point.

We also note that, even when we start the algorithm from the normalized channel covariance matrix eigenvalues, we observe from Figure 3.4 that it may still take some time for the algorithm to converge. In this case, this occurs mainly because one of the optimum eigenvalues is equal to zero. In order to improve the convergence rate, we can check if any one of the optimum eigenvalues will be zero, before we start the algorithm. We can use the beamforming optimality conditions from [14], and from the Chapter 2 in order to check if the second component of the eigenvalue vector is zero. For the rest of the components, similar conditions can easily be derived by

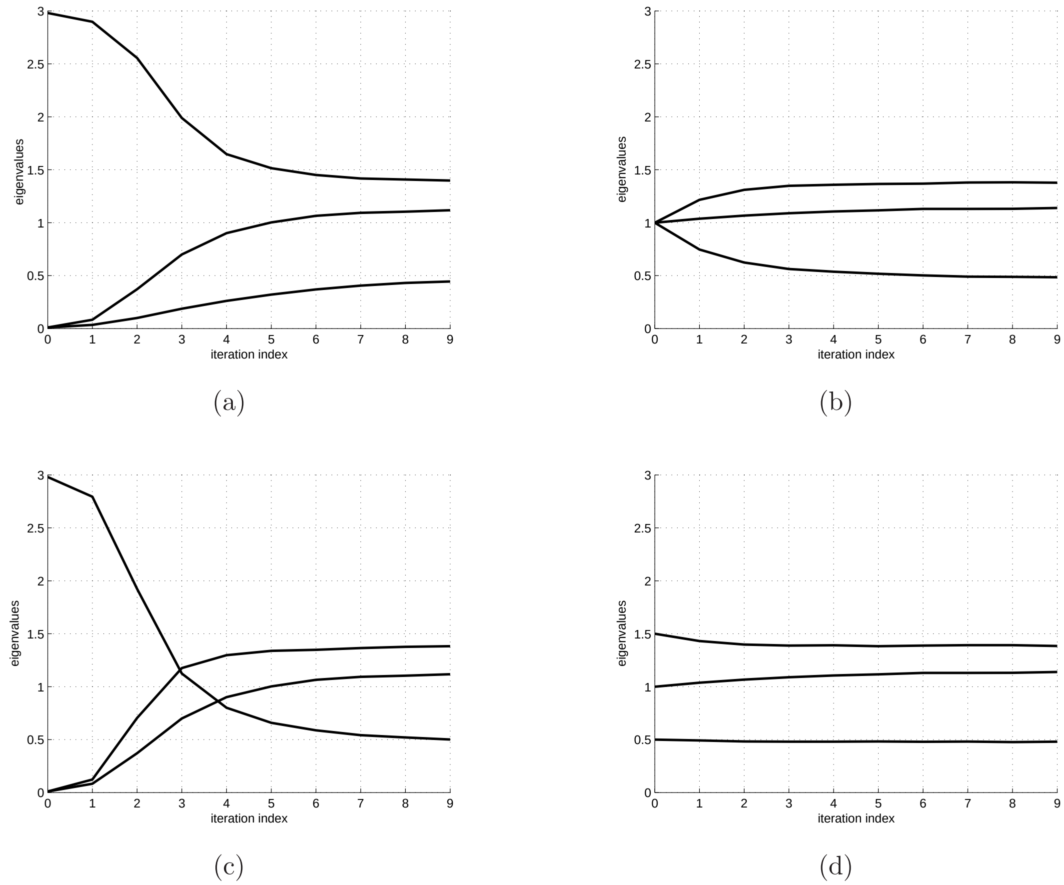


Figure 3.3: The convergence of the single-user algorithm starting from various points, when all of the optimal eigenvalues are non-zero: (a) convergence of all three eigenvalues from $(P - 2\epsilon, \epsilon, \epsilon)$; (b) convergence of all three eigenvalues from $(\frac{P}{n_T}, \frac{P}{n_T}, \frac{P}{n_T})$; (c) convergence of all three eigenvalues from $(\epsilon, \epsilon, P - 2\epsilon)$; (d) convergence of all three eigenvalues from the normalized channel eigenvalue vector.

using the ideas in [14], and Chapter 2. If there are any eigenvalues that will be zero at the optimum, we can drop them from the optimization problem, and solve a reduced problem with fewer dimensions. In Figure 3.5, we have selected the eigenvalues of the channel covariance matrix so that the third eigenvalue of the optimum transmit covariance matrix happens to be zero. We considered two different initial points: the normalized channel covariance eigenvalue vector, and a vector obtained by setting the third component of the channel covariance eigenvalue vector to zero, before the

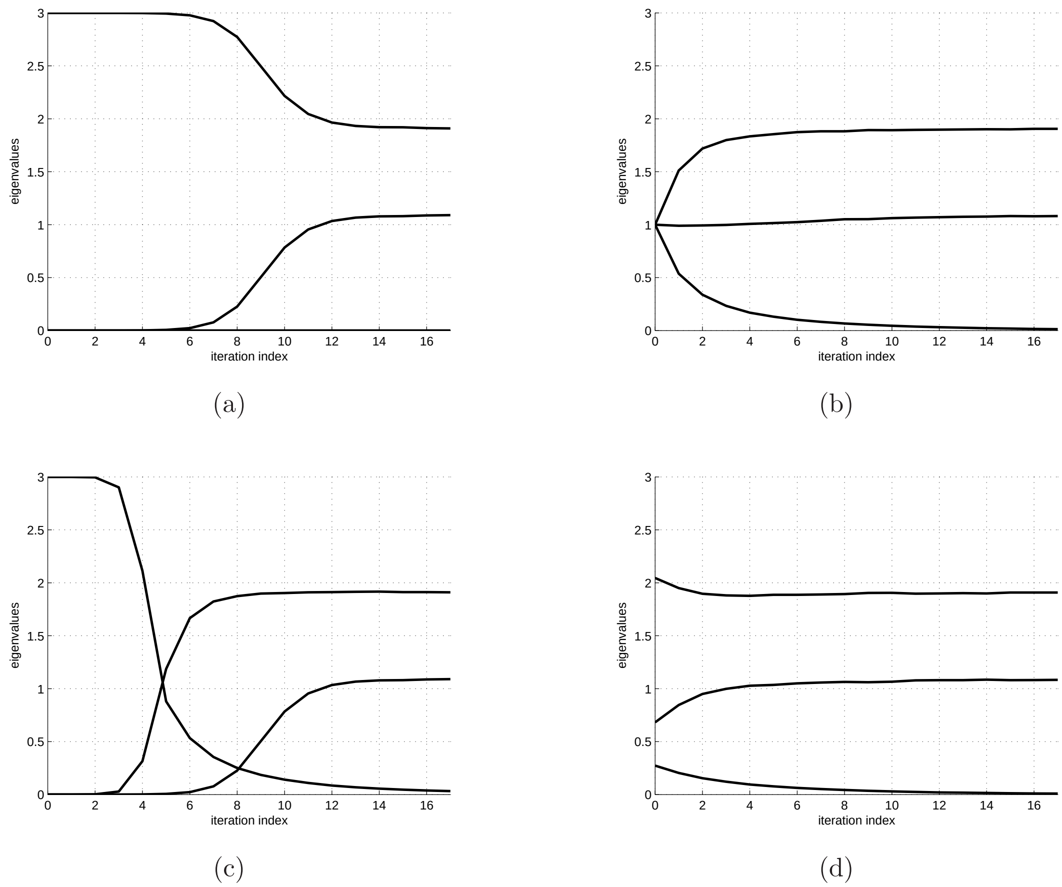


Figure 3.4: The convergence of the single-user algorithm starting from various points, when one of the optimal eigenvalues is zero: (a) convergence of all three eigenvalues from $(P-2\epsilon, \epsilon, \epsilon)$; (b) convergence of all three eigenvalues from $(\frac{P}{n_T}, \frac{P}{n_T}, \frac{P}{n_T})$; (c) convergence of all three eigenvalues from $(\epsilon, \epsilon, P-2\epsilon)$; (d) convergence of all three eigenvalues from the normalized channel eigenvalue vector.

normalization. We observe that the algorithm converges much faster if we identify the components that will be zero at the convergence point and remove them from the iterations.

Finally, we consider a multi-user MIMO-MAC scenario. Note that, for a given user, the multi-user algorithm given in (3.38) demonstrates the same convergence behavior as in Figure 3.3 and Figure 3.4, when the eigenvalues of the other users are kept constant. Therefore, we plot Figure 3.6 by running the multi-user algorithm

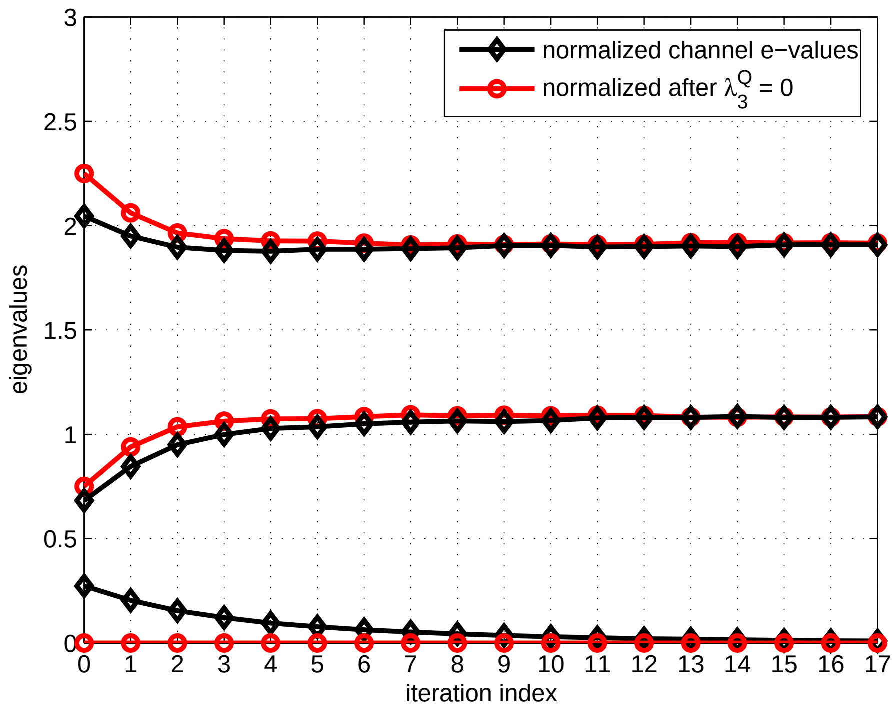


Figure 3.5: The convergence of the single-user algorithm when one of the optimum eigenvalues is zero.

proposed in (3.40). In this figure, we consider 3 users with different channel covariance matrices. The algorithm is started at the normalized channel covariance eigenvalue vectors of the users. Each iteration in the figure corresponds to an update of the eigenvalues of the transmit covariance matrices of all users. At the end of the first iteration, all users have run the algorithm in (3.40) once. We can see in Figure 3.6 that the multi-user algorithm converges quite quickly, and at the end of the fourth iteration, all users are almost at their optimum eigenvalue points.

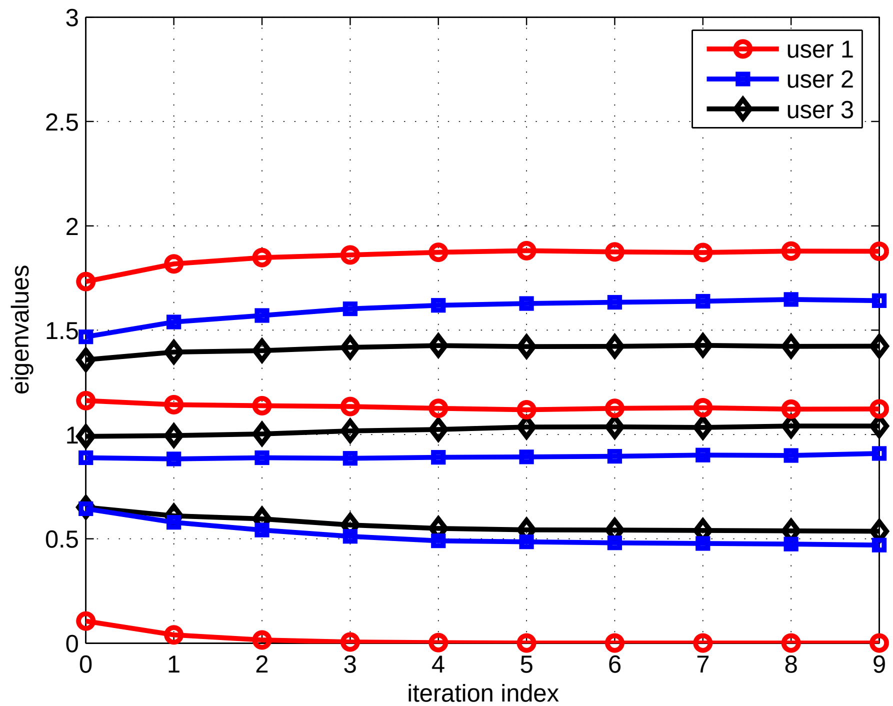


Figure 3.6: The convergence of the multi-user algorithm where each iteration corresponds to a single update of all users.

3.6 Discussions

Due to the nature of our optimization problem, our algorithms include calculation of some expectations at each iteration. Direct calculation of these expectations is sometimes difficult. However, by exploiting the ergodicity of the system and using sample averages, we can get very fast results. Although the number of expectations that has to be calculated increases as the number of users increases, fortunately, we can eliminate most of the components inside the expectations using the results of Chapter 2, which state that beamforming becomes optimal as the number of users in the system increases. As it can be seen in Chapter 2, even for a fairly low number of users, beamforming is almost optimal. Therefore, by combining beamforming optimality conditions with the proposed algorithms, we can find the optimum power allocations of the users much faster. Figure 3.5 shows that the number of iterations is significantly less when beamforming optimality conditions are utilized. Although it cannot be seen in the figure, each iteration takes less time as well, i.e., the expectations are computed faster, since there is less randomness in the system as a result of setting some eigenvalues to zero.

Another issue that we want to discuss here is the possibility of having a channel with double-sided correlation. In our model, as a result of the assumption that the receiver (e.g., a base station) is not physically limited and one can place the antenna elements sufficiently away from each other, the receiver side correlation matrix becomes the identity matrix. In a different model with receiver side correlation present in the system, similar results can be found. For the single-user scenario, it is already known that the transmit directions are still the eigenvectors of the transmitter side channel correlation matrix, even when there is receiver side channel correlation in the system [18]. Beamforming optimality condition for this case is also found previously [18]. For the power allocation problem, an approach similar to the one in this chapter can be applied and a similar but more cumbersome algorithm can be found. This algorithm includes extra terms that are similar to the terms in beamforming optimality conditions that are given in [18]. For the multi-user scenario, our approach generalizes to the case where there is receiver side channel correlation in the system, when the receiver side channel correlation matrices of all users are the same. This might be motivated by assuming that the receiver side channel correlation is only a result of the physical structure of the receiver and the environment around the receiver, therefore it is the same for all users. In this case, it is possible to find similar but again more cumbersome algorithms in order to solve the optimum power allocation policies of all users.

3.7 Conclusions

We proposed globally convergent algorithms for finding the optimum power allocation policies for both single-user MIMO and MIMO-MAC systems. Combining this with our previous results in Chapter 2 on the optimum transmit directions and the asymptotic behavior of MIMO-MAC systems, the sum capacity maximization problem is completely solved for a finite or infinite sized MIMO-MAC with the full CSI at the receiver and partial CSI at the transmitters in the form of channel covariance information. In this chapter, for a single-user case, we proved the convergence and the uniqueness of the convergence point of a pre-existing algorithm. This proof handles the complications arising from the existence of the artificial fixed points, and it gives some insights to the classical water-filling solution. For the multi-user case, we derived and proved the convergence of a multi-user algorithm, which finds the optimum power allocations of all users.

The results in this chapter are published in [37], [38].

3.8 Appendix

3.8.1 Proof of Lemma 3

Without loss of generality, let us take $j = 1$. We will show that $\frac{\partial\left(\sum_{k=1}^{n_T} \lambda_k^Q E_k(\boldsymbol{\lambda}^Q)\right)}{\partial \lambda_1^Q} > 0$.

$$\frac{\partial \lambda_1^Q E_1(\boldsymbol{\lambda}^Q)}{\partial \lambda_1^Q} = E\left[\frac{\lambda_1^\Sigma \mathbf{z}_1^\dagger \mathbf{A}_1^{-1}\mathbf{z}_1}{\left(1 + \lambda_1^Q \lambda_1^\Sigma \mathbf{z}_1^\dagger \mathbf{A}_1^{-1}\mathbf{z}_1\right)^2}\right] \tag{3.41}$$

Now, for $k = 2, \ldots, n_T$, let us consider $\lambda_k^Q E_k(\boldsymbol{\lambda}^Q) = \lambda_k^Q \lambda_k^\Sigma E[\mathbf{z}_k \mathbf{A}^{-1}\mathbf{z}_k]$. Applying the matrix inversion lemma [12, page 19] to $\mathbf{A} = \mathbf{A}_1 + \lambda_1^Q \lambda_1^\Sigma \mathbf{z}_1 \mathbf{z}_1^\dagger$, we get

$$\lambda_k^Q E_k(\boldsymbol{\lambda}^Q) = \lambda_k^Q \lambda_k^\Sigma \left(E[\mathbf{z}_k \mathbf{A}_1^{-1}\mathbf{z}_k] - E\left[\frac{\lambda_1^Q \lambda_1^\Sigma (\mathbf{z}_k \mathbf{A}_1^{-1}\mathbf{z}_1)^2}{1 + \lambda_1^Q \lambda_1^\Sigma \mathbf{z}_1^\dagger \mathbf{A}_1^{-1}\mathbf{z}_1}\right]\right) \tag{3.42}$$

By taking the derivative of (3.42) with respect to λ_1^Q, we get

$$\frac{\partial \lambda_k^Q E_k(\boldsymbol{\lambda}^Q)}{\partial \lambda_1^Q} = -\lambda_k^Q \lambda_k^\Sigma E\left[\frac{\lambda_1^\Sigma (\mathbf{z}_k^\dagger \mathbf{A}_1^{-1}\mathbf{z}_1)^2}{(1 + \lambda_1^Q \lambda_1^\Sigma \mathbf{z}_1^\dagger \mathbf{A}_1^{-1}\mathbf{z}_1)^2}\right] \tag{3.43}$$

Combining (3.41) and (3.43) with $\mathbf{s}_k = (\lambda_k^\Sigma \lambda_k^Q)^{1/2}\mathbf{z}_k$, we have

$$\frac{\partial\left(\sum_{k=1}^{n_T} \lambda_k^Q E_k(\boldsymbol{\lambda}^Q)\right)}{\partial \lambda_1^Q} = \frac{1}{\lambda_1^Q} E\left[\frac{\mathbf{s}_1^\dagger \mathbf{A}_1^{-1}\mathbf{s}_1 - \sum_{k=2}^{n_T}(\mathbf{s}_k^\dagger \mathbf{A}_1^{-1}\mathbf{s}_1)^2}{(1 + \mathbf{s}_1^\dagger \mathbf{A}_1^{-1}\mathbf{s}_1)^2}\right] \tag{3.44}$$

We note that $\mathbf{A}_1 = \mathbf{I} + \mathbf{S}_1\mathbf{S}_1^\dagger$, where $\mathbf{S}_1 = [\mathbf{s}_2, \ldots, \mathbf{s}_{n_T}]$. Then, by using the matrix inversion lemma, we have $\mathbf{A}_1^{-1} = \mathbf{I} - \mathbf{S}_1(\mathbf{I} + \mathbf{S}_1^\dagger\mathbf{S}_1)^{-1}\mathbf{S}_1^\dagger$. Finally, note that $\sum_{k=2}^{n_T}(\mathbf{s}_k^\dagger\mathbf{A}_1^{-1}\mathbf{s}_1)^2 = \mathbf{s}_1^\dagger\mathbf{A}_1^{-1}\mathbf{S}_1\mathbf{S}_1^\dagger\mathbf{A}_1^{-1}\mathbf{s}_1$. Now, we will find equivalent expressions for the numerator of (3.44). Let us first look at $\mathbf{s}_1^\dagger\mathbf{A}_1^{-1}\mathbf{s}_1$,

$$\mathbf{s}_1^\dagger\mathbf{A}_1^{-1}\mathbf{s}_1 = \mathbf{s}_1^\dagger\mathbf{s}_1 - \mathbf{s}_1^\dagger\mathbf{S}_1(\mathbf{I} + \mathbf{S}_1^\dagger\mathbf{S}_1)^{-1}\mathbf{S}_1^\dagger\mathbf{s}_1 \tag{3.45}$$

Now, let us look at $\mathbf{s}_1^\dagger\mathbf{A}_1^{-1}\mathbf{S}_1$,

$$\mathbf{s}_1^\dagger\mathbf{A}_1^{-1}\mathbf{S}_1 = \mathbf{s}_1^\dagger\mathbf{S}_1 - \mathbf{s}_1^\dagger\mathbf{S}_1(\mathbf{I} + \mathbf{S}_1^\dagger\mathbf{S}_1)^{-1}\mathbf{S}_1^\dagger\mathbf{S}_1 \tag{3.46}$$

$$= \mathbf{s}_1^\dagger\mathbf{S}_1(\mathbf{I} + \mathbf{S}_1^\dagger\mathbf{S}_1)^{-1}(\mathbf{I} + \mathbf{S}_1^\dagger\mathbf{S}_1) - \mathbf{s}_1^\dagger\mathbf{S}_1(\mathbf{I} + \mathbf{S}_1^\dagger\mathbf{S}_1)^{-1}\mathbf{S}_1^\dagger\mathbf{S}_1 \tag{3.47}$$

$$= \mathbf{s}_1^\dagger\mathbf{S}_1(\mathbf{I} + \mathbf{S}_1^\dagger\mathbf{S}_1)^{-1} \tag{3.48}$$

Inserting (3.45) and (3.48) into (3.44), it is sufficient to show

$$\mathbf{s}_1^\dagger\mathbf{s}_1 - \mathbf{s}_1^\dagger\mathbf{S}_1(\mathbf{I} + \mathbf{S}_1^\dagger\mathbf{S}_1)^{-1}\mathbf{S}_1^\dagger\mathbf{s}_1 - \mathbf{s}_1^\dagger\mathbf{S}_1(\mathbf{I} + \mathbf{S}_1^\dagger\mathbf{S}_1)^{-2}\mathbf{S}_1^\dagger\mathbf{s}_1 \geq 0 \tag{3.49}$$

In order to proceed, we note that $\mathbf{s}_1^\dagger\mathbf{s}_1 \geq \mathbf{s}_1^\dagger\mathbf{S}_1(\mathbf{S}_1^\dagger\mathbf{S}_1)^{-1}\mathbf{S}_1^\dagger\mathbf{s}_1$ holds. This can be seen by noting that the matrix $\mathbf{S}_1(\mathbf{S}_1^\dagger\mathbf{S}_1)^{-1}\mathbf{S}_1^\dagger$ is idempotent, and therefore its eigenvalues are either zero or one. Hence, $\mathbf{I} - \mathbf{S}_1(\mathbf{S}_1^\dagger\mathbf{S}_1)^{-1}\mathbf{S}_1^\dagger$ is positive definite. Using this inequality, the condition becomes,

$$\mathbf{s}_1^\dagger\mathbf{S}_1\left[(\mathbf{S}_1^\dagger\mathbf{S}_1)^{-1} - (\mathbf{I} + \mathbf{S}_1^\dagger\mathbf{S}_1)^{-1} - (\mathbf{I} + \mathbf{S}_1^\dagger\mathbf{S}_1)^{-2}\right]\mathbf{S}_1^\dagger\mathbf{s}_1 \geq 0 \tag{3.50}$$

Now, let us look at the term between the square brackets,

$$(\mathbf{S}_1^\dagger \mathbf{S}_1)^{-1} - (\mathbf{I} + \mathbf{S}_1^\dagger \mathbf{S}_1)^{-1} - (\mathbf{I} + \mathbf{S}_1^\dagger \mathbf{S}_1)^{-2} = \tag{3.51}$$

$$= (\mathbf{S}_1^\dagger \mathbf{S}_1)^{-1}(\mathbf{I} + \mathbf{S}_1^\dagger \mathbf{S}_1)(\mathbf{I} + \mathbf{S}_1^\dagger \mathbf{S}_1)^{-1} - (\mathbf{I} + \mathbf{S}_1^\dagger \mathbf{S}_1)^{-1} - (\mathbf{I} + \mathbf{S}_1^\dagger \mathbf{S}_1)^{-2} \tag{3.52}$$

$$= (\mathbf{S}_1^\dagger \mathbf{S}_1)^{-1}(\mathbf{I} + \mathbf{S}_1^\dagger \mathbf{S}_1)^{-1} - (\mathbf{I} + \mathbf{S}_1^\dagger \mathbf{S}_1)^{-2} \tag{3.53}$$

$$= \left((\mathbf{S}_1^\dagger \mathbf{S}_1)^{-1} - (\mathbf{I} + \mathbf{S}_1^\dagger \mathbf{S}_1)^{-1}\right)(\mathbf{I} + \mathbf{S}_1^\dagger \mathbf{S}_1)^{-1} \tag{3.54}$$

$$= (\mathbf{S}_1^\dagger \mathbf{S}_1)^{-1}(\mathbf{I} + \mathbf{S}_1^\dagger \mathbf{S}_1)^{-1}(\mathbf{I} + \mathbf{S}_1^\dagger \mathbf{S}_1)^{-1} \tag{3.55}$$

$$= (\mathbf{I} + \mathbf{S}_1^\dagger \mathbf{S}_1)^{-1}(\mathbf{S}_1^\dagger \mathbf{S}_1)^{-1}(\mathbf{I} + \mathbf{S}_1^\dagger \mathbf{S}_1)^{-1} \tag{3.56}$$

Now, let the singular value decomposition of $\mathbf{S}_1$ be $\mathbf{S}_1 = \mathbf{U}\mathbf{D}\mathbf{V}^\dagger$, then $\mathbf{S}_1^\dagger \mathbf{S}_1 = \mathbf{V}\mathbf{D}^2\mathbf{V}^\dagger$. Inserting this into (3.56), and (3.56) into (3.50), we get

$$\mathbf{s}_1^\dagger \mathbf{U}\mathbf{D}\mathbf{V}^\dagger \mathbf{V}(\mathbf{I} + \mathbf{D}^2)^{-1}\mathbf{V}^\dagger \mathbf{V}\mathbf{D}^{-2}\mathbf{V}^\dagger \mathbf{V}(\mathbf{I} + \mathbf{D}^2)^{-1}\mathbf{V}^\dagger \mathbf{V}\mathbf{D}\mathbf{U}^\dagger \mathbf{s}_1 \geq 0 \tag{3.57}$$

$$\mathbf{s}_1^\dagger \mathbf{U}\mathbf{D}(\mathbf{I} + \mathbf{D}^2)^{-1}\mathbf{D}^{-2}(\mathbf{I} + \mathbf{D}^2)^{-1}\mathbf{D}\mathbf{U}^\dagger \mathbf{s}_1 \geq 0 \tag{3.58}$$

$$\mathbf{s}_1^\dagger \mathbf{U}(\mathbf{I} + \mathbf{D}^2)^{-1}(\mathbf{I} + \mathbf{D}^2)^{-1}\mathbf{U}^\dagger \mathbf{s}_1 \geq 0 \tag{3.59}$$

Finally, since $(\mathbf{I} + \mathbf{D}^2)^{-2}$ is positive definite, (3.59) holds and (3.44) is greater than zero.

3.8.2 Proof of Lemma 5

For arbitrary number of antennas, we will assume that we are at some artificial fixed point, which is not a solution of the original optimization problem, with possibly more than one zero components. Let this artificial fixed point be $\mathbf{p} = (a_1, a_2, \ldots, a_{n_T})$, and let S be the index set of the zero components so that $a_j = 0$ for all $j \in S$. Since $\mathbf{p}$ is a fixed point, the following equalities hold for $i \notin S$

$$a_i = \frac{a_i E_i(\mathbf{p})}{\sum_{j \notin S} a_j E_j(\mathbf{p})} P \tag{3.60}$$

From above, we find that $\sum_{j \notin S} a_j E_j(\mathbf{p}) = P E_i(\mathbf{p}) = P\mu'$, for all $i \notin S$. This is equivalent to saying that the KKT conditions of the reduced optimization problem corresponding to components, $i \notin S$, are satisfied with equality, where μ' is possibly different than μ. We will show that some conditions on $E_j(\mathbf{p})$, $j \in S$ cannot hold. The case where $E_j(\mathbf{p}) = \mu'$ for all $j \in S$ cannot hold, because this would mean that the KKT conditions of the original optimization problem are all satisfied with equality with $\mu' = \mu$, and this can only happen when optimal λ_i^Q's for all i are non-zero. Now, let k be the smallest index in S, then because of the ordering of the eigenvalues of the channel covariance matrix, $E_k(\mathbf{p})$ is greater than all $E_j(\mathbf{p})$, for all $j \neq k$, and $j \in S$. The case where $E_k(\mathbf{p}) = \mu'$ and $E_j(\mathbf{p}) \leq \mu'$, for all $j \neq k$, and $j \in S$ cannot hold, because that would mean that the KKT conditions of the reduced optimization problem is violated. The case where $E_k(\mathbf{p}) < \mu'$ and $E_j(\mathbf{p}) < E_k(\mathbf{p}) < \mu'$, for all $j \neq k$, and $j \in S$ cannot hold, because that would mean that we satisfy all KKT conditions of the original optimization problem with $\mu' = \mu$. This contradicts with our assumption that we are at an artificial fixed point that is not the solution of the original optimization problem. Therefore, in all other possibilities, we have at least $E_k(\mathbf{p}) > \mu'$, where k is the smallest index in S. Now, we will show that by perturbing the artificial fixed point by an ϵ amount, we move further away from that artificial fixed point. We run the algorithm for $\mathbf{p}'$ which is different from $\mathbf{p}$ in two components: the k^{th} component is ϵ, and any i^{th} component, for $i \notin S$, is $a_i - \epsilon$. By using the same Taylor series arguments, we can say that $E_i(\mathbf{p}') = E_i(\mathbf{p}) + O(\epsilon)$, for $i = 1, \ldots, n_T$. If we insert these into $f_k(\mathbf{p}')$, we have

$$f_k(\mathbf{p}') = \frac{\epsilon E_k(\mathbf{p})}{\sum_{i \notin S} a_i E_i(\mathbf{p})} P + O(\epsilon^2) \tag{3.61}$$

We know from (3.60) that $\sum_{i \notin S} a_i E_i(\mathbf{p}) = P\mu'$. Inserting this into the above equation, we have

$$f_k(\mathbf{p}') = \epsilon \frac{E_k(\mathbf{p})}{\mu'} + O(\epsilon^2) \tag{3.62}$$

$$> \epsilon \tag{3.63}$$

where the last inequality follows from the fact that $E_k(\mathbf{p}) > \mu'$. This result tells us that starting from ϵ away from an artificial fixed point, the k^{th} component of the updated vector, and therefore the updated vector itself moves further away from the artificial fixed point. By using Lemma 3, the algorithm will move away from the artificial fixed point at each iteration. Therefore, this artificial fixed point is unstable.

Chapter 4

Channel Estimation and Noisy CSI at the Receiver

In wireless communication scenarios, the achievable rate of a system depends crucially on the amount of 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. If the transmitter adapts its transmission scheme to the received CSI estimate, it is possible to obtain higher rates, especially in MIMO links. In practice, the channel estimation is always noisy, and the amount of feedback to the transmitter is limited.

Measuring the CSI and feeding it back to the transmitter uses communication resources, which could otherwise be used for useful information transmission. One way of measuring the CSI is that the transmitter sends a known training sequence, from which the receiver measures the channel. This estimated CSI is used by the receiver in decoding the messages, however the estimation process uses up time and power.

Optimizing the achievable rate in a fading channel has been widely studied under various assumptions on the channel estimation process and the CSI available at the transmitter side. With perfect CSI at the receiver and the instantaneous knowledge of perfect CSI at the transmitter, the optimum adaptation scheme becomes water-filling [7, 42, 49]. In some cases, especially in 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, 14, 46].

Another line of research considers the actual estimation of the channel at the receiver, which is noisy. When the CSI available at the receiver is not perfect, most of the research focuses on single-user systems. The capacity and the corresponding optimum signaling scheme for this case are not known. However, lower and upper bounds for the capacity can be obtained. A common approach in finding an achievable rate for such situations involves assuming Gaussian signaling. Reference [28] finds bounds for the achievable rate of a single-user system without CSI feedback, under the assumption that there exists a separate channel, that does not consume communication resources, for the estimation process. This work has been extended to the case where there is error free feedback in the system [20], where it was shown that the optimum power allocation that achieves the lower bound is a form of water-filling. Reference [47] extends [20, 28] to a MIMO system, where the power allocation is done in two steps: first, the sum power values for all channel realizations are found, and then the sum power is spatially water-filled over the antennas at each channel state.

It is important to note that [20, 28, 47] assume the existence of a separate channel that does not consume system resources for channel estimation. Consequently, [20, 28, 47] do not consider optimizing the channel estimation process. For a single-user multiple-antenna system with no CSI available at the transmitter, [9] considers optimizing the achievable rate as a function of both the training and the data transmission phases. Since there is no CSI feedback, the transmitter power allocation is constant over the channel states and the antennas. In this case, optimizing the achievable rate involves finding the optimal power allocation between the training and data transmission phases, determining the optimal training sequence length, and the optimal training symbols. Reference [9] shows that using more training symbols than the number of transmit antennas is sub-optimal, and that orthonormal training symbols are optimal.

In the first part of this chapter, we consider a single-user, block-fading, correlated MIMO channel with noisy channel estimation at the receiver, and partial CSI available at the transmitter. The CSI feedback that we consider lies somewhere between perfect CSI [47] and no CSI [9], and it is similar to [3, 14, 46], and is the same as in the previous chapters in this thesis, i.e., covariance feedback. 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 [9] 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.

The training phase is characterized by three parameters, namely, the training signal, the training sequence length and the training sequence power. Similarly, the data transmission phase is characterized by the data carrying input signal, data transmission length, and the data transmission power. Assuming that the receiver uses linear 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 [9]. However, unlike [9], 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 the 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.

In a multi-user setting, the amount of resources required to measure the channel and to feed the estimated channel back to the transmitter increases substantially. When perfect channel information is assumed to be available at the receiver and the transmitters at no cost, [49] finds the optimum transmission strategy, which is a multi-user water-filling scheme. Under a more practical assumption, when there is perfect CSI at the receiver but only partial CSI available at the transmitters, we found the optimum transmit strategies for all users in Chapters 2 and 3.

When the perfect CSI assumption at the receiver is relaxed, i.e., when the channel estimation at the receiver is noisy, most of the research focuses on single-user systems [9], [28], [47]. In the second half of this chapter, we extend the single-user results to multiple access channels. In a multi-user setting, we first consider the channel estimation process and find the optimum training signals for all users. Although all of the users are allowed to use the available training duration simultaneously, we find that the training signals of the users should be non-overlapping in time. Since the total block length, and therefore the total training duration is limited, each user can only train a fraction of its available channel dimensions, which might result in shorter individual training signal durations compared to the single-user case. However, as a result of having shorter individual training signal duration and the conservation of energy, the training signal power that is used by a particular user in a multi-user case could be larger than the training signal power that the same user would use in a single-user case. Therefore, although fewer dimensions of the channel are estimated, the channel estimation error corresponding to those estimated dimensions will be smaller.

Next, we move to the data transmission phase, and derive an achievable sum-rate expression that includes the channel estimation and data transmission parameters of all users. We first determine the optimum transmit directions for all users. Then, we develop an algorithm that maximizes the sum-rate jointly over the individual training durations of all users, the allocation of power of each user between training and data transmission phases, and also the allocation of the data transmission power of each user over its transmit directions. Finally, we provide detailed simulation results that investigates the effects of the power constraint, coherence interval (block length), and channel covariance matrices on our results.

Our contributions in this chapter provide a solution to the data-rate optimization problem jointly over the training and data transmission phases. In both single-user MIMO and MIMO-MAC cases, we first find the optimum training signal that minimizes the mean square error of the channel estimation. Then, we develop algorithms that maximize the achievable rate of the data transmission phase jointly in terms the training and data transmission parameters.

4.1 System Model

We consider a multiple access channel (MAC) with multiple transmit antennas at every user and multiple receive antennas at the receiver. The channel between user k and the receiver is represented by a random matrix $\mathbf{H}_k$ with dimensions of $n_R \times n_T$, where n_R and n_T are the number of antennas at the receiver and at the transmitters, respectively. We consider a block fading scenario where 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 channels, the receiver performs a linear MMSE estimation for the channels of the users 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 fading channel, the transmitters have only the statistical model of the channel. At time n, each transmitter sends a vector $\mathbf{x}_{kn}$, and the received vector is

$$\mathbf{r}_n = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{x}_{kn} + \mathbf{n}_n, \quad n = 1, \ldots, T \tag{4.1}$$

where K is the number of users, $\mathbf{n}_n$ is a zero-mean, identity-covariance complex Gaussian vector at time n, and the entries of $\mathbf{H}_k$ are complex Gaussian random variables. Each user has a power constraint of P_k, averaged over T symbols.

The statistical model that we consider in this chapter, as in the previous chapters, is the "partial CSI with covariance feedback" model. The channel of user k is written as [4]

$$\mathbf{H}_k = \mathbf{Z}_k \mathbf{\Sigma}_k^{1/2} \tag{4.2}$$

where the entries of $\mathbf{Z}_k$ are i.i.d., zero-mean, unit-variance complex Gaussian random variables.

4.2 Joint Optimization for Single-user MIMO

In this section, we will assume that $K = 1$. In our model, a coherence interval, over which the channel is fixed, is divided into two phases: training phase and data transmission phase; see Figure 4.1. 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 $PT = P_t T_t + P_d T_d$.

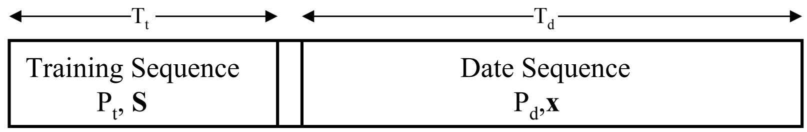


Figure 4.1: Illustration of a single coherence time, over which the channel is fixed.

In a single-user system with partial CSI in the form of the channel covariance matrix at the transmitter, and channel estimation error at the receiver, the optimization problem is to maximize the achievable rate of the data transmission phase. Unlike the case with perfect channel estimation, the data rate here depends on the channel 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, since the channel estimation error contributes to the effective noise. 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.

We will first consider the channel estimation process during the training phase, and choose the training signals to minimize the channel estimation error. Then, we will consider the data transmission phase and develop a lower bound to the capacity, which can be achieved by Gaussian signaling. We will optimize this rate jointly over the rest of the channel estimation parameters and the data transmission parameters.

4.2.1 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 known 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

$$\mathbf{R}_t = \mathbf{H}\mathbf{S} + \mathbf{N}_t \tag{4.3}$$

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 n^{th} column of the matrix equation in (4.3) represents the input-output relationship at time n. The power constraint for the training input signal is $\frac{1}{T_t}\text{tr}(\mathbf{S}\mathbf{S}^\dagger) \leq P_t$.

Due to our channel model in (4.2), 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 $E[\mathbf{h}_i\mathbf{h}_i^\dagger] = \mathbf{\Sigma}, i = 1, \ldots n_R$. Since rows are i.i.d., the receiver can estimate each of them independently using the same training signal. Re-writing (4.3), we get

$$\begin{bmatrix} \mathbf{r}_{t1}^{\dagger} \\ \vdots \\ \mathbf{r}_{tn_R}^{\dagger} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_{1}^{\dagger} \\ \vdots \\ \mathbf{h}_{n_R}^{\dagger} \end{bmatrix} \mathbf{S} + \begin{bmatrix} \mathbf{n}_{t1}^{\dagger} \\ \vdots \\ \mathbf{n}_{tn_R}^{\dagger} \end{bmatrix}. \tag{4.4}$$

Now, the i^{th} row of the above equation can be written as

$$\mathbf{r}_{ti} = \mathbf{S}^{\dagger}\mathbf{h}_i + \mathbf{n}_{ti}. \tag{4.5}$$

The receiver will estimate the i^{th} row of the channel matrix using the received signal $\mathbf{r}_{ti}$, and the training signal $\mathbf{S}$. In general, the estimate $\hat{\mathbf{h}}_i$ can be set to any function of $\mathbf{S}$ and $\mathbf{r}_{ti}$. That is, $\hat{\mathbf{h}}_i = f(\mathbf{S}, \mathbf{r}_{ti})$. However, it is common to use and easier to implement linear MMSE estimation. Also, when the random variables involved in the estimation are Gaussian, as in Rayleigh fading channels, linear MMSE estimation is optimal. In order to find the linear MMSE estimator, we solve the following optimization problem with $\hat{\mathbf{h}}_i = \mathbf{M}\mathbf{r}_{ti}$ as the estimate of $\mathbf{h}_i$, and $\tilde{\mathbf{h}}_i = \mathbf{h}_i - \hat{\mathbf{h}}_i$ as the channel estimation error,

$$\min_{\mathbf{M}} E\left[\tilde{\mathbf{h}}_i^{\dagger}\tilde{\mathbf{h}}_i\right] = \min_{\mathbf{M}} E\left[\mathrm{tr}\left(\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^{\dagger}\right)\right] \tag{4.6}$$

$$= \min_{\mathbf{M}} E\left[\mathrm{tr}\left((\mathbf{h}_i - \mathbf{M}\mathbf{r}_{ti})(\mathbf{h}_i - \mathbf{M}\mathbf{r}_{ti})^{\dagger}\right)\right]. \tag{4.7}$$

Solving the optimum $\mathbf{M}$ from (4.7) is equivalent to solving $\mathbf{M}$ from the orthogonality principle for vector random variables, which is given as [19, page 91],

$$E\left[(\mathbf{h}_i - \mathbf{M}\mathbf{r}_{ti})\mathbf{r}_{ti}^\dagger\right] = \mathbf{0} \tag{4.8}$$

where $\mathbf{0}$ is the $n_T \times T_t$ zero matrix. We can solve $\mathbf{M}$ from (4.8) as

$$\mathbf{M} = E\left[\mathbf{h}_i\mathbf{r}_{ti}^\dagger\right]\left(E\left[\mathbf{r}_{ti}\mathbf{r}_{ti}^\dagger\right]\right)^{-1}. \tag{4.9}$$

By using (4.5), we calculate $E[\mathbf{h}_i\mathbf{r}_{ti}^\dagger] = \mathbf{\Sigma S}$, and $E[\mathbf{r}_{ti}\mathbf{r}_{ti}^\dagger] = \mathbf{S}^\dagger\mathbf{\Sigma S} + \mathbf{I}$. Then, the optimum $\mathbf{M}$ becomes $\mathbf{M} = \mathbf{\Sigma S}(\mathbf{S}^\dagger\mathbf{\Sigma S} + \mathbf{I})^{-1}$. Using this, the mean square error in (4.7) becomes,

$$\min_{\mathbf{M}} E\left[\tilde{\mathbf{h}}_i^\dagger\tilde{\mathbf{h}}_i\right] = \mathrm{tr}\left(\mathbf{\Sigma} - \mathbf{\Sigma S}(\mathbf{S}^\dagger\mathbf{\Sigma S} + \mathbf{I})^{-1}\mathbf{S\Sigma}\right) \tag{4.10}$$

$$= \mathrm{tr}\left(\left(\mathbf{\Sigma}^{-1} + \mathbf{SS}^\dagger\right)^{-1}\right) \tag{4.11}$$

where the last line follows from the matrix inversion lemma [12, page 19]. Note that the mean square error of the channel estimation process can be further decreased by choosing the training signal $\mathbf{S}$ to minimize (4.11). In addition, it is stated in [9] 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 [9]. 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 9 *For given* $\mathbf{\Sigma} = \mathbf{U}_\Sigma \mathbf{\Lambda}_\Sigma \mathbf{U}_\Sigma^\dagger$*,* P_t*,* T_t*, and the power constraint* $tr(\mathbf{S}\mathbf{S}^\dagger) \leq P_t T_t$*, the optimum training input that minimizes the power of the channel estimation error vector is* $\mathbf{S} = \mathbf{U}_\Sigma \mathbf{\Lambda}_S^{1/2}$ *with*

$$\lambda_i^S = \left(\frac{1}{\mu_S} - \frac{1}{\lambda_i^\Sigma} \right)^+, \qquad i = 1, \ldots, \min(n_T, T_t) \tag{4.12}$$

where μ_S^2 *is the Lagrange multiplier that satisfies the power constraint with*

$$\mu_S = \frac{J}{P_t + \sum_{i=1}^{J} \frac{1}{\lambda_i^\Sigma}} \tag{4.13}$$

where J *is the largest index that has non-zero* λ_i^S*.*

Proof: Let us have $\mathbf{S} = \mathbf{U}_S \mathbf{\Lambda}_S^{1/2} \mathbf{V}_S^\dagger$. The expression in (4.11) is minimized when $\mathbf{\Sigma}^{-1}$ and $\mathbf{S}\mathbf{S}^\dagger$ have the same eigenvectors [22]. Therefore, we have $\mathbf{U}_S = \mathbf{U}_\Sigma$. Since, $\mathbf{S}\mathbf{S}^\dagger = \mathbf{U}_S \mathbf{\Lambda}_S \mathbf{U}_S^\dagger$, and the unitary matrix $\mathbf{V}_S$ does not appear in the objective function and the constraint, we can choose $\mathbf{V}_S = \mathbf{I}$. Inserting this into (4.11), the optimization can be written as

$$\tilde{\sigma} = \min_{\mathrm{tr}(\mathbf{\Lambda}_S) \leq P_t T_t} \mathrm{tr} \left(\left(\mathbf{\Lambda}_\Sigma^{-1} + \mathbf{\Lambda}_S \right)^{-1} \right). \tag{4.14}$$

The Langrangian of the problem in (4.14) can be written as

$$\sum_{i=1}^{n_T} \frac{1}{\frac{1}{\lambda_i^\Sigma} + \lambda_i^S} + \mu_S^2 \left(\sum_{i=1}^{n_T} \lambda_i^S - P_t T_t \right) \tag{4.15}$$

where μ_S^2 is the Lagrange multiplier. The solution that satisfies the KKT conditions is water-filling over the eigenvalues of the channel covariance matrix, which can be written as

$$\lambda_i^S = \left(\frac{1}{\mu_S} - \frac{1}{\lambda_i^\Sigma} \right)^+, \qquad i = 1, \ldots, \min(n_T, T_t). \tag{4.16}$$

In order to calculate μ_S, we sum both sides of (4.16) over all antennas to get

$$\mu_S = \frac{J}{P_t + \sum_{i=1}^{J} \frac{1}{\lambda_i^\Sigma}} \tag{4.17}$$

where J is the largest index that has non-zero λ_i^S. $\square$

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 chapter.

Theorem 9 tells us that the optimum transmit directions of the training signal are the eigenvectors of the channel covariance matrix, and the right eigenvector matrix of the training signal is identity. As a result, 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 (4.12). 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.

Note that μ_S is a function of only P_t and T_t, which are given to the problem in Theorem 9, and will be picked as a result of the achievable rate maximization problem in the data transmission phase. 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 9.

Before moving on to the next section, we will calculate the eigenvalues of the covariance matrices of the estimated channel vector, and the channel estimation error vector. Plugging $\mathbf{S}$ into the covariance of the channel estimation error, $\tilde{\mathbf{\Sigma}} = E\left[\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^\dagger\right] = \left(\mathbf{\Sigma}^{-1} + \mathbf{S}\mathbf{S}^\dagger\right)^{-1}$, we find the eigenvectors,

$$\tilde{\mathbf{\Sigma}} = \mathbf{U}_\Sigma \left(\mathbf{\Lambda}_\Sigma^{-1} + \mathbf{\Lambda}_S\right)^{-1} \mathbf{U}_\Sigma^\dagger, \tag{4.18}$$

and by plugging (4.12) into (4.18), we find the eigenvalues of the covariance of the channel estimation error, $\tilde{\mathbf{\Sigma}}$

$$\tilde{\lambda}_i^{\Sigma} = \begin{cases} \mu_S, & \mu_S < \lambda_i^{\Sigma}; \\ \lambda_i^{\Sigma}, & \mu_S > \lambda_i^{\Sigma} \end{cases} = \min\left(\lambda_i^{\Sigma}, \mu_S\right). \tag{4.19}$$

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 [19, page 91]. We have,

$$E\left[\mathbf{h}_i\mathbf{h}_i^{\dagger}\right] = E\left[\hat{\mathbf{h}}_i\hat{\mathbf{h}}_i^{\dagger}\right] + E\left[\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^{\dagger}\right] \tag{4.20}$$

$$\mathbf{\Sigma} = E\left[\hat{\mathbf{h}}_i\hat{\mathbf{h}}_i^{\dagger}\right] + \tilde{\mathbf{\Sigma}}. \tag{4.21}$$

Now, the covariance matrix of the estimated channel becomes,

$$E\left[\hat{\mathbf{h}}_i\hat{\mathbf{h}}_i^\dagger\right] \triangleq \hat{\mathbf{\Sigma}} = \mathbf{U}_\Sigma\mathbf{\Lambda}_\Sigma\mathbf{U}_\Sigma^\dagger - \mathbf{U}_\Sigma\tilde{\mathbf{\Lambda}}_\Sigma\mathbf{U}_\Sigma^\dagger \tag{4.22}$$

$$= \mathbf{U}_\Sigma\left(\mathbf{\Lambda}_\Sigma - \tilde{\mathbf{\Lambda}}_\Sigma\right)\mathbf{U}_\Sigma^\dagger \tag{4.23}$$

$$\triangleq \mathbf{U}_\Sigma\hat{\mathbf{\Lambda}}_\Sigma\mathbf{U}_\Sigma^\dagger. \tag{4.24}$$

The covariance matrix of the estimated channel 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

$$\hat{\lambda}_i^\Sigma = \lambda_i^\Sigma - \tilde{\lambda}_i^\Sigma \tag{4.25}$$

$$= \lambda_i^\Sigma - \min\left(\lambda_i^\Sigma, \mu_S\right) \tag{4.26}$$

$$= \min\left(0, \lambda_i^\Sigma - \mu_S\right). \tag{4.27}$$

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 9.

4.2.2 Data Transmission Phase

When the CSI at the receiver is noisy, the optimum input signaling that achieves the capacity is not known. Following [9, 20, 28, 47], 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 (4.1) as

$$\mathbf{r} = \hat{\mathbf{H}}\mathbf{x} + \tilde{\mathbf{H}}\mathbf{x} + \mathbf{n}. \tag{4.28}$$

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, $\mathrm{tr}(\mathbf{Q}) \leq P_d$. Although the optimum input distribution is not known, we achieve the following rate with Gaussian $\mathbf{x}$ for a MIMO channel [47],

$$C_{lb} = I(\mathbf{r};\mathbf{x}|\hat{\mathbf{H}}) = E\left[\log\left|\mathbf{I} + \mathbf{R}^{-1}_{\tilde{\mathbf{H}}\mathbf{x}+\mathbf{n}}\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^\dagger\right|\right] \tag{4.29}$$

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

$$\mathbf{R}_{\tilde{\mathbf{H}}\mathbf{x}+\mathbf{n}} = E\left[\tilde{\mathbf{H}}\mathbf{x}\mathbf{x}^\dagger\tilde{\mathbf{H}}^\dagger\right] + \mathbf{I} = E\left[\tilde{\mathbf{H}}\mathbf{Q}\tilde{\mathbf{H}}^\dagger\right] + \mathbf{I}. \tag{4.30}$$

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,

$$E\left[\tilde{\mathbf{h}}_i^\dagger\mathbf{Q}\tilde{\mathbf{h}}_j\right] = \mathrm{tr}\left(\mathbf{Q}E\left[\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_j^\dagger\right]\right) \tag{4.31}$$

$$= \begin{cases} \mathrm{tr}(\mathbf{Q}\tilde{\mathbf{\Sigma}}), & \text{when} \quad i=j \\ 0, & \text{when} \quad i\neq j \end{cases} \tag{4.32}$$

which results in $E\left[\tilde{\mathbf{H}}\mathbf{Q}\tilde{\mathbf{H}}^\dagger\right] = \mathrm{tr}(\mathbf{Q}\tilde{\mathbf{\Sigma}})\mathbf{I}$. Now, the rate in (4.29) can be written as

$$C_{lb} = E\left[\log\left|\mathbf{I} + \frac{\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^\dagger}{1+\mathrm{tr}(\mathbf{Q}\tilde{\mathbf{\Sigma}})}\right|\right]. \tag{4.33}$$

Since our goal is to find the largest such achievable rate, the rate maximization problem over the entire block becomes

$$R = \max_{\substack{(\mathbf{Q},P_t,T_t)\in\mathcal{S} \\ \mathrm{tr}(\mathbf{Q})\leq P_d}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^\dagger}{1+\mathrm{tr}(\mathbf{Q}\tilde{\mathbf{\Sigma}})}\right|\right] \tag{4.34}$$

where $\mathcal{S} = \left\{(\mathbf{Q},P_t,T_t)\middle|\mathrm{tr}(\mathbf{Q})T_d + P_tT_t = PT\right\}$, 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.

Transmit Directions

Unlike the case with no-CSI at the transmitters [9], in a single-user system with partial CSI in the form of channel covariance matrix at the transmitter, and noisy CSI at the receiver, 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 to maximize (4.34). The channel covariance matrix $\hat{\mathbf{\Sigma}}$, which is known at the transmitter, and the transmit covariance matrix $\mathbf{Q}$ have the eigenvalue decompositions $\hat{\mathbf{\Sigma}} = \mathbf{U}_\Sigma \hat{\mathbf{\Lambda}}_\Sigma \mathbf{U}_\Sigma^\dagger$, and $\mathbf{Q} = \mathbf{U}_Q \mathbf{\Lambda}_Q \mathbf{U}_Q^\dagger$, respectively.

When the CSI at the receiver is perfect, [14] 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 10 *Let* $\mathbf{\Sigma} = \mathbf{U}_\Sigma \mathbf{\Lambda}_\Sigma \mathbf{U}_\Sigma^\dagger$ *be the spectral decomposition of the covariance feedback matrix of the channel. Then, the optimum transmit covariance matrix* $\mathbf{Q}$ *has the form* $\mathbf{Q} = \mathbf{U}_\Sigma \mathbf{\Lambda}_Q \mathbf{U}_\Sigma^\dagger$.

Proof: In (4.18) and (4.24), we have shown that, when $\mathbf{\Sigma} = \mathbf{U}_\Sigma \mathbf{\Lambda}_\Sigma \mathbf{U}_\Sigma^\dagger$, we have $\hat{\mathbf{\Sigma}} = \mathbf{U}_\Sigma \hat{\mathbf{\Lambda}}_\Sigma \mathbf{U}_\Sigma^\dagger$, and $\tilde{\mathbf{\Sigma}} = \mathbf{U}_\Sigma \tilde{\mathbf{\Lambda}}_\Sigma \mathbf{U}_\Sigma^\dagger$. By using (4.2), we have $\hat{\mathbf{H}} = \hat{\mathbf{Z}} \mathbf{U}_\Sigma \hat{\mathbf{\Lambda}}_\Sigma^{1/2} \mathbf{U}_\Sigma^\dagger$. Inserting these into (4.34), we obtain

$$R = \max_{\substack{(\mathbf{Q},P_t,T_t)\in\mathcal{S} \\ \mathrm{tr}(\mathbf{Q})\leq P_d}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\hat{\mathbf{Z}}\hat{\mathbf{\Lambda}}_\Sigma^{1/2}\mathbf{U}_\Sigma^\dagger\mathbf{Q}\mathbf{U}_\Sigma\hat{\mathbf{\Lambda}}_\Sigma^{1/2}\hat{\mathbf{Z}}^\dagger}{1+\mathrm{tr}\left(\mathbf{U}_\Sigma^\dagger\mathbf{Q}\mathbf{U}_\Sigma\tilde{\mathbf{\Lambda}}_\Sigma\right)}\right|\right] \tag{4.35}$$

where we used the fact that the random matrices $\hat{\mathbf{Z}}\mathbf{U}_\Sigma$ and $\hat{\mathbf{Z}}$ have the same distribution for zero-mean identity-covariance Gaussian $\hat{\mathbf{Z}}$ and unitary $\mathbf{U}_\Sigma$ [42]. We may spectrally decompose the expression sandwiched between $\hat{\mathbf{Z}}$ and its conjugate transpose in (4.35) as

$$\hat{\mathbf{\Lambda}}_\Sigma^{1/2}\mathbf{U}_\Sigma^\dagger\mathbf{Q}\mathbf{U}_\Sigma\hat{\mathbf{\Lambda}}_\Sigma^{1/2} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger. \tag{4.36}$$

Using (4.36), and the identity $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$, we can write the trace expression in the denominator of (4.35) as $\mathrm{tr}\left(\mathbf{U}_\Sigma^\dagger\mathbf{Q}\mathbf{U}_\Sigma\tilde{\mathbf{\Lambda}}_\Sigma\right) = \mathrm{tr}\left(\mathbf{U}^\dagger\hat{\mathbf{\Lambda}}_\Sigma^{-1}\tilde{\mathbf{\Lambda}}_\Sigma\mathbf{U}\mathbf{\Lambda}\right)$, and the optimization problem in (4.35) can be written as

$$R = \max_{\substack{(\mathbf{Q},P_t,T_t)\in\mathcal{S} \\ \mathrm{tr}(\mathbf{Q})\leq P_d}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\hat{\mathbf{Z}}\mathbf{\Lambda}\hat{\mathbf{Z}}^\dagger}{1+\mathrm{tr}\left(\mathbf{U}^\dagger\hat{\mathbf{\Lambda}}_\Sigma^{-1}\tilde{\mathbf{\Lambda}}_\Sigma\mathbf{U}\mathbf{\Lambda}\right)}\right|\right] \tag{4.37}$$

where we again used the fact that the random matrices $\hat{\mathbf{Z}}\mathbf{U}$ and $\hat{\mathbf{Z}}$ have the same distribution. Since, in (4.37), the numerator of the objective function does not involve $\mathbf{U}$, and using [26, Theorem 9.H.1.h, page 249], we know for the denominator that $\mathrm{tr}(\hat{\mathbf{\Lambda}}_\Sigma^{-1}\tilde{\mathbf{\Lambda}}_\Sigma\mathbf{\Lambda}) \leq \mathrm{tr}(\mathbf{U}^\dagger\hat{\mathbf{\Lambda}}_\Sigma^{-1}\tilde{\mathbf{\Lambda}}_\Sigma\mathbf{U}\mathbf{\Lambda})$, for all unitary $\mathbf{U}$, we can choose $\mathbf{U} = \mathbf{I}$ to maximize the rate as long as this choice is feasible. In order to check for the feasibility, we write the trace constraint on $\mathbf{Q}$ using (4.36) as

$$\mathrm{tr}(\mathbf{Q}) = \mathrm{tr}(\mathbf{U}_\Sigma\hat{\mathbf{\Lambda}}_\Sigma^{-1/2}\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger\hat{\mathbf{\Lambda}}_\Sigma^{-1/2}\mathbf{U}_\Sigma^\dagger) \tag{4.38}$$

$$= \mathrm{tr}(\mathbf{U}^\dagger\hat{\mathbf{\Lambda}}_\Sigma^{-1}\mathbf{U}\mathbf{\Lambda}). \tag{4.39}$$

Again from [26, Theorem 9.H.1.h, page 249], $\mathrm{tr}(\hat{\mathbf{\Lambda}}_\Sigma^{-1}\mathbf{\Lambda}) \leq \mathrm{tr}(\mathbf{U}^\dagger\hat{\mathbf{\Lambda}}_\Sigma^{-1}\mathbf{U}\mathbf{\Lambda}) \leq P_d$, for all unitary $\mathbf{U}$. Therefore, we conclude that $\mathbf{U} = \mathbf{I}$ choice is feasible. Then, using $\mathbf{U} = \mathbf{I}$, from (4.36), we have the desired result:

$$\mathbf{Q} = \mathbf{U}_\Sigma\hat{\mathbf{\Lambda}}_\Sigma^{-1}\mathbf{\Lambda}\mathbf{U}_\Sigma \tag{4.40}$$

with $\mathbf{\Lambda}_Q = \hat{\mathbf{\Lambda}}_\Sigma^{-1}\mathbf{\Lambda}$. $\square$

Using Theorem 10, we can write the optimization problem in (4.34) as,

$$R = \max_{\substack{(\mathbf{Q},P_t,T_t)\in\mathcal{S}\\ \mathrm{tr}(\mathbf{Q})\leq P_d}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\hat{\mathbf{Z}}\mathbf{\Lambda}_Q\hat{\mathbf{\Lambda}}_\Sigma\hat{\mathbf{Z}}^\dagger}{1+\mathrm{tr}\left(\mathbf{\Lambda}_Q\tilde{\mathbf{\Lambda}}_\Sigma\right)}\right|\right] \tag{4.41}$$

$$= \max_{(\boldsymbol{\lambda}^Q,P_t,T_t)\in\mathcal{P}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\sum_{i=1}^{n_T}\lambda_i^Q\hat{\lambda}_i^\Sigma\hat{\mathbf{z}}_i\hat{\mathbf{z}}_i^\dagger}{1+\sum_{i=1}^{n_T}\lambda_i^Q\tilde{\lambda}_i^\Sigma}\right|\right] \tag{4.42}$$

where $\boldsymbol{\lambda}^Q = [\lambda_1^Q, \ldots, \lambda_{n_T}^Q]$, $\mathcal{P} = \left\{(\boldsymbol{\lambda}^Q, P_t, T_t) \;\middle|\; \left(\sum_{i=1}^{n_T}\lambda_i^Q\right)T_d + P_tT_t = PT\right\}$, 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 i^{th} column of $\hat{\mathbf{Z}}$. Although the constraint set of the optimization problem is $\left(\sum_{i=1}^{n_T} \lambda_i^Q\right) T_d + P_t T_t \leq PT$, we know that the optimum value is obtained when the summation is equal to PT. Since $\frac{x}{a+x}$ is an increasing function in x, if the summation was strictly less than PT, we could increase the value of the objective function by increasing any one of the λ_i^Q's, while keeping the rest fixed. Therefore, it is sufficient to search over a constraint set, where the inequality is satisfied with equality.

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 (4.42). Although Theorem 10 gives us a very simple closed form solution for the optimum transmit directions, solving (4.42) 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 solves (4.42) for $\boldsymbol{\lambda}^Q$.

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 Chapter 3. In this section, we extend the algorithm in Chapter 3 to the case when there is channel estimation error at the receiver, or in other words, we have the training signal power and the training signal duration in the sum-rate expression. The algorithm in Chapter 3 cannot be trivially generalized to the model in this chapter, since, here we have the training power P_t, and the training duration T_t as additional parameters. Using the algorithm in Chapter 3, we cannot update the value of the training power.

By plugging (4.25) and (4.19) into (4.42), we get

$$R = \max_{(\boldsymbol{\lambda}^Q, P_t, T_t) \in \mathcal{P}} \frac{T - T_t}{T} E \left[\log \left| \mathbf{I} + \frac{\sum_{i=1}^{J} \lambda_i^Q (\lambda_i^\Sigma - \mu_S) \hat{\mathbf{z}}_i \hat{\mathbf{z}}_i^\dagger}{1 + \sum_{i=1}^{J} \lambda_i^Q \mu_S + \sum_{i=J+1}^{n_T} \lambda_i^Q \lambda_i^\Sigma} \right| \right]. \tag{4.43}$$

Note that J and μ_S are functions of P_t and T_t. Since λ_i^Q, for $i = J+1, \ldots, n_T$ does not contribute to the numerator, we should choose $\lambda_i^Q = 0$, for $i = J+1, \ldots, n_T$. This means that the number of unknowns in $\boldsymbol{\lambda}^Q$ that we should solve for is J, i.e., the unknowns are $\lambda_1^Q, \ldots, \lambda_J^Q$. This is to be expected, because we have trained only J transmit directions, and we should now solve for J power values along those directions. Consequently, we have

$$R = \max_{(\boldsymbol{\lambda}^Q, P_t, T_t) \in \mathcal{P}} \frac{T - T_t}{T} E \left[\log \left| \mathbf{I} + \frac{\sum_{i=1}^{J} \lambda_i^Q (\lambda_i^\Sigma - \mu_S) \hat{\mathbf{z}}_i \hat{\mathbf{z}}_i^\dagger}{1 + \mu_S P_d} \right| \right]. \tag{4.44}$$

From Theorem 9, we know that $J \leq T_t$. We further claim that while optimizing the rate, it is sufficient to search over those (P_t, T_t) pairs that result in $J = T_t$. In other words, for any pair (P_t, T_t) that results in $J < T_t$, we can find another pair $(P_t, T_t^{'})$ that results in a higher achievable rate. In order to see this consider a pair (P_t, T_t) that results in $J < T_t$, then let us choose $T_t^{'} = J$. For this choice, the result of Theorem 9 is the same, since the available power can only fill J of the parallel channels, and the amount of power filled over those J channels does not depend in the number of empty channels. Therefore with $(P_t, T_t^{'}) = (P_t, J)$, the estimation process yields the same channel estimate. When we look at (4.44), we see that inside of the expectation is the same for both (P_t, T_t) and $(P_t, T_t^{'})$. However, the coefficient in front of the expectation is higher with $(P_t, T_t^{'})$, since $J = T_t^{'} < T_t$. Therefore $(P_t, T_t^{'})$ yields a higher achievable rate and it is sufficient to search over those (P_t, T_t) pairs that result in $J = T_t$. We can now write (4.44) as

$$R = \max_{(\boldsymbol{\lambda}^Q, P_t, T_t) \in \mathcal{R}} \frac{T - T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\sum_{i=1}^{T_t} \lambda_i^Q (\lambda_i^\Sigma - \mu_S)\hat{\mathbf{z}}_i \hat{\mathbf{z}}_i^\dagger}{1 + \mu_S P_d}\right|\right] \tag{4.45}$$

where $\mathcal{R} = \left\{ \left(\boldsymbol{\lambda}^Q, P_t, T_t\right) \middle| \left(\sum_{i=1}^{n_T} \lambda_i^Q\right) T_d + P_t T_t = PT, P_t > \sum_{i=1}^{T_t} \left(\frac{1}{\lambda_{T_t}^\Sigma} - \frac{1}{\lambda_i^\Sigma}\right)\right\}$, and the condition $P_t > \sum_{i=1}^{T_t} \left(\frac{1}{\lambda_{T_t}^\Sigma} - \frac{1}{\lambda_i^\Sigma}\right)$ 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 (4.45) over are discrete valued T_t, and continuous valued P_t, and $\boldsymbol{\lambda}^Q$. Since, for every value of T_t, both the coefficient in front of the expectation, and the number of terms in the sum in the numerator of (4.45) are different, the form of the objective function is also different. 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 in each one. Then, we take the solution that results in the maximum rate, i.e.,

$$R = \max_{1 \le T_t \le n_T} \max_{(\boldsymbol{\lambda}^Q, P_t) \in \mathcal{R}_{T_t}} \frac{T - T_t}{T} E \left[\log \left| \mathbf{I} + \frac{\sum_{i=1}^{T_t} \lambda_i^Q (\lambda_i^\Sigma - \mu_S) \hat{\mathbf{z}}_i \hat{\mathbf{z}}_i^\dagger}{1 + \mu_S P_d} \right| \right] \tag{4.46}$$

where $\mathcal{R}_{T_t} = \left\{ \left(\boldsymbol{\lambda}^Q, P_t \right) \middle| \left(\sum_{i=1}^{n_T} \lambda_i^Q \right) T_d + P_t T_t = PT, P_t > \sum_{i=1}^{T_t} \left(\frac{1}{\lambda_{T_t}^\Sigma} - \frac{1}{\lambda_i^\Sigma} \right) \right\}$.

While solving the inner maximization problem, we define $f_i(P_t) = \frac{\lambda_i^\Sigma - \mu_S}{1 + \mu_S P_d}$, for $i = 1, \ldots, T_t$. In this case, the inner optimization problem becomes

$$R_{T_t} = \max_{(\boldsymbol{\lambda}^Q, P_t) \in \mathcal{R}_{T_t}} \frac{T - T_t}{T} E \left[\log \left| \mathbf{I} + \sum_{i=1}^{T_t} \lambda_i^Q f_i(P_t) \hat{\mathbf{z}}_i \hat{\mathbf{z}}_i^\dagger \right| \right]. \tag{4.47}$$

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 (4.47) would become exactly the same as the corresponding problem with perfect CSI assumption at the receiver as in Chapter 3, where here, $f_i(P_t)$ replaces λ_i^Σ in (3.7). In the optimization problem in (4.47), we have $T_t + 1$ optimization variables, $\lambda_1^Q, \ldots, \lambda_{T_t}^Q$, and P_t. The Lagrangian for (4.47) can be written as

$$\frac{T - T_t}{T} E \left[\log \left| \mathbf{I} + \sum_{i=1}^{T_t} \lambda_i^Q f_i(P_t) \hat{\mathbf{z}}_i \hat{\mathbf{z}}_i^\dagger \right| \right] - \mu \left(\left(\sum_{i=1}^{T_t} \lambda_i^Q \right) T_d + P_t T_t - PT \right) \tag{4.48}$$

where μ is the Lagrange multiplier, and we omitted the complementary slackness conditions related to the positiveness of λ_i^Q, and $P_t - \sum_{i=1}^{T_t} \left(\frac{1}{\lambda_{T_t}^\Sigma} - \frac{1}{\lambda_i^\Sigma} \right)$. Using the identity in (2.25), the KKT conditions can be written as

$$\frac{T_d}{T} f_i(P_t) E\left[\mathbf{z}_i^\dagger \mathbf{A}^{-1} \mathbf{z}_i\right] \leq \mu T_d, \quad i = 1, \ldots, T_t \tag{4.49}$$

$$\frac{T_d}{T} \sum_{i=1}^{T_t} \lambda_i^Q E\left[\mathbf{z}_i^\dagger \mathbf{A}^{-1} \mathbf{z}_i\right] \frac{\partial f_i(P_t)}{\partial P_t} = \mu T_t \tag{4.50}$$

where $\mathbf{A} = \mathbf{I} + \sum_{i=1}^{T_t} \lambda_i^Q f_i(P_t) \hat{\mathbf{z}}_i \hat{\mathbf{z}}_i^\dagger$, and the equality of the last equation follows from the complementary slackness condition, which says $P_t > \sum_{i=1}^{T_t} \left(\frac{1}{\lambda_{T_t}^\Sigma} - \frac{1}{\lambda_i^\Sigma} \right)$. If the complementary slackness condition is not satisfied, i.e., we had $P_t \leq \sum_{i=1}^{T_t} \left(\frac{1}{\lambda_{T_t}^\Sigma} - \frac{1}{\lambda_i^\Sigma} \right)$, then at least one of the channels out of T_t channels could not be filled, i.e., $J < T_t$, which means this choice of (P_t, T_t) pair is not optimal. Therefore, the complementary slackness condition is always satisfied, resulting in the equality in (4.50).

Note that when the optimum λ_i^Q is non-zero, the corresponding inequality in (4.49) will be satisfied with equality due to its corresponding complementary slackness condition. Therefore, we pull the expectation terms from (4.49) for those equations with non-zero λ_i^Q's, and insert them into (4.50). Since those indices with $\lambda_i^Q = 0$ do not contribute to (4.50), we have

$$\frac{T_d}{T} \sum_{i=1}^{T_t} \lambda_i^Q \frac{\mu T}{f_i(P_t)} \frac{\partial f_i(P_t)}{\partial P_t} = \mu T_t. \tag{4.51}$$

By canceling out μ's on both sides, we get

$$\sum_{i=1}^{T_t} \lambda_i^Q \frac{f_i'(P_t)}{f_i(P_t)} = \frac{T_t}{T_d}. \tag{4.52}$$

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 (4.49) is the same as the KKT conditions in Chapter 3. Therefore, we propose to update λ_i^Q in the same way as in Chapter 3, and between the iterations solve (4.52) to update P_t. At any given iteration, our algorithm first solves $P_t(n+1)$ from

$$\sum_{i=1}^{T_t} \lambda_i^Q(n) \frac{f_i'(P_t(n+1))}{f_i(P_t(n+1))} = \frac{T_t}{T_d} \tag{4.53}$$

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

$$\lambda_i^Q(n+1) = \frac{\lambda_i^Q(n) f_i(P_t(n+1)) E\left[\mathbf{z}_i^\dagger \mathbf{A}^{-1} \mathbf{z}_i\right]}{\sum_{j=1}^{n_T} \lambda_j^Q(n) f_j(P_t(n+1)) E\left[\mathbf{z}_j^\dagger \mathbf{A}^{-1} \mathbf{z}_j\right]} \frac{(PT - P_t(n+1)T_t)}{T_d}, \quad i = 1, \ldots, T_t \tag{4.54}$$

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 $1 \leq T_t \leq n_T$. We run n_T such algorithms, and the solution of (4.45) is found by taking the one that results in the largest rate, which gives us the solution for the training phase duration T_t.

As a result, we solved the joint channel estimation and resource allocation problem that we considered in this chapter. Through the solutions for T_t and P_t, we find the solution for the allocation of available time and power over the training and data transmission phases, since total block length and power is fixed. Through Theorem 10, we find the optimum transmit directions, and through $\lambda_1^Q, \ldots, \lambda_{T_t}^Q$, we find the solution for the allocation of data transmission power over these transmit directions. Finally, the optimum training signal $\mathbf{S}$ that minimizes the mean square error is determined by T_t and P_t through Theorem 9.

4.2.3 Numerical Results for Single-user MIMO

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, and seems to be intractable for now. However, in our extensive simulations, we observed that the algorithm always converged.

We start our numerical analysis with the single-user case. We first consider a system having $n_T = n_R = 2$ with 10 dB total average power and block length $T = 4$. In Figure 4.2, we plot the eigenvalues of the data transmit matrix and the 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 results in the highest rate.

Next, we investigate the effect of total average power on the number of estimated channel dimensions. We observe that if we keep the block length small at $T = 4$, the amount of total power required in order to estimate the second channel dimension is very high. In Figure 4.3, for a 40 dB total average power, we plot the eigenvalues of the data transmit matrix and the training power as a function of the iteration index for both possible values of the training signal duration, and we see that achievable

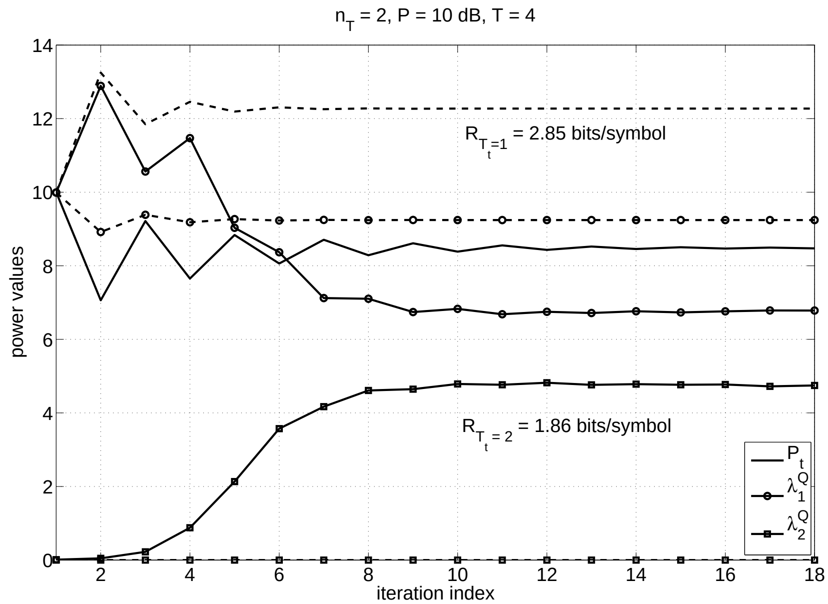


Figure 4.2: The convergence of the single-user algorithm with $n_T = n_R = 2$, 10 dB total average power and $T = 4$. The dashed curves correspond to one symbol long training, $T_t = 1$, and solid curves correspond to two symbols long training, $T_t = 2$.

rate with two symbols of training is barely higher than the achievable rate with one symbol of training. We repeat this experiment with different numbers of antennas and channel eigenvalues, and we see that we need very high power levels in order to use more than one symbol of training. This suggests that the block length, i.e., the coherence interval, is more important for determining the duration of the training phase.

In order to investigate the effect of the block length, in Figure 4.4, we consider 10 dB total average power, and block length $T = 20$. We observe that similar to the high SNR case, in this case as well, having two symbols long training phase results in higher rates. We repeat this experiment with different numbers of antennas, and channel eigenvalues for long block lengths, and we see that moderate block lengths are

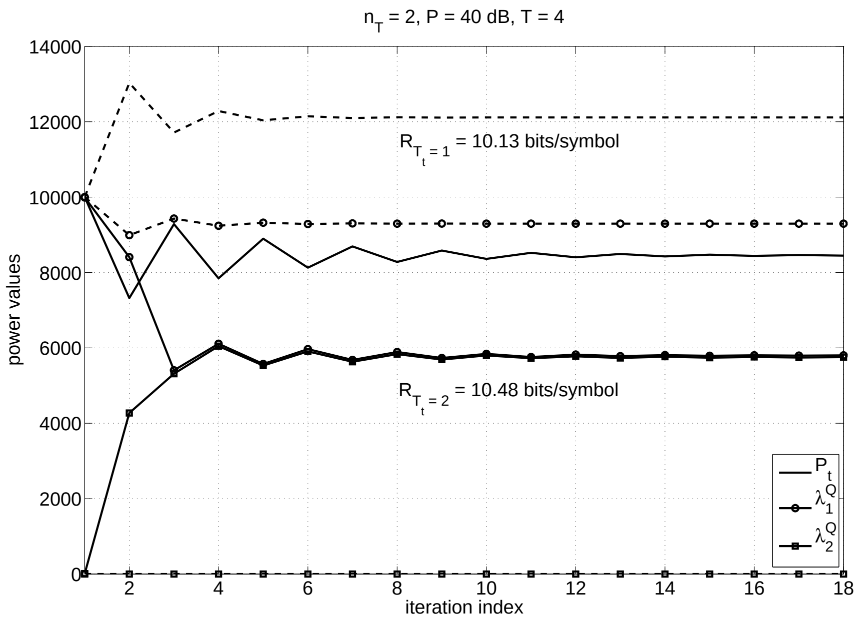


Figure 4.3: The convergence of the single-user algorithm with $n_T = n_R = 2$, 40 dB total average power and $T = 4$. The dashed curves correspond to one symbol long training, $T_t = 1$, and solid curves correspond to two symbols long training, $T_t = 2$.

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, and for low SNR systems, 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, 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 4.5, we consider 10 dB average power, and a channel covariance matrix that has a first eigen-direction much stronger than the second eigen-direction, i.e., the largest eigenvalue of the channel covariance matrix is much larger than the second largest

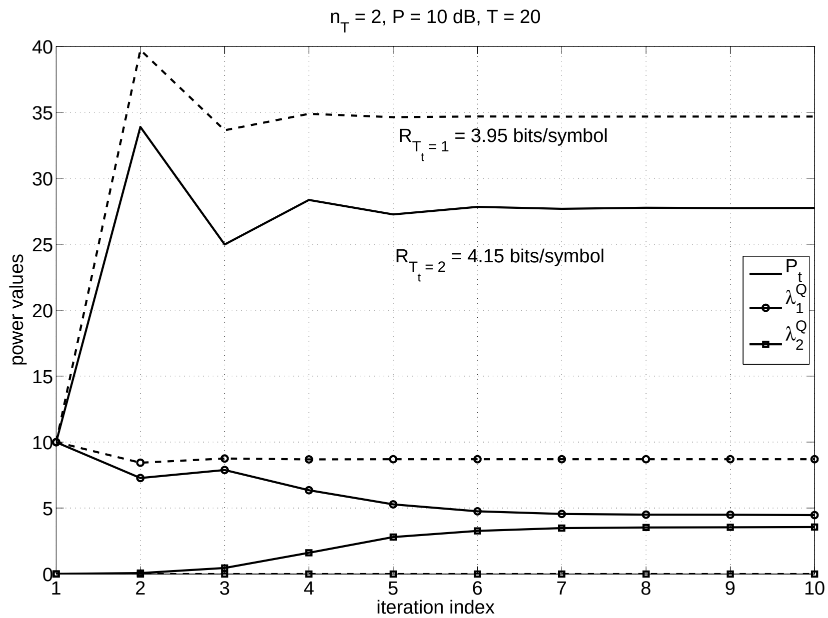
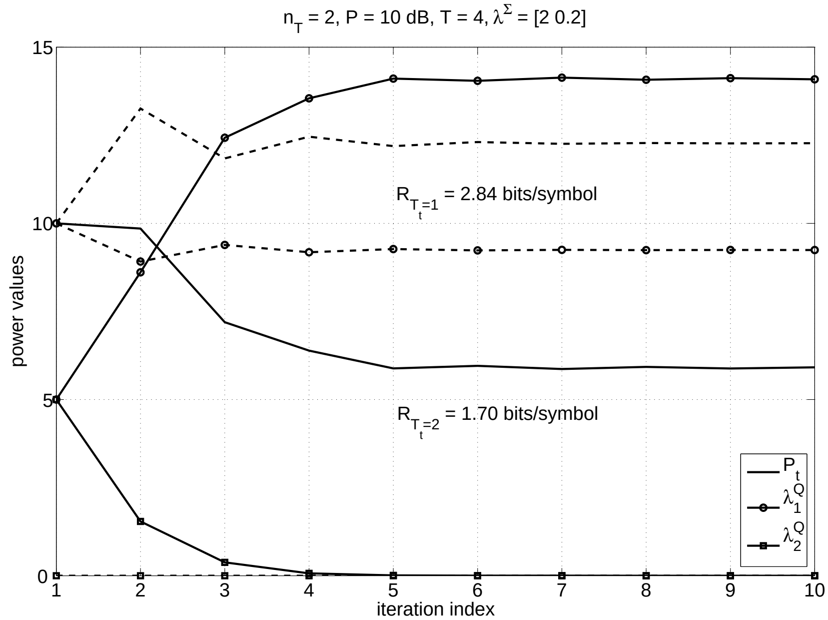


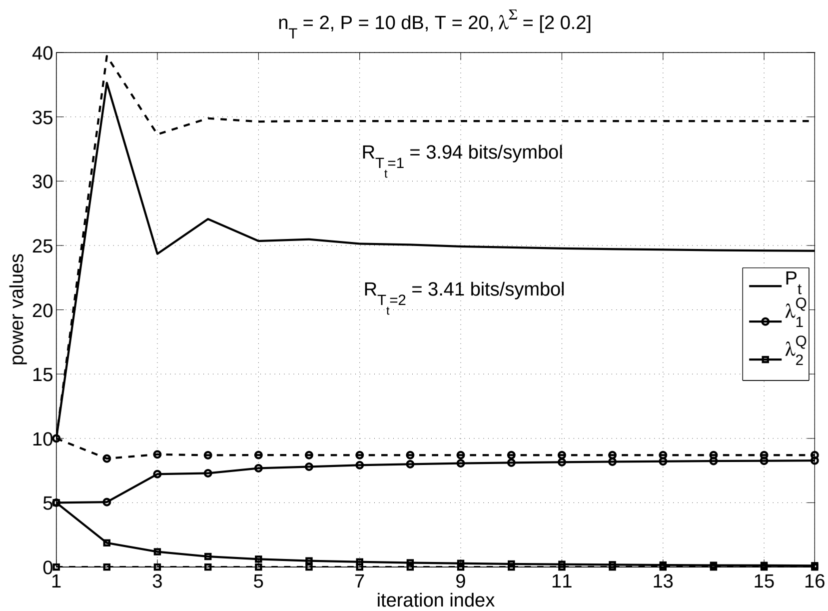
Figure 4.4: The convergence of the single-user algorithm with $n_T = n_R = 2$, 10 dB total average power and $T = 20$. The dashed curves correspond to one symbol long training, $T_t = 1$, and solid curves correspond to two symbols long training, $T_t = 2$.

eigenvalue. In such scenarios, even if the block length is large, beamforming turns out to be the optimal strategy for the data transmission period. Therefore, estimating the second dimension is a waste of resources, because no power will be allocated to that direction in the data transmission phase. Confirming this intuition, in Figure 4.5, for the cases when $T_t = 2$, the power allocated to the second eigen-direction is zero, although the training power is large enough to estimate both channels.

Another extreme for the eigenvalues of the channel covariance matrix is the case when both eigenvalues are equal, which is considered in Figure 4.6. Note that this case is exactly the case considered [9]. However, in this thesis, we do not assume the restriction that $T_t \geq n_T$ as it was assumed in [9] by reasoning that one needs at least $T_t \geq n_T$ measurements in order to estimate n_T variables. Although this reasoning

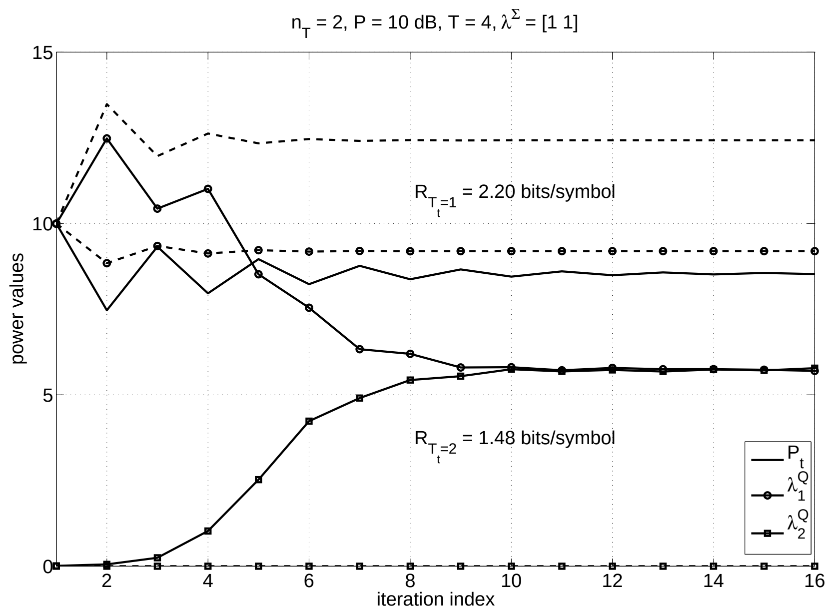


(a)

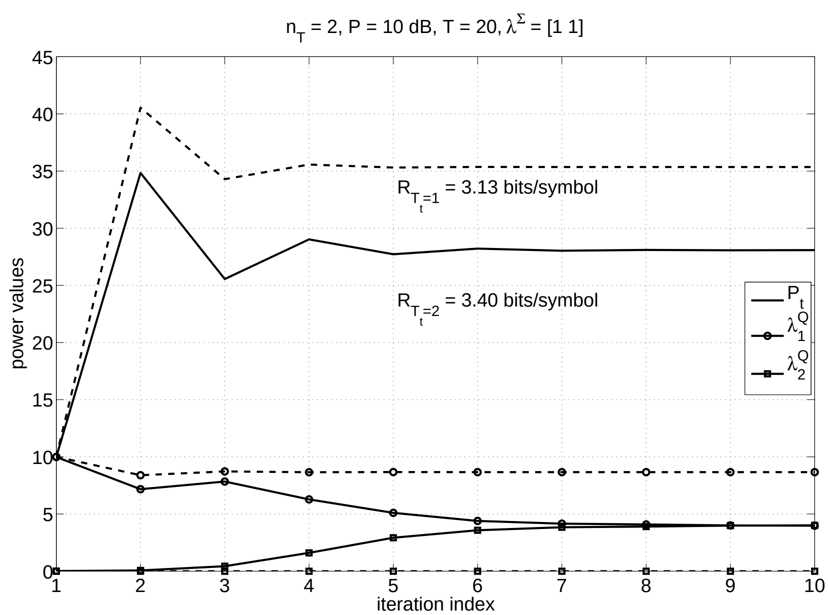


(b)

Figure 4.5: The convergence of the single-user algorithm with $n_T = n_R = 2$, 10 dB total average power, and channel eigenvalues $\boldsymbol{\lambda}^{\Sigma} = [2, 0.2]$, 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) $T = 20$.



(a)



(b)

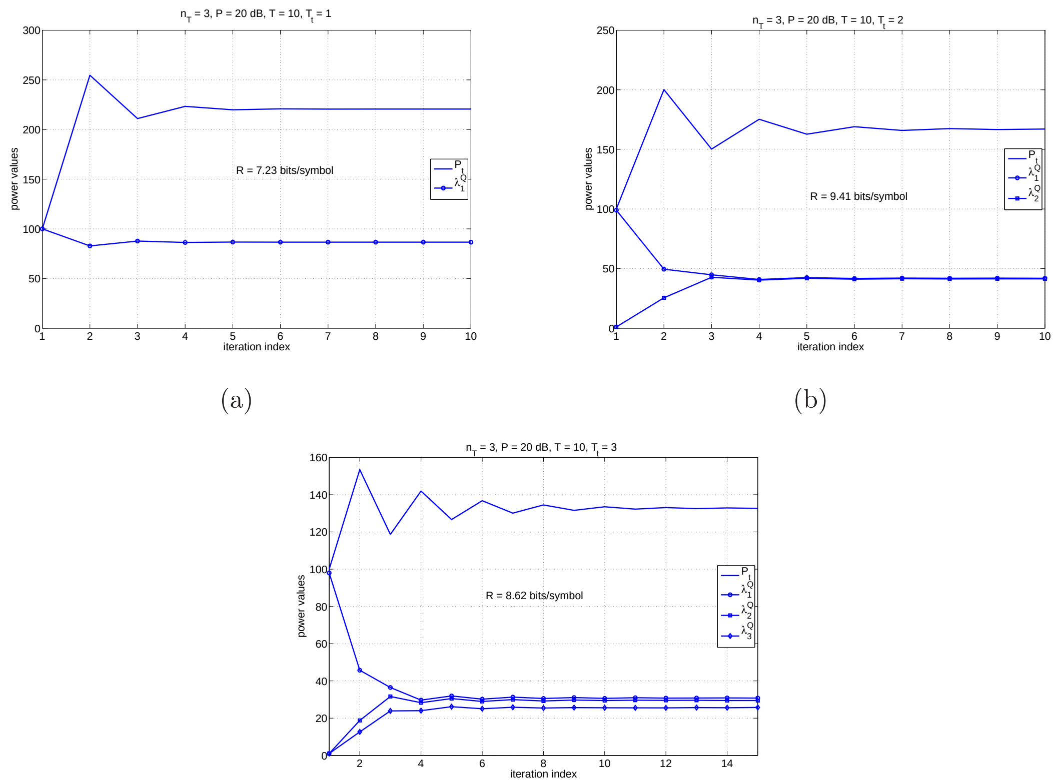
Figure 4.6: The convergence of the single-user algorithm with $n_T = n_R = 2$, 10 dB total average power, and channel eigenvalues $\boldsymbol{\lambda}^{\Sigma} = [1, 1]$, where dashed curves correspond to one symbol long training, $T_t = 1$, and solid curves correspond to two symbols long training, $T_t = 2$: (a) $T = 4$; (b) $T = 20$.

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 thesis, 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 4.6 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.

Finally, we consider a larger system with $n_T = n_R = 3$ having power, $P = 20$ dB, and block length, $T = 10$. For this system, we run our algorithm for all three possible values of the training symbol duration, i.e., $T_t = 1, 2, 3$. We observe in Figure 4.7 that estimating two of the three dimensions of the channel results in the highest rate for this setting.

4.3 Joint Optimization for Multi-user MIMO

In this section, we will consider the multi-user case, where there are K users in the system and a single receiver. Note that in our model, a transmission block is divided into training and data transmission phases. During the training phase, each user has training signal $\mathbf{S}_k$, training signal power P_{t_k}, and training signal duration T_t. During the data transmission phase, each user has data transmission power P_{d_k}, which appears as a constraint on the trace of the transmit covariance matrix. Our goal in



(a) (b) (c)

Figure 4.7: The convergence of the single-user algorithm with $n_T = n_R = 3$, 20 dB total average power and $T = 10$: (a) one symbol long training, $T_t = 1$; (b) two symbols long training, $T_t = 2$; (c) three symbols long training, $T_t = 3$.

this section is to find the optimum values of these training and the data transmission parameters for all users.

In a MIMO-MAC with partial CSI in the form of the channel covariance matrix at the transmitters, and channel estimation error at the receiver, the optimization problem is to maximize the achievable sum-rate of the data transmission phase jointly over the channel estimation parameters and the data transmission parameters. Similar to the single-user case, we will first consider the channel estimation problem during the training phase, and choose the training signals to minimize the channel estimation error. Then, we will consider the data transmission phase and develop a lower bound to the sum-capacity which can be achieved by Gaussian signaling. We will optimize this achievable rate jointly over both the channel estimation and the data transmission parameters.

4.3.1 Training and Channel Estimation Phase

For a multiple access channel, we write the input-output relationship during the training phase as

$$\mathbf{R}_t = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{S}_k + \mathbf{N}_t \tag{4.55}$$

where $\mathbf{S}_k$ is an $n_T \times T_t$ dimensional training signal for user k that will be chosen and known at both ends, $\mathbf{R}_t$ and $\mathbf{N}_t$ are $n_R \times T$ dimensional received signal and noise matrices, respectively. The n^{th} column of the matrix equation in (4.55) represents the input-output relationship at time n. The power constraint for the training input signal for user k is $\frac{1}{T_t}\mathrm{tr}(\mathbf{S}\mathbf{S}^\dagger) \leq P_{t_k}$.

Since the receiver is supposed to estimate the channels of all users during the same training phase with the knowledge of all training symbols, it can regard the multi-user channel as a single-user channel, where the channel and the training signal matrices of users are stacked together. We can then write (4.55) equivalently as

$$\mathbf{R}_t = \bar{\mathbf{H}}\bar{\mathbf{S}} + \mathbf{N}_t \tag{4.56}$$

where $\bar{\mathbf{H}} = [\mathbf{H}_1, \ldots, \mathbf{H}_K]$ is an $n_R \times Kn_T$ dimensional channel matrix, and $\bar{\mathbf{S}} = [\mathbf{S}_1^\dagger, \ldots, \mathbf{S}_K^\dagger]^\dagger$ is a $Kn_T \times T_t$ dimensional training signal matrix. Note that, we put the channel matrices next to each other to form longer rows, and the training symbols on top of each other to form longer columns. In this equivalent problem, the receiver will estimate $\bar{\mathbf{H}}$ using the output $\mathbf{R}_t$ and the training signal $\bar{\mathbf{S}}$.

Due to our channel model in (4.2), the entries in a row of $\mathbf{H}_k$ are correlated, and the entries in a column of $\mathbf{H}_k$ are uncorrelated. In other words, for each user, row i of the channel matrix is i.i.d. with row j. This also holds for the stacked matrix, $\bar{\mathbf{H}}$. Let us represent row i of $\mathbf{H}_k$ as $\mathbf{h}_{ki}^\dagger$, where $E[\mathbf{h}_{ki}\mathbf{h}_{ki}^\dagger] = \mathbf{\Sigma}_k, i = 1, \ldots n_R$, and row i of $\bar{\mathbf{H}}$ as $\bar{\mathbf{h}}_i = [\mathbf{h}_{1i}^\dagger, \ldots, \mathbf{h}_{Ki}^\dagger]^\dagger$, where $\bar{\mathbf{\Sigma}} = E[\bar{\mathbf{h}}_i\bar{\mathbf{h}}_i^\dagger] = \mathrm{diag}\{\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_K\}$ is a block diagonal matrix, having $\mathbf{\Sigma}_k$ on its diagonals.

Let the eigenvalue representation of the channel covariance matrix of user k be $\mathbf{\Sigma}_k = \mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{\Sigma_k}\mathbf{U}_{\Sigma_k}^\dagger$, then the eigenvectors of the stacked channel covariance matrix $\bar{\mathbf{\Sigma}} = \bar{\mathbf{U}}_\Sigma\bar{\mathbf{\Lambda}}_\Sigma\bar{\mathbf{U}}_\Sigma^\dagger$ can also be written as $\bar{\mathbf{U}}_\Sigma = \mathrm{diag}\{\mathbf{U}_{\Sigma_1}, \ldots, \mathbf{U}_{\Sigma_K}\}$ [12, Lemma 1.3.10], which is a block diagonal matrix as well.

Since a row of $\bar{\mathbf{H}}$ is formed by combining the rows of all $\mathbf{H}_k$ into a single, and longer row, we can conclude that the rows of $\bar{\mathbf{H}}$ are also i.i.d., and the receiver can estimate each of them independently using the same training symbols. The i^{th} row of (4.56) can be written as

$$\mathbf{r}_{ti} = \bar{\mathbf{S}}^\dagger\bar{\mathbf{h}}_i + \mathbf{n}_{ti}. \tag{4.57}$$

Since this is equivalent to to a single-user channel estimation problem in (4.4) with the exception of a block diagonal channel covariance matrix, we can use the MMSE estimation results of the single-user case. Denoting the estimate of $\bar{\mathbf{h}}_i$ as $\hat{\mathbf{h}}_i = \bar{\mathbf{M}}\mathbf{r}_{ti}$, and the channel estimation error as $\tilde{\mathbf{h}}_i = \bar{\mathbf{h}}_i - \hat{\mathbf{h}}_i$, the MMSE estimation problem can be written as

$$\min_{\bar{\mathbf{M}}} E\left[\tilde{\mathbf{h}}_i^\dagger \tilde{\mathbf{h}}_i\right] = \min_{\bar{\mathbf{M}}} E\left[\operatorname{tr}\left(\tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^\dagger\right)\right] \tag{4.58}$$

$$= \min_{\bar{\mathbf{M}}} E\left[\operatorname{tr}\left((\bar{\mathbf{h}}_i - \bar{\mathbf{M}}\mathbf{r}_{ti})(\bar{\mathbf{h}}_i - \bar{\mathbf{M}}\mathbf{r}_{ti})^\dagger\right)\right]. \tag{4.59}$$

Using the orthogonality principle [19, page 91] as in the single-user case, we can find the optimum estimator as

$$\bar{\mathbf{M}}^* = \bar{\mathbf{\Sigma}}\bar{\mathbf{S}}\left(\bar{\mathbf{S}}^\dagger \bar{\mathbf{\Sigma}}\bar{\mathbf{S}} + \mathbf{I}\right)^{-1}. \tag{4.60}$$

Using this, the mean square error in (4.59) becomes,

$$\min_{\bar{\mathbf{M}}} E\left[\tilde{\mathbf{h}}_i^\dagger \tilde{\mathbf{h}}_i\right] = \operatorname{tr}\left(\bar{\mathbf{\Sigma}} - \bar{\mathbf{\Sigma}}\bar{\mathbf{S}}(\bar{\mathbf{S}}^\dagger \bar{\mathbf{\Sigma}}\bar{\mathbf{S}} + \mathbf{I})^{-1}\bar{\mathbf{S}}\bar{\mathbf{\Sigma}}\right) \tag{4.61}$$

$$= \operatorname{tr}\left(\left(\bar{\mathbf{\Sigma}}^{-1} + \bar{\mathbf{S}}\bar{\mathbf{S}}^\dagger\right)^{-1}\right) \tag{4.62}$$

where the last line follows from the matrix inversion lemma [12, page 19]. Note that mean square error of the channel estimation process can be further decreased by choosing the training signal $\bar{\mathbf{S}}$ to minimize (4.62). The following theorem finds $\bar{\mathbf{S}}$, and training signals of individual users $\mathbf{S}_k$, for a given training power and training duration.

Theorem 11 *For given* $\mathbf{\Sigma}_k = \mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{\Sigma_k}\mathbf{U}_{\Sigma_k}^{\dagger}$, P_{t_k}, T_t, *and the power constraints* $tr(\mathbf{S}_k\mathbf{S}_k^{\dagger}) \leq P_{t_k}T_t$, *the* $Kn_T \times T_t$ *dimensional optimum stacked training signal* $\bar{\mathbf{S}}$ *that minimizes the total power of the channel estimation error vector is* $\bar{\mathbf{S}} = \bar{\mathbf{U}}_{\Sigma}\bar{\mathbf{\Lambda}}_S^{1/2}$, *and the* $n_T \times K$ *dimensional optimum training signal of user* k *is* $\mathbf{S}_k = \Big[\mathbf{0}, \ldots, \mathbf{0}, \mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{S_k}^{1/2}, \mathbf{0}, \ldots, \mathbf{0}\Big]$ *with*

$$\lambda_{ki}^S = \left(\frac{1}{\mu_k^S} - \frac{1}{\lambda_{ki}^{\Sigma}}\right)^+, \qquad i = 1, \ldots, \min(n_T, T_{t_k}) \tag{4.63}$$

where $(\mu_k^S)^2$ *is the Lagrange multiplier that satisfies the power constraint with*

$$\mu_k^S = \frac{J_k}{P_{t_k} + \sum_{i=1}^{J_k} \frac{1}{\lambda_{ki}^{\Sigma}}} \tag{4.64}$$

where J_k *is the largest index that has non-zero* λ_{ki}^S *for user* k.

Proof: Let us have $\bar{\mathbf{S}} = \bar{\mathbf{U}}_S\bar{\mathbf{\Lambda}}_S^{1/2}\bar{\mathbf{V}}_S^{\dagger}$. The expression in (4.62) is minimized when $\bar{\mathbf{\Sigma}}^{-1}$ and $\bar{\mathbf{S}}\bar{\mathbf{S}}^{\dagger}$ have the same eigenvectors [22]. Therefore, we have $\bar{\mathbf{U}}_S = \bar{\mathbf{U}}_{\Sigma}$. Since, $\bar{\mathbf{S}}\bar{\mathbf{S}}^{\dagger} = \bar{\mathbf{U}}_S\bar{\mathbf{\Lambda}}_S\bar{\mathbf{U}}_S^{\dagger}$, and the unitary matrix $\bar{\mathbf{V}}_S$ does not appear in the objective function and the constraint, we can choose $\bar{\mathbf{V}}_S = \mathbf{I}$. Now, we have

$$\bar{\mathbf{S}} = \bar{\mathbf{U}}_\Sigma \bar{\mathbf{\Lambda}}_S^{1/2} \tag{4.65}$$

$$\begin{bmatrix} \mathbf{S}_1 \\ \vdots \\ \mathbf{S}_K \end{bmatrix} = \begin{bmatrix} \mathbf{U}_{\Sigma_1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{U}_{\Sigma_K} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_{S_1}^{1/2} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{\Lambda}_{S_K}^{1/2} \end{bmatrix} \tag{4.66}$$

where each user has $\mathbf{S}_k = \left[\mathbf{0}, \ldots, \mathbf{0}, \mathbf{U}_{\Sigma_k} \mathbf{\Lambda}_{S_k}^{1/2}, \mathbf{0}, \ldots, \mathbf{0}\right]$. Note that $\mathbf{S}_k$ is an $n_T \times T_t$ dimensional matrix, and $\mathbf{U}_{\Sigma_k}$ is an $n_T \times n_T$ dimensional matrix. Let us denote the dimension of $\mathbf{\Lambda}_{S_k}$ as $n_T \times T_{t_k}$ in such a way that $\sum_{k=1}^{K} T_{t_k} = T_t$.

Inserting $\bar{\mathbf{S}}$ into (4.62), the optimization problem can be written as

$$\tilde{\sigma}_{sum} = \min_{\substack{\mathrm{tr}(\mathbf{\Lambda}_{S_k}) \leq P_{t_k} T_{t_k} \\ k=1,\ldots,K}} \mathrm{tr}\left(\left(\bar{\mathbf{\Lambda}}_\Sigma^{-1} + \bar{\mathbf{\Lambda}}_S \right)^{-1} \right) \tag{4.67}$$

$$= \min_{\substack{\mathrm{tr}(\mathbf{\Lambda}_{S_k}) \leq P_{t_k} T_{t_k} \\ k=1,\ldots,K}} \sum_{k=1}^{K} \mathrm{tr}\left(\left(\mathbf{\Lambda}_{\Sigma_k}^{-1} + \mathbf{\Lambda}_{S_k} \right)^{-1} \right). \tag{4.68}$$

The Langrangian of the problem in (4.68) can be written as

$$\sum_{k=1}^{K} \sum_{i=1}^{n_T} \frac{1}{\frac{1}{\lambda_{ki}^{\Sigma}} + \lambda_{ki}^{S}} + \sum_{k=1}^{K} (\mu_k^S)^2 \left(\sum_{i=1}^{n_T} \lambda_{ki}^S - P_{t_k} T_{t_k} \right) \tag{4.69}$$

where $(\mu_k^S)^2$ are the Lagrange multipliers. The solution that satisfies the KKT conditions is water-filling the available power of each user over the eigenvalues of its own channel covariance matrix. The solution for user k can be written as

$$\lambda_{ki}^S = \left(\frac{1}{\mu_k^S} - \frac{1}{\lambda_{ki}^\Sigma}\right)^+, \qquad i = 1, \ldots, \min(n_T, T_{t_k}) \tag{4.70}$$

In order to calculate μ_k^S, we sum both sides of (4.70) for user k, over all antennas to get

$$\mu_S = \frac{J_k}{P_{t_k} + \sum_{i=1}^{J_k} \frac{1}{\lambda_i^\Sigma}} \tag{4.71}$$

where J_k is the largest index that has non-zero λ_{ki}^S. $\square$

Similar to the single-user case, for any given P_{t_k}, and $T_{t_k} > n_T$, the effect of training length is completely eliminated from the problem, i.e., increasing T_{t_k} beyond n_T does not result in better channel estimates. However, larger T_{t_k} will result in smaller data transmission length, and will decrease the achievable rate of the data transmission phase. Therefore, it is sufficient to consider only $T_{t_k} \leq n_T$, which we will assume for the rest of this chapter.

Theorem 11 states that orthogonality in the time domain holds over the users in a multi-user setting as well. Since the receiver can stack up the channels for the channel estimation process, and the resulting stacked channel covariance matrix is block-diagonal, the problem is equivalent to a single-user problem where all transmit antennas are on the same unit, but antennas are put into K groups. Each group is correlated within the group, but groups are uncorrelated, which results in a block diagonal channel covariance matrix. Since this is an equivalent single-user problem, training signals of different users are orthogonal in time. This is achieved by transmitting the training signal of user k during its own time slot for T_{t_k} channel uses. Although this might seem counter-intuitive at first, after the diagonalization of the channel, both in the multi-user and single-user cases, we are left with orthogonal channels. Therefore, in order to estimate orthogonal channels, sending orthogonal training signals is sufficient.

Due to the constraint on the training duration, fewer dimensions of the individual channels will be estimated for each user, which will result in shorter individual training durations compared to a single-user case. However, by the conservation of energy, the training signal power of a particular user will be larger compared to the training signal power of the same user in a single-user environment. Therefore, although fewer dimensions of the channel are estimated, the channel estimation error corresponding to those estimated dimension will be smaller.

It was shown in other contexts as well that the degrees of freedom of a MAC does not increase by increasing the number of users. For example in [48], the degrees of freedom is limited by the number of receive antennas. In our case, it is limited by the duration of the training phase, which itself depends on several variables including the number of receive antennas.

Note that μ_k^S is a function of only P_{t_k} and T_{t_k}, both of which will be chosen to maximize the sum-rate of the data transmission phase. The value of T_{t_k} determines the total number of available parallel channels for user k, and the value of P_{t_k} determines the number of channels that will be estimated. The parametric values of P_{t_k} and T_{t_k} will appear in the sum-rate formula. After the sum-rate maximization is performed, the optimum P_{t_k} and T_{t_k} will be found, and this in turn, will give us the optimum $\mathbf{S}_k$ through Theorem 11.

Before moving on to the next section, we will calculate the eigenvalues of the covariance matrices of the estimated channel vector, and the channel estimation error vector for all users. Plugging $\bar{\mathbf{S}}$ into the covariance of the channel estimation error, $\tilde{\mathbf{\Sigma}} = E\left[\tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^\dagger\right] = \left(\bar{\mathbf{\Sigma}}^{-1} + \bar{\mathbf{S}}\bar{\mathbf{S}}^\dagger\right)^{-1}$, we find the eigenvectors,

$$\tilde{\mathbf{\Sigma}} = \bar{\mathbf{U}}_\Sigma \left(\bar{\mathbf{\Lambda}}_\Sigma^{-1} + \bar{\mathbf{\Lambda}}_S\right)^{-1} \bar{\mathbf{U}}_\Sigma^\dagger, \tag{4.72}$$

from where, we conclude

$$\tilde{\mathbf{\Sigma}}_k = \mathbf{U}_{\Sigma_k} \left(\mathbf{\Lambda}_{\Sigma_k}^{-1} + \mathbf{\Lambda}_{S_k}\right)^{-1} \mathbf{U}_{\Sigma_k}^\dagger, \tag{4.73}$$

and by plugging (4.70) into (4.73), we find the eigenvalues of the covariance of the channel estimation error of user k, $\tilde{\mathbf{\Sigma}}_k$

$$\tilde{\lambda}_{ki}^\Sigma = \begin{cases} \mu_k^S, & \mu_k^S < \lambda_{ki}^\Sigma; \\ \lambda_{ki}^\Sigma, & \mu_k^S > \lambda_{ki}^\Sigma \end{cases} = \min\left(\lambda_{ki}^\Sigma, \mu_k^S\right). \tag{4.74}$$

Next, we will calculate the eigenvalues of the covariance of the channel estimate of user k. Using the orthogonality property of the MMSE estimation, $\hat{\mathbf{h}}_i$ and $\tilde{\mathbf{h}}_i$ are uncorrelated [19, page 91]. We have,

$$E\left[\bar{\mathbf{h}}_i\bar{\mathbf{h}}_i^\dagger\right] = E\left[\hat{\mathbf{h}}_i\hat{\mathbf{h}}_i^\dagger\right] + E\left[\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^\dagger\right] \tag{4.75}$$

$$\bar{\mathbf{\Sigma}} = E\left[\hat{\mathbf{h}}_i\hat{\mathbf{h}}_i^\dagger\right] + \tilde{\mathbf{\Sigma}}. \tag{4.76}$$

Now, the covariance matrix of the estimated stacked channel becomes,

$$E\left[\hat{\mathbf{h}}_i\hat{\mathbf{h}}_i^\dagger\right] \triangleq \hat{\mathbf{\Sigma}} = \bar{\mathbf{U}}_\Sigma\bar{\mathbf{\Lambda}}_\Sigma\bar{\mathbf{U}}_\Sigma^\dagger - \bar{\mathbf{U}}_\Sigma\tilde{\mathbf{\Lambda}}_\Sigma\bar{\mathbf{U}}_\Sigma^\dagger \tag{4.77}$$

$$= \bar{\mathbf{U}}_\Sigma\left(\bar{\mathbf{\Lambda}}_\Sigma - \tilde{\mathbf{\Lambda}}_\Sigma\right)\bar{\mathbf{U}}_\Sigma^\dagger \tag{4.78}$$

$$\triangleq \bar{\mathbf{U}}_\Sigma\hat{\mathbf{\Lambda}}_\Sigma\bar{\mathbf{U}}_\Sigma^\dagger \tag{4.79}$$

from where, we conclude

$$\hat{\mathbf{\Sigma}}_k = \mathbf{U}_{\Sigma_k}\hat{\mathbf{\Lambda}}_{\Sigma_k}\mathbf{U}_{\Sigma_k}^\dagger. \tag{4.80}$$

We can write each eigenvalue of the covariance matrix of the estimated channel of user k as

$$\hat{\lambda}_{ki}^{\Sigma} = \min\left(\lambda_{ki}^{\Sigma} - \mu_k^S, 0\right). \tag{4.81}$$

In the next section, we will plug in the values of the channel estimation error matrix and its covariance, the estimate of the channel and its covariance, and the training parameters into the sum-rate formula and develop an algorithm to solve the sum-rate maximization problem jointly over all of the involved parameters.

4.3.2 Data Transmission Phase

The sum-rate of a multiple access channel can be derived using the stacked channel and input matrices. We can write (4.55) as

$$
\begin{aligned}
\mathbf{r} &= \sum_{k=1}^{K} \hat{\mathbf{H}}_k \mathbf{x}_k + \sum_{k=1}^{K} \tilde{\mathbf{H}}_k \mathbf{x}_k + \mathbf{n} &\quad (4.82)\\
&= \hat{\mathbf{H}}\bar{\mathbf{x}} + \tilde{\mathbf{H}}\bar{\mathbf{x}} + \mathbf{n} &\quad (4.83)
\end{aligned}
$$

where $\hat{\mathbf{H}} = [\hat{\mathbf{H}}_1, \ldots, \hat{\mathbf{H}}_K]$, $\tilde{\mathbf{H}} = [\tilde{\mathbf{H}}_1, \ldots, \tilde{\mathbf{H}}_K]$ are $n_R \times Kn_T$ dimensional, and $\bar{\mathbf{x}} = [\mathbf{x}_1^\dagger, \ldots, \mathbf{x}_K^\dagger]^\dagger$ is $Kn_T \times 1$ dimensional. Although the optimum input distribution is not known, we achieve the following lower bound with Gaussian $\bar{\mathbf{x}}$ [47],

$$
C_{sum}^{lb} = I(\mathbf{r}; \bar{\mathbf{x}}|\hat{\mathbf{H}}) = E\left[\log\left|\mathbf{I} + \mathbf{R}_{\tilde{\mathbf{H}}\bar{\mathbf{x}}+\mathbf{n}}^{-1} \hat{\mathbf{H}}\bar{\mathbf{Q}}\hat{\mathbf{H}}^\dagger\right|\right] \quad (4.84)
$$

where $\mathbf{R}_{\tilde{\mathbf{H}}\bar{\mathbf{x}}+\mathbf{n}}$ is the covariance matrix of the effective noise, $\tilde{\mathbf{H}}\bar{\mathbf{x}}+\mathbf{n}$, and $\bar{\mathbf{Q}} = E[\bar{\mathbf{x}}\bar{\mathbf{x}}^\dagger]$. Since the inputs for different users are independent from each other, $\bar{\mathbf{Q}}$ is a block diagonal matrix, having $\mathbf{Q}_k$ in its diagonals with $\text{tr}(\mathbf{Q}_k) \leq P_{d_k}$. As a result, we have $\bar{\mathbf{H}}\bar{\mathbf{Q}}\bar{\mathbf{H}}^\dagger = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^\dagger$,

$$
C_{sum}^{lb} = E\left[\log\left|\mathbf{I} + \mathbf{R}_{\tilde{\mathbf{H}}\bar{\mathbf{x}}+\mathbf{n}}^{-1} \sum_{k=1}^{K} \hat{\mathbf{H}}_k \mathbf{Q}_k \hat{\mathbf{H}}_k^\dagger\right|\right]. \quad (4.85)
$$

The covariance of the effective noise can be calculated as

$$\mathbf{R}_{\tilde{\mathbf{H}}\bar{\mathbf{x}}+\mathbf{n}} = \mathbf{I} + E\left[\tilde{\mathbf{H}}\bar{\mathbf{x}}\bar{\mathbf{x}}^{\dagger}\tilde{\mathbf{H}}^{\dagger}\right] = \mathbf{I} + \sum_{k=1}^{K} E\left[\tilde{\mathbf{H}}_k \mathbf{Q}_k \tilde{\mathbf{H}}_k^{\dagger}\right]. \tag{4.86}$$

From (4.31), we know that $E\left[\tilde{\mathbf{H}}_k \mathbf{Q}_k \tilde{\mathbf{H}}_k^{\dagger}\right] = \mathrm{tr}(\mathbf{Q}_k \tilde{\mathbf{\Sigma}}_k)\mathbf{I}$. Using this, the achievable rate in (4.85) can be written as

$$C_{sum}^{lb} = E\left[\log\left|\mathbf{I} + \frac{\sum_{k=1}^{K} \hat{\mathbf{H}}_k \mathbf{Q}_k \hat{\mathbf{H}}_k^{\dagger}}{1 + \sum_{k=1}^{K} \mathrm{tr}(\mathbf{Q}_k \tilde{\mathbf{\Sigma}}_k)}\right|\right] \tag{4.87}$$

Since our goal is to find the largest lower bound, i.e., the largest achievable sum-rate with Gaussian signaling, the sum-rate maximization problem over the entire block becomes

$$R_{sum} = \max_{\substack{(\mathbf{Q}_k, P_{t_k}, T_{t_k}) \in \mathcal{S}_k \\ \mathrm{tr}(\mathbf{Q}_k) \le P_{d_k}, \forall k}} \frac{T - T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\sum_{k=1}^{K} \hat{\mathbf{H}}_k \mathbf{Q}_k \hat{\mathbf{H}}_k^{\dagger}}{1 + \sum_{k=1}^{K} \mathrm{tr}(\mathbf{Q}_k \tilde{\mathbf{\Sigma}}_k)}\right|\right] \tag{4.88}$$

where $\mathcal{S}_k = \left\{(\mathbf{Q}_k, P_{t_k}, T_{t_k}) \middle| \mathrm{tr}(\mathbf{Q}_k)T_d + P_{t_k}T_{t_k} = P_k T\right\}$, and the coefficient $\frac{T-T_t}{T}$ reflects the amount of time that is spend during the training phase. Note that the maximization is over the parameters of all users, where user k has the training parameters P_{t_k}, and T_{t_k}, and the data transmission parameter $\mathbf{Q}_k$, 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 all users during the data transmission phase, which are given by the eigenvectors of $\mathbf{Q}_k$. We will then focus on the joint optimization of the sum-rate over the eigenvalues of the transmit covariance matrix of all users, the transmit powers of the training phase of all users, and the duration of the training phase.

Transmit Directions

When the CSI at the receiver is perfect, we showed in Chapter 3 that the eigenvectors of the transmit covariance matrix of each user must be equal to the eigenvectors of the channel covariance matrix of that user, i.e., $\mathbf{U}_{Q_k} = \mathbf{U}_{\Sigma_k}$. In other words, single-user transmit directions strategy is optimum in a multi-user case as well. In the following theorem, we show that this is also true when there is channel estimation error at the receiver.

Theorem 12 *Let $\mathbf{\Sigma}_k = \mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{\Sigma_k}\mathbf{U}_{\Sigma_k}^\dagger$ be the spectral decomposition of the covariance matrix of the channel of user k. Then, the optimum transmit covariance matrix $\mathbf{Q}_k$ of user k has the form $\mathbf{Q}_k = \mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{Q_k}\mathbf{U}_{\Sigma_k}^\dagger$.*

Proof: In (4.73), and (4.80), we have shown that, when $\mathbf{\Sigma}_k = \mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{\Sigma_k}\mathbf{U}_{\Sigma_k}^\dagger$, we have $\hat{\mathbf{\Sigma}}_k = \mathbf{U}_{\Sigma_k}\hat{\mathbf{\Lambda}}_{\Sigma_k}\mathbf{U}_{\Sigma_k}^\dagger$, and $\tilde{\mathbf{\Sigma}}_k = \mathbf{U}_{\Sigma_k}\tilde{\mathbf{\Lambda}}_{\Sigma_k}\mathbf{U}_{\Sigma_k}^\dagger$. By using (4.2), we have $\hat{\mathbf{H}}_k = \hat{\mathbf{Z}}_k\mathbf{U}_{\Sigma_k}\hat{\mathbf{\Lambda}}_{\Sigma_k}^{1/2}\mathbf{U}_{\Sigma_k}^\dagger$. Inserting these into (4.34), we obtain

$$R_{sum} = \max_{\substack{(\mathbf{Q}_k, P_{t_k}, T_{t_k}) \in \mathcal{S}_k \\ \mathrm{tr}(\mathbf{Q}_k) \leq P_{d_k}, \forall k}} \frac{T - T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\sum_{k=1}^K \hat{\mathbf{Z}}_k\hat{\mathbf{\Lambda}}_{\Sigma_k}^{1/2}\mathbf{U}_{\Sigma_k}^\dagger\mathbf{Q}_k\mathbf{U}_{\Sigma_k}\hat{\mathbf{\Lambda}}_{\Sigma_k}^{1/2}\hat{\mathbf{Z}}_k^\dagger}{1 + \sum_{k=1}^K \mathrm{tr}\left(\mathbf{U}_{\Sigma_k}^\dagger\mathbf{Q}_k\mathbf{U}_{\Sigma_k}\tilde{\mathbf{\Lambda}}_{\Sigma_k}\right)}\right|\right] \qquad (4.89)$$

where we used the fact that the random matrices $\left\{\hat{\mathbf{Z}}_k\right\}_{k=1}^{K}$ and $\left\{\hat{\mathbf{Z}}_k\mathbf{U}_{\Sigma_k}\right\}_{k=1}^{K}$ have the same joint distribution for zero-mean identity covariance Gaussian $\hat{\mathbf{Z}}_k$ and unitary $\mathbf{U}_{\Sigma_k}$. We may spectrally decompose the expression sandwiched between $\hat{\mathbf{Z}}_k$ and its conjugate transpose in (4.89) as

$$\hat{\mathbf{\Lambda}}_{\Sigma_k}^{1/2}\mathbf{U}_{\Sigma_k}^{\dagger}\mathbf{Q}_k\mathbf{U}_{\Sigma_k}\hat{\mathbf{\Lambda}}_{\Sigma_k}^{1/2} = \mathbf{U}_k\mathbf{\Lambda}_k\mathbf{U}_k^{\dagger} \tag{4.90}$$

Using (4.36), we have $\mathrm{tr}\left(\mathbf{U}_{\Sigma_k}^{\dagger}\mathbf{Q}_k\mathbf{U}_{\Sigma_k}\tilde{\mathbf{\Lambda}}_{\Sigma_k}\right) = \mathrm{tr}\left(\mathbf{U}_k^{\dagger}\hat{\mathbf{\Lambda}}_{\Sigma_k}^{-1}\tilde{\mathbf{\Lambda}}_{\Sigma_k}\mathbf{U}_k\mathbf{\Lambda}_k\right)$. The optimization problem becomes,

$$R_{sum} = \max_{\substack{(\mathbf{Q}_k, P_{t_k}, T_{t_k}) \in \mathcal{S}_k \\ \mathrm{tr}(\mathbf{Q}_k) \le P_{d_k}, \forall k}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\sum_{k=1}^{K}\hat{\mathbf{Z}}_k\mathbf{\Lambda}_k\hat{\mathbf{Z}}_k^{\dagger}}{1+\sum_{k=1}^{K}\mathrm{tr}\left(\mathbf{U}_k^{\dagger}\hat{\mathbf{\Lambda}}_{\Sigma_k}^{-1}\tilde{\mathbf{\Lambda}}_{\Sigma_k}\mathbf{U}_k\mathbf{\Lambda}_k\right)}\right|\right] \tag{4.91}$$

where we again used the fact that the random matrices $\left\{\hat{\mathbf{Z}}_k\right\}_{k=1}^{K}$ and $\left\{\hat{\mathbf{Z}}_k\mathbf{U}_{\Sigma_k}\right\}_{k=1}^{K}$ have the same joint distribution. Note that in the optimization in (4.91), the numerator of the objective function does not involve $\mathbf{U}_k$. For the denominator, using [26, Theorem 9.H.1.h, page 249], we know $\mathrm{tr}(\hat{\mathbf{\Lambda}}_{\Sigma_k}^{-1}\tilde{\mathbf{\Lambda}}_{\Sigma_k}\mathbf{\Lambda}_k) \le \mathrm{tr}(\mathbf{U}_k^{\dagger}\hat{\mathbf{\Lambda}}_{\Sigma_k}^{-1}\tilde{\mathbf{\Lambda}}_{\Sigma_k}\mathbf{U}_k\mathbf{\Lambda}_k)$, for all unitary $\mathbf{U}_k$. Therefore, we can choose $\mathbf{U}_k = \mathbf{I}$ for all k to minimize the denominator, and hence maximize the sum-rate as long as this choice is feasible. In order to check for the feasibility, we write the trace constraint on $\mathbf{Q}_k$ using (4.90) as

$$\begin{aligned}
\mathrm{tr}(\mathbf{Q}_k) &= \mathrm{tr}(\mathbf{U}_{\Sigma_k}\hat{\mathbf{\Lambda}}_{\Sigma_k}^{-1/2}\mathbf{U}_k\mathbf{\Lambda}_k\mathbf{U}_k^{\dagger}\hat{\mathbf{\Lambda}}_{\Sigma_k}^{-1/2}\mathbf{U}_{\Sigma_k}^{\dagger}) && (4.92)\\
&= \mathrm{tr}(\mathbf{U}_k^{\dagger}\hat{\mathbf{\Lambda}}_{\Sigma_k}^{-1}\mathbf{U}_k\mathbf{\Lambda}_k). && (4.93)
\end{aligned}$$

Again from [26, Theorem 9.H.1.h, page 249], $\text{tr}(\hat{\mathbf{\Lambda}}_{\Sigma_k}^{-1}\mathbf{\Lambda}_k) \leq \text{tr}(\mathbf{U}_k^\dagger \hat{\mathbf{\Lambda}}_{\Sigma_k}^{-1}\mathbf{U}_k\mathbf{\Lambda}_k) \leq P_{d_k}$, for all unitary $\mathbf{U}_k$ and for all k. Therefore, we conclude that $\mathbf{U}_k = \mathbf{I}$ choice is feasible for all k. Then, using $\mathbf{U}_k = \mathbf{I}$, from (4.90), we have the desired result:

$$\mathbf{Q}_k = \mathbf{U}_{\Sigma_k}\hat{\mathbf{\Lambda}}_{\Sigma_k}^{-1}\mathbf{\Lambda}_k\mathbf{U}_{\Sigma_k} \tag{4.94}$$

with $\mathbf{\Lambda}_{Q_k} = \hat{\mathbf{\Lambda}}_{\Sigma_k}^{-1}\mathbf{\Lambda}_k$. $\square$

Using Theorem 12, we can write the optimization problem in (4.88) as,

$$R_{sum} = \max_{\substack{(\boldsymbol{\lambda}_k^Q, P_{t_k}, T_{t_k}) \in \mathcal{P}_k \\ k=1,\ldots,K}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\sum_{k=1}^K \sum_{i=1}^{n_T} \lambda_{ki}^Q \hat{\lambda}_{ki}^\Sigma \hat{\mathbf{z}}_{ki}\hat{\mathbf{z}}_{ki}^\dagger}{1 + \sum_{k=1}^K \sum_{i=1}^{n_T} \lambda_{ki}^Q \tilde{\lambda}_{ki}^\Sigma}\right|\right] \tag{4.95}$$

where $\boldsymbol{\lambda}_k^Q = [\lambda_{k1}^Q, \ldots, \lambda_{kn_T}^Q]$, $\mathcal{P}_k = \left\{\left(\boldsymbol{\lambda}_k^Q, P_{t_k}, T_{t_k}\right) \,\middle|\, \left(\sum_{i=1}^{n_T} \lambda_{ki}^Q\right) T_d + P_{t_k}T_{t_k} = P_k T\right\}$, and $\hat{\mathbf{z}}_{ki}$, which is an $n_R \times 1$ dimensional i.i.d., zero-mean, identity-covariance Gaussian random vector, is the i^{th} column of $\hat{\mathbf{Z}}_{ki}$.

Joint Power Allocation Policy

For a MIMO-MAC system with perfect CSI at the receiver and partial CSI at the transmitters, we propose an algorithm to find the optimum power allocation policy in Chapter 3. However, the algorithm in Chapter 3 is not suitable to find the optimum values of P_{t_k} and T_{t_k}, if directly applied to the model in this chapter. Existence of P_{t_k} and T_{t_k} violates the symmetry in Chapter 3, and changes the form of the objective function. Therefore, in this chapter, we modify the algorithm in Chapter 3 so that the new algorithm finds the solutions for P_{t_k} and T_{t_k} as well as the powers along the transmit directions.

By plugging (4.81) and (4.74) into (4.95), we get

$$R_{sum} = \max_{\substack{(\boldsymbol{\lambda}_k^Q, P_{t_k}, T_{t_k}) \in \mathcal{P}_k \\ k=1,\ldots,K}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\sum_{k=1}^{K}\sum_{i=1}^{J_k} \lambda_{ki}^Q(\lambda_{ki}^\Sigma - \mu_k^S)\hat{\mathbf{z}}_{ki}\hat{\mathbf{z}}_{ki}^\dagger}{1+\sum_{k=1}^{K}\left(\sum_{i=1}^{J_k}\lambda_{ki}^Q\mu_k^S + \sum_{i=J_k+1}^{n_T}\lambda_{ki}^Q\lambda_{ki}^\Sigma\right)}\right|\right] \tag{4.96}$$

Since λ_{ki}^Q, for $i = J_k+1,\ldots,n_T$ does not contribute to the numerator, we should choose $\lambda_{ki}^Q = 0$, for $i = J_k+1,\ldots,n_T$. We have,

$$R_{sum} = \max_{\substack{(\boldsymbol{\lambda}_k^Q, P_{t_k}, T_{t_k}) \in \mathcal{P}_k \\ k=1,\ldots,K}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\sum_{k=1}^{K}\sum_{i=1}^{J_k} \lambda_{ki}^Q(\lambda_{ki}^\Sigma - \mu_k^S)\hat{\mathbf{z}}_{ki}\hat{\mathbf{z}}_{ki}^\dagger}{1+\sum_{k=1}^{K}\mu_k^S P_{d_k}}\right|\right]. \tag{4.97}$$

In (4.97), the parameters of the optimization problem are the powers of all users $\lambda_{k1}^Q,\ldots,\lambda_{kT_{t_k}}^Q$ along the transmit directions, the training signal powers P_{t_k} of all users, and the training durations T_{t_k} of all users. Solving for all these variables simultaneously seems intractable. Therefore, we propose a Gauss-Seidel type, round-robin algorithm that solves (4.97) iteratively over the users as in Chapter 3. When updating the parameters corresponding to user k, we assume that the parameters of the rest of the users are fixed. For an update of a given user, the optimization problem becomes

$$R_{sum}^k = \max_{(\boldsymbol{\lambda}_k^Q, P_{t_k}, T_{t_k}) \in \mathcal{P}_k} \frac{T-T_t}{T} E\left[\log\left|\mathbf{\Phi} + \frac{\sum_{i=1}^{J_k} \lambda_{ki}^Q(\lambda_{ki}^\Sigma - \mu_k^S)\hat{\mathbf{z}}_{ki}\hat{\mathbf{z}}_{ki}^\dagger}{\phi+\mu_k^S P_{d_k}}\right|\right], \tag{4.98}$$

where $\mathbf{\Phi} = \mathbf{I} + \frac{\sum_{l \neq k}^{K} \sum_{i=1}^{J_l} \lambda_{li}^{Q} (\lambda_{li}^{\Sigma} - \mu_l^S) \hat{\mathbf{z}}_{li} \hat{\mathbf{z}}_{li}^{\dagger}}{1 + \sum_{l=1}^{K} \mu_l^S P_{d_l}}$, and $\phi = 1 + \sum_{l \neq k}^{K} \mu_l^S P_{d_l}$. Note that the optimization problem in (4.98) is now a single-user problem with fixed interference from the other users. Therefore, we can follow arguments similar to those in the single-user case. Since for any pair (P_{t_k}, T_{t_k}) that results in $J_k < T_{t_k}$, we can find another pair $(P_{t_k}, T_{t_k}^{'})$ that results in a higher rate, it is sufficient to search over those (P_{t_k}, T_{t_k}) pairs that results in $J_k = T_{t_k}$. We can now write (4.98) as

$$R_{sum}^k = \max_{(\boldsymbol{\lambda}_k^Q, P_{t_k}, T_{t_k}) \in \mathcal{R}_k} \frac{T - T_t}{T} E \left[\log \left| \mathbf{\Phi} + \frac{\sum_{i=1}^{T_{t_k}} \lambda_{ki}^Q (\lambda_{ki}^{\Sigma} - \mu_k^S) \hat{\mathbf{z}}_{ki} \hat{\mathbf{z}}_{ki}^{\dagger}}{\phi + \mu_k^S P_{d_k}} \right| \right], \tag{4.99}$$

where $\mathcal{R}_k = \left\{ \left(\boldsymbol{\lambda}_k^Q, P_{t_k}, T_{t_k} \right) \middle| \left(\sum_{i=1}^{n_T} \lambda_{ki}^Q \right) T_d + P_{t_k} T_{t_k} = P_k T, P_{t_k} > \sum_{i=1}^{T_{t_k}} \left(\frac{1}{\lambda_{T_{t_k}}^{\Sigma}} - \frac{1}{\lambda_i^{\Sigma}} \right) \right\}$, and the condition $P_{t_k} > \sum_{i=1}^{T_{t_k}} \left(\frac{1}{\lambda_{T_{t_k}}^{\Sigma}} - \frac{1}{\lambda_i^{\Sigma}} \right)$ guarantees that, using the pair (P_{t_k}, T_{t_k}), all T_{t_k} channels are filled, i.e., $J_k = T_{t_k}$. Note that the parameters that we want to optimize (4.99) over are discrete valued T_{t_k}, and continuous valued P_{t_k}, and λ_{ki}^Q, for $i = 1, \ldots, T_{t_k}$. Since, for every value of T_{t_k}, both the coefficient in front of the expectation, and the number of terms in the sum in the numerator of (4.99) are different, the form of the objective function is also different. Since T_{t_k} is discrete, and $1 \leq T_{t_k} \leq n_T$, we can perform an exhaustive search over T_{t_k} and solve n_T reduced optimization problems with fixed T_{t_k} in each one. Then, we take the solution that results in the maximum rate, i.e.,

$$R_{sum}^k = \max_{1 \leq T_{t_k} \leq n_T} \max_{(\boldsymbol{\lambda}^Q, P_t) \in \mathcal{R}_{k T_{t_k}}} \frac{T - T_t}{T} E \left[\log \left| \mathbf{\Phi} + \frac{\sum_{i=1}^{T_{t_k}} \lambda_{ki}^Q (\lambda_{ki}^{\Sigma} - \mu_k^S) \hat{\mathbf{z}}_{ki} \hat{\mathbf{z}}_{ki}^{\dagger}}{\phi + \mu_k^S P_{d_k}} \right| \right]. \tag{4.100}$$

where $\mathcal{R}_{kT_{t_k}} = \left\{ \left(\boldsymbol{\lambda}_k^Q, P_{t_k} \right) \middle| \left(\sum_{i=1}^{n_T} \lambda_{ki}^Q \right) T_d + P_{t_k} T_{t_k} = P_k T, P_{t_k} > \sum_{i=1}^{T_{t_k}} \left(\frac{1}{\lambda_{T_{t_k}}^\Sigma} - \frac{1}{\lambda_i^\Sigma} \right) \right\}$. While solving for the inner maximization problem, we define $f_{ki}(P_{t_k}) = \frac{\lambda_{ki}^\Sigma - \mu_k^S}{\phi + \mu_k^S P_{d_k}}$, for $i = 1, \ldots, T_{t_k}$. In this case, the inner optimization problem becomes

$$R_{sum}^{k,T_{t_k}} = \max_{(\boldsymbol{\lambda}^Q, P_t) \in \mathcal{R}_{kT_{t_k}}} \frac{T - T_t}{T} E \left[\log \left| \boldsymbol{\Phi} + \sum_{i=1}^{T_{t_k}} \lambda_{ki}^Q f_{ki}(P_{t_k}) \hat{\mathbf{z}}_{ki} \hat{\mathbf{z}}_{ki}^\dagger \right| \right] \tag{4.101}$$

In the optimization problem in (4.101), we have $T_{t_k} + 1$ optimization variables, $\lambda_{k1}^Q, \ldots, \lambda_{T_{t_k}}^Q$, and P_{t_k}. The Lagrangian for the optimization problem in (4.101) can be written as

$$\frac{T - T_t}{T} E \left[\log \left| \boldsymbol{\Phi} + \sum_{i=1}^{T_{t_k}} \lambda_{ki}^Q f_{ki}(P_{t_k}) \hat{\mathbf{z}}_{ki} \hat{\mathbf{z}}_{ki}^\dagger \right| \right] - \mu_k \left(\left(\sum_{i=1}^{T_{t_k}} \lambda_{ki}^Q \right) T_d + P_{t_k} T_{t_k} - P_k T \right). \tag{4.102}$$

where μ_k is the Lagrange multiplier, and we omitted the complementary slackness conditions related to the positiveness of λ_{ki}^Q, and $P_{t_k} - \sum_{i=1}^{T_{t_k}} \left(\frac{1}{\lambda_{T_{t_k}}^\Sigma} - \frac{1}{\lambda_i^\Sigma} \right)$. Using (2.25), the KKT conditions can be written as

$$\frac{T_d}{T} f_{ki}(P_{t_k}) E \left[\mathbf{z}_{ki}^\dagger \mathbf{B}^{-1} \mathbf{z}_{ki} \right] \leq \mu_k T_d, \quad i = 1, \ldots, T_{t_k} \tag{4.103}$$

$$\frac{T_d}{T} \sum_{i=1}^{T_{t_k}} \lambda_{ki}^Q E \left[\mathbf{z}_{ki}^\dagger \mathbf{B}^{-1} \mathbf{z}_{ki} \right] \frac{\partial f_{ki}(P_{t_k})}{\partial P_{t_k}} = \mu_k T_{t_k} \tag{4.104}$$

where $\mathbf{B} = \boldsymbol{\Phi} + \sum_{i=1}^{T_{t_k}} \lambda_{ki}^Q f_{ki}(P_{t_k}) \hat{\mathbf{z}}_{ki} \hat{\mathbf{z}}_{ki}^\dagger$, and the equality of the last equation again follows from the complementary slackness condition. Note that when the optimum λ_{ki}^Q is non-zero, the corresponding inequality in (4.103) will be satisfied with equality.

Therefore, we pull the expectation terms from (4.103) for those equations with non-zero λ_{ki}^Q's, and insert them into (4.104). Since those indices with $\lambda_{ki}^Q = 0$ do not contribute to (4.104), we have

$$\sum_{i=1}^{T_{t_k}} \lambda_{ki}^Q \frac{f'_{ki}(P_{t_k})}{f_{ki}(P_{t_k})} = \frac{T_{t_k}}{T_d}. \tag{4.105}$$

Now, we have a fixed-point equation which does not include any expectation terms. We can use this to solve P_{t_k} in terms of λ_{ki}^Q's. Using the single-user results in the previous section, we propose the following algorithm that first solves $P_{t_k}(n+1)$ from

$$\sum_{i=1}^{T_{t_k}} \lambda_{ki}^Q(n) \frac{f'_{ki}(P_{t_k}(n+1))}{f_{ki}(P_{t_k}(n+1))} = \frac{T_{t_k}}{T_d} \tag{4.106}$$

then, updates $\lambda_{ki}^Q(n+1)$ using

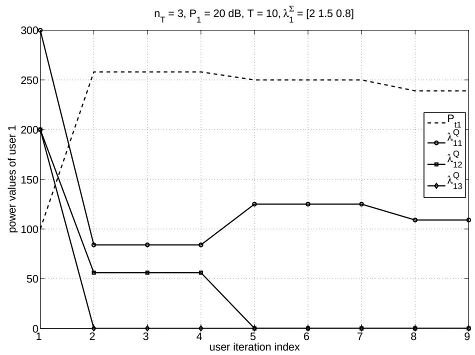
$$\lambda_{ki}^Q(n+1) = \frac{\lambda_{ki}^Q(n) f_{ki}(P_{t_k}(n+1)) E\left[\mathbf{z}_{ki}^\dagger \mathbf{B}^{-1} \mathbf{z}_{ki}\right]}{\sum_{j=1}^{T_{t_k}} \lambda_{kj}^Q(n) f_{kj}(P_{t_k}(n+1)) E\left[\mathbf{z}_{kj}^\dagger \mathbf{B}^{-1} \mathbf{z}_{kj}\right]} \frac{(P_k T - P_{t_k}(n+1) T_{t_k})}{T_d} \tag{4.107}$$

This algorithm finds the solution of the inner optimization problem in (4.101) in terms of the training power P_{t_k}, and the eigenvalues of the transmit covariance matrix $\lambda_{k1}^Q, \ldots, \lambda_{kT_{t_k}}^Q$ of user k, when T_{t_k} and the parameters of the rest of the users are fixed. We run n_T such algorithms simultaneously for user k. The solution of (4.99) can be found by taking the one that results in the largest rate, which gives us the solution for T_{t_k}. Now, we know the parameters $\boldsymbol{\lambda}_k^Q, P_{t_k}, T_{t_k}$, that solve (4.99), when the parameters of the rest of the users are fixed. We then move to another user, and perform the same inner maximization for this user keeping the parameters of the rest of the users fixed. In this manner we iterate over the users in a round-robin fashion. Finally, this iterative algorithm gives us the parameters of all users that solve (4.97).

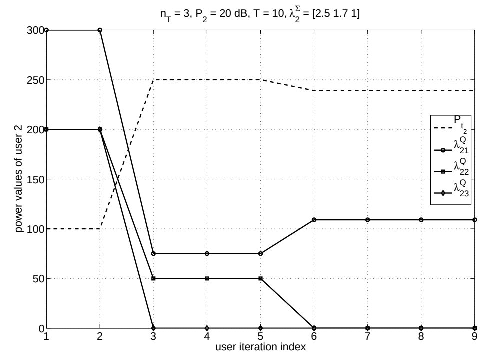
As a result, we solved the joint channel estimation and resource allocation problem in a MIMO multiple access channel with noisy channel estimation and partial CSI available at the transmitter. For user k, through the solution for P_{t_k}, we find the solution for the allocation of available power of user k over the training and data transmission phases. Through T_{t_k}, we find the portion of the training duration that is allocated to user k, and through the sum of these portions $T_t = \sum_{k=1}^{K} T_{t_k}$, we find the solution for the allocation of available time over the training and data transmission phases. Through Theorem 12, we find the optimum transmit directions of user k, and through $\lambda_{k1}^{Q}, \ldots, \lambda_{T_{t_k}}^{Q}$, we find the solution for the allocation of data transmission power of user k over these transmit directions. Finally, the optimum training signal of user k, $\mathbf{S}_k$, is determined by T_{t_k} and P_{t_k} through Theorem 11.

4.3.3 Numerical Results for MIMO-MAC

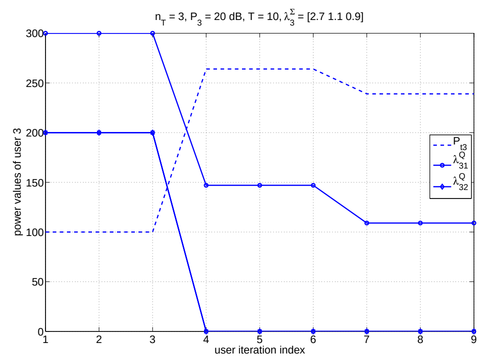
In a MIMO-MAC case, proving the convergence of our algorithm becomes even harder. However, again, we observed through extensive simulations that the proposed algorithm always converges. In Figure 4.8, we considered a system of $K = 3$ users with $n_T = n_R = 3$, all having moderate power, $P = 20$ dB, and moderate block



(a)



(b)

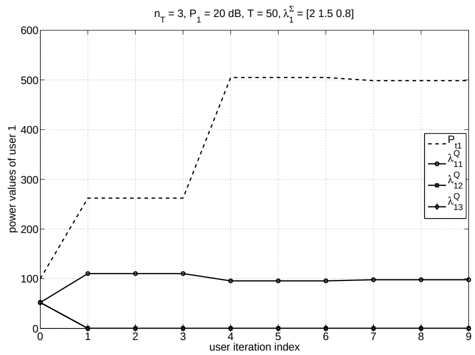


(c)

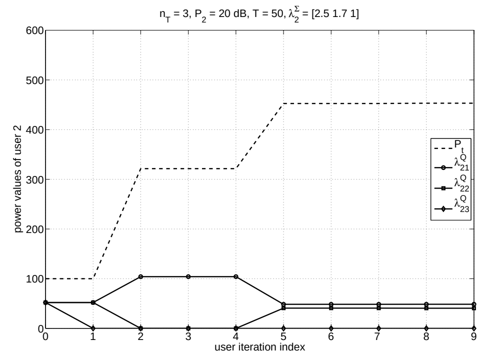
Figure 4.8: The convergence of the multi-user algorithm with $n_T = n_R = 3$, 20 dB total average power and $T = 10$: (a) convergence of user 1; (b) convergence of user 2; (c) convergence of user 3.

length, $T = 10$. Each iteration in Figure 4.8, corresponds to solving (4.99) for one of the users, while the parameters of the rest of the users are fixed. Although we observe in Figure 4.7 that in the same system with a single user, estimating two dimensions of the channel gives the highest rate, in this multi-user case, we observe in Figure 4.8 that, all users estimate only one dimension of the channel.

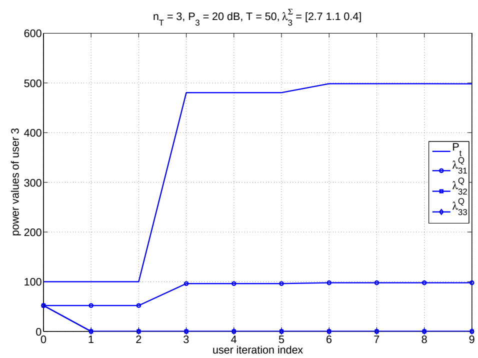
We observed through extensive simulations that for a large set of channel eigenvalues, total available power and the block length, all users estimate only one dimension of the channel. In order to estimate a second dimension, either very large levels of power or a long enough coherence time is needed. For example, we see in Figure



(a)



(b)



(c)

Figure 4.9: The convergence of the multi-user algorithm with $n_T = n_R = 3$, 20 dB total average power and $T = 50$: (a) convergence of user 1; (b) convergence of user 2; (c) convergence of user 3.

4.9 that, for a 3-user system, one of the users start estimating the second dimension, when T gets large enough, i.e., when $T = 50$. However, when the number of users increases, total number of channels estimated by all users also increases, since each user has to spend its power.

4.4 Conclusions

In this chapter, we considered both the training and data transmission phases of a transmission block, for single-user MIMO and MIMO-MAC scenarios in a block-fading channel where the receiver has a noisy estimate of the channel and the transmitters have partial CSI in the form of covariance feedback. We analyzed the joint optimization of the channel estimation and the data transmission parameters in both single-user and multi-user cases. In the single-user case, we formulated the joint optimization problem over the eigenvalues of the transmit covariance matrix and the estimation process parameters. We solved this problem by introducing a number of reduced optimization problems, each of which can be solved efficiently using the proposed algorithm. Through simulations, we observed that the algorithm converges and it converges to the same point regardless of the initial point of the iterations. In the multi-user case, we considered to optimize an achievable sum-rate jointly over the training and data transmission parameters. The proposed multi-user algorithm solves the problem iteratively over the users, while utilizing the single-user algorithm for an update of each user. The theoretical convergence proofs of these algorithms remain as open problems.

The results in this chapter are submitted for publication in [33], [34], [39], [41].

Chapter 5

Conclusions

The information theoretic promise of high data rates when utilizing multiple antennas in a wireless communications link motivated significant amount of research on the design of optimum transmit strategies that can achieve those rates. However, achievable rates depend crucially on how well the channel state is estimated at the receiver and how much of the channel state is available at the transmitters. For most practical systems, the assumption of having perfect channel knowledge at both the receiver and the transmitter is unrealistic.

In this thesis we have addressed the effects of having partial CSI at the transmitter side, and noisy channel estimation at the receiver side on the optimum transmit strategies that maximize the achievable data rates in wireless MIMO communications. The analysis in this thesis combines methods from information theory, optimization theory, estimation theory, parallel and distributed algorithms, matrix analysis, probability, and statistics. The main contributions of this thesis can be summarized as follows.

Transmit Directions and Optimality of Beamforming

The use of multiple antennas at both the transmitters and the receivers in wireless communications promises very large information rates when the perfect knowledge of the channel is known at the receivers and the transmitters. However, in most of the wireless communication scenarios, especially in wireless MIMO communications, it is unrealistic to assume that the transmitter side has the perfect knowledge of the instantaneous CSI. In such scenarios, it might be more realistic to assume that only the receiver side can perfectly estimate the instantaneous CSI, while the transmitter side has only a statistical knowledge of the channel. When the fading in the channel is assumed to be a Gaussian process, statistics of the channel reduce to mean and covariance information of the channel. Since the capacity achieving input signaling is Gaussian, the capacity maximization problem reduces to finding the optimum transmit covariance matrices, i.e., the optimum transmit directions and the optimum power allocation policies. In this case the transmit directions and the power allocations are not functions of the channel states, but they are functions of the statistics of the channel states, that are fed by the receiver side back to the transmitter side.

This thesis provides a thorough analysis of the effects of partial and no CSI on the capacity of a single-user and the sum-capacity of a multi-user MIMO channel. The results show that even in a multi-user scenario, each user should maintain its single-user transmit directions, which means that multi-user interference does not affect the directions that the signals are transmitted. Furthermore, when the number of users in the system increases, beamforming becomes optimal for a greater range of channel parameters, and finally becomes unconditionally optimal for asymptotically large systems. Consequently, in large multi-user systems, the optimum transmit strategies of the users get simplified, and therefore the overall complexity of large multi-user systems may ramain in reasonable limits.

Optimum Power Allocation Policies

In a MIMO system, a transmit strategy is a combination of a transmit direction strategy and a transmit power allocation strategy. We have shown in this thesis that the optimal multi-user transmit direction strategies are decoupled into a set of single-user transmit direction strategies. However, in general, this is not true for the optimal transmit power allocation strategies. The amount of power each user allocates in each direction depends on both the transmit directions and the power allocations of other users. Optimum power allocation policy, in effect, determines the number of spatial dimensions that is required to achieve the capacity, through the number of components with non-zero power allocation. If this number is one, beamforming is optimal. If this number is greater than one, then either vector coding or parallel processing of scalar codes is needed.

Although having beamforming optimality conditions is extremely helpful, in a single-user MIMO or in a MIMO-MAC with finite number of users, the channel statistics might be such that beamforming may never be optimal. For such scenarios, we proposed efficient and globally convergent algorithms in order to solve for the optimum eigenvalues of the transmit covariance matrices in both single-user and multi-user scenarios. These algorithms find the optimum eigenvalues of the transmit covariance matrices, by enforcing the KKT optimality conditions at each iteration. We proved that the convergence points of these algorithms are unique and equal to the optimum eigenvalue allocations. The proposed algorithms converge to this unique point starting from any point in the space of feasible eigenvalues. With these algorithms and the convergence results, the optimization of the transmit strategies for single-user MIMO and MIMO-MAC systems with partial CSI at the transmitters, is solved.

Noisy Channel Estimation at the Receiver

Although one can obtain very high rates with perfect CSI at the receiver, when the channel knowledge is not perfect, achievable rates decrease significantly. This decrease is especially pronounced for MIMO channels. Moreover, measuring the CSI and feeding it back to the transmitter uses communication resources, which could otherwise be used for useful information transmission. One way of measuring the CSI is that the transmitters send known training sequences, from which the receivers measure the channels. The receivers, then, extract the information (according to the feedback model) to be fed back from the estimated channel, and feed it back to the transmitters. This overall process of estimating and feeding back CSI uses up time, bandwidth and power. In order to take this loss of resources into account, we have considered a block fading channel, where the block length and the available power is divided between the training and data transmission phases.

When the CSI at the receiver is noisy, the capacity and the corresponding optimum signalling scheme are not known. Therefore, for both the single-user and multi-user scenarios, we first developed a lower bound to the capacity that can be achieved with Gaussian signaling. We optimized the achievable rate of the data transmission phase jointly over the parameters of the training and data transmission phases. We first found the optimum training signal that minimizes the mean square error of the channel estimation process. Then, we developed an algorithm to maximize the achievable rate. This algorithm finds the solution for the partition of the given total transmitter power and the block length between the training and the data transmission phases, and also the solution for the allocation of the data transmission power over the antennas.

BIBLIOGRAPHY

[1] I. C. Abou-Faycal, M. D. Trott, and S. Shamai. The capacity of discrete-time memoryless rayleigh-fading channels. *IEEE Transactions on Information Theory*, 47(4):1290–1301, May 2001.

[2] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.

[3] 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.

[4] 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.

[5] G. J. Foschini and M. J Gans. On limits of wireless communication in a fading environment when using multiple antennas. *Wireless Personal Communications*, 6:311–335, 1998.

[6] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath. Capacity limits of MIMO channels. *IEEE Journal on Selected Areas in Communications*, 21(5), June 2003.

[7] 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.

[8] S. V. Hanly and D. N. C. Tse. Resource pooling and effective bandwidths in CDMA networks with multiuser receivers and spatial diversity. *IEEE Transactions on Information Theory*, 47(4):1328–1351, May 2001.

[9] 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.

[10] B. M. Hochwald, T. L. Marzetta, and V. Tarokh. Multiple-antenna channel hardening and its implications for rate feedback and scheduling. *IEEE Transactions on Information Theory*, 50(9):1893–1909, September 2004.

[11] D. Hoesli, Y. Kim, and A. Lapidoth. Monotonicity results for coherent MIMO Ricean channels. *IEEE Transactions on Information Theory*, 51(12):4334–4339, December 2005.

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

[13] D. Hösli and A. Lapidoth. The capacity of a MIMO Ricean channel is monotonic in the singular values of the mean. In 5^{th} *International ITG Conference on Source and Channel Coding*, January 2004.

[14] 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.

[15] W.C. Jakes, editor. *Microwave Mobile Communications*. Wiley, 1974.

[16] N. Jindal. MIMO broadcast channels with finite-rate feedback. *IEEE Transactions on Information Theory*, 52(11):5045–5060, November 2006.

[17] E. Jorswieck and H. Boche. Optimal transmission with imperfect channel state information at the transmit antenna array. *Wireless Personal Communications*, 27(1):33–56, 2003.

[18] E. Jorswieck and H. Boche. Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback. *IEEE Transactions on Wireless Communications*, 3(5):1543–1553, September 2004.

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

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

[21] R. Knopp and P. A. Humblet. Information capacity and power control in single-cell multiuser communications. In *IEEE International Conference on Communications*, June 1995.

[22] J. H. Kotecha and A. M. Sayeed. Transmit signal design for optimal estimation of correlated MIMO channels. *IEEE Transactions on Information Theory*, 49(10):2562–2579, October 2003.

[23] A. Lozano, F. R. Farrokhi, and R. A Valenzuela. Asymptotically optimal open-loop space-time architecture adaptive to scattering conditions. In *Proceedings of Vehicular Technology Conference*, Spring 2001.

[24] S. C. Malik and S. Arora. *Mathematical Analysis*. Wiley, 1992.

[25] A. Mantravadi, V. V. Veeravalli, and H. Viswanathan. Spectral efficiency of MIMO multiaccess systems with single-user decoding. *IEEE Journal on Selected Areas in Communication*, 21(3):382–394, April 2003.

[26] A. W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and Its Applications*. New York:Academic, 1979.

[27] T. L. Marzetta and B. M. Hochwald. Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading. *IEEE Transactions on Information Theory*, 45(1):139–157, January 1999.

[28] 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.

[29] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang. On beamforming with finite rate feedback in multiple-antenna systems. *IEEE Transactions on Information Theory*, 49(10):2562–2579, October 2003.

[30] W. Rhee and J. M. Cioffi. On the capacity of multiuser wireless channels with multiple antennas. *IEEE Transactions on Information Theory*, 49(10):2580–2595, October 2003.

[31] W. Rhee, W. Yu, and J. M. Cioffi. The optimality of beamforming in uplink multiuser wireless systems. *IEEE Transactions on Wireless Communications*, 3(1):86–96, January 2004.

[32] R. J. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley, 1980.

[33] 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.

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

[35] A. Soysal and S. Ulukus. Asymptotic optimality of beamforming in multi-user MIMO-MAC with no or partial CSI at the transmitters. In *Vehicular Technology Conference*, Stockholm, Sweden, May 2005.

[36] A. Soysal and S. Ulukus. Transmit directions and optimality of beamforming in MIMO-MAC with partial CSI at the transmitters. In *Conference on Information Sciences and Systems*, Baltimore, MD, March 2005.

[37] A. Soysal and S. Ulukus. Optimum power allocation in fading MIMO multiple access channels with partial CSI at the transmitters. In *Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, October 2006.

[38] 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.

[39] A. Soysal and S. Ulukus. MIMO multiple access channels with noisy channel estimation and partial CSI feedback. In *IEEE Globecom*, 2008. Submitted.

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

[41] A. Soysal and S. Ulukus. Optimizing the rate of a correlated MIMO link jointly over channel estimation and data transmission parameters. In *Conference on Information Sciences and Systems*, 2008.

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

[43] D. N. C. Tse and S. V. Hanly. Multiaccess fading channels-part I: Polymatroid structure, optimal resource allocation and throughput capacities. *IEEE Transactions on Information Theory*, 44(7):2796–2815, November 1998.

[44] A. M. Tulino, A. Lozano, and S. Verdú. Power allocation in multi-antenna communication with statistical channel information at the transmitter. In *IEEE Int. Symp. on PIMRC*, Barcelona, Spain, September 2004.

[45] A. M. Tulino, A. Lozano, and S. Verdú. Capacity-achieving input covariance for single-user multi-antenna channels. *IEEE Transactions on Wireless Communications*, 5(3):662–671, March 2006.

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

[47] 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.

[48] W. Yu and W. Rhee. Degrees of freedom in wireless multiuser spatial multiplex systems with multiple antennas. *IEEE Transactions on Communications*, 54(10):1747–1753, October 2006.

[49] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi. Iterative water-filling for Gaussian vector multiple access channels. *IEEE Transactions on Information Theory*, 50(1):145–151, January 2004.

[50] W. Yu, W. Rhee, and J.M. Cioffi. Optimal power control in multiple access fading channels with multiple antennas. In *IEEE International Conference on Communications*, pages 575–579, 2001.

[51] L. Zheng and D. N. C. Tse. Packing spheres in the Grassmann manifold: A geometric approach to the noncoherent multi-antenna channel. *IEEE Transactions on Information Theory*, 48(2):359–383, February 2002.