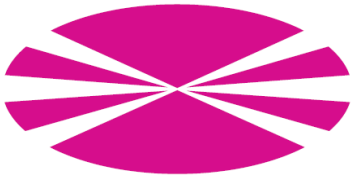
Quality of Service Optimization in the Broadcast Channel with Imperfect Transmit Channel State Information

José Pablo González Coma

Department of Electronics and Systems
University of A Coruña, Spain

Department of Electronics and Systems
University of A Coruña, Spain



PhD Thesis

Quality of Service Optimization in the Broadcast Channel with Imperfect Transmit Channel State Information

José Pablo González Coma

27 de marzo de 2015

PhD Advisors:
Luis Castedo Ribas
Paula M. Castro Castro
Michael Joham

D. Luis Castedo Ribas, Dña. Paula M. Castro Castro y D. Michael Joham

CERTIFICAN:

Que la memoria titulada "Quality of Service Optimization in the Broadcast Channel with Imperfect Transmit Channel State Information" ha sido realizada por D. José Pablo González Coma bajo nuestra dirección en el Departamento de Electrónica y Sistemas de la Universidade da Coruña y concluye la Tesis que presenta para optar al grado de Doctor con Mención Internacional.

A Coruña, a 18 de marzo de 2015

Fdo: Dr. Luis Castedo Ribas
Director de la Tesis Doctoral
Catedrático de Universidad
Dpto. de Electrónica y Sistemas
Universidade da Coruña

Fdo: Dra. Paula M. Castro Castro
Directora de la Tesis Doctoral
Contratada Doctora
Dpto. de Electrónica y Sistemas
Universidade da Coruña

Fdo: Dr. Michael Joham
Director de la Tesis Doctoral
Senior Researcher
Associate Institute for Signal Processing
Technische Universität München

Tesis Doctoral:	Quality of Service Optimization in the Broadcast Channel with Imperfect Transmit Channel State Information
Autor:	D. José Pablo González Coma
Directores:	Dr. Luis Castedo Ribas, Dra. Paula M. Castro Castro y Dr. Michael Joham
Fecha:	23 de marzo de 2015

Tribunal

Presidente:

Vocal 1:

Vocal 2:

Vocal 3:

Secretario:

A mi familia

Agradecimientos

Quiero expresar mi agradecimiento a todos los miembros del GTEC con los que tantos momentos he compartido. En especial me gustaría destacar el apoyo recibido y los buenos ratos pasados durante aquellas comidas con Ángel, Pedro, Tiago y Néstor. También a Óscar, con sus espirales, y a Ismael por esas merendas a media mañá. Quisiera agradecer también los buenos ratos transcurridos durante las reuniones para practicar inglés con todos ellos y con Paula Fraga, Xose y José Antonio.

No puedo dejar de mencionar la ayuda recibida de los restantes profesores de Tecnoloxía Electrónica, Lamas, Julio, Fran, Barreiro y Miguel, y de la asignatura de Xestión de Infraestructuras, Adriana. Tampoco quisiera olvidarme de otros miembros del grupo como Valentín, Javi, Carlos, Dani y Cris.

Mi más sincero agradecimiento a los excelentes investigadores y profesores de la Universidad Técnica de Múnich: Wolfgang Utschick, David Neumann, Thomas Wiese, Alexander Krebs, Andreas Gründinger, Lennart Gerdes, Christoph Hellings, Andreas Dotzler y Maximilian Riemensberger. Y especialmente a Michael Joham, por su enorme colaboración en esta tesis, tanto por sus aportaciones científicas sobre álgebra de matrices, estadísitca, teoría de la información…, como por su genialidad, apoyo y su gran talento como docente desde mi estancia en Alemania en el año 2012.

Quiero agradecer a los miembros del tribunal de evaluación de mi tesis, los profesores doctores Javier Rodríguez Fonollosa, Adriana Dapena Janeiro, Mónica Fernández Bugallo, Wolfgang Utschick, y John Thompson, el trabajo realizado y su excelente disposición en todo momento.

Por último, quiero agradecer a mi directora de tesis, Paula Castro, todo el trabajo que hemos compartido desde el principio, su colaboración en todo lo relativo a la realización de esta tesis y los ánimos aportados en los momentos más duros. Y a mi director de tesis, Luis Castedo, quiero agradecerle la oportunidad que me ha ofrecido de poder trabajar durante todos estos años en este departamento, esas reuniones en las que han surgido ideas

tan valiosas y su calidad como persona. A ambos quiero agradecer todas esas laboriosas correcciones sobre los trabajos a presentar, su paciencia y los conocimientos que me han transmitido.

A mis amigos, con los que tantas alegrías y penas he compartido: Dami, Trassy, Lucía, Isma, Airis, Sabi, Alberto, Jorge, Alba, Miguel, Dinorah . . . Y muchos otros que también debiera mencionar aquí.

A mi familia, a quienes además dedico este trabajo. Dedicárselo a mi familia me permite no dejar a nadie fuera. Mi familia son mis amigos, todas las personas con las que he convivido desde pequeño y, por supuesto, mi familia actual, Andrea. Quiero dedicar mi trabajo a mis padres y mis hermanos, Begoña y Jose, Juan, Javi; mis suegros Isidro y Susi, y mis cuñadas Maira y Marta; algunas de las personas más importantes con las que afortunadamente me ha tocado vivir. A mis abuelos, fuente de inspiración, esfuerzo y apoyo desde los primeros momentos de mi infancia, Isabel, Maximiano, Sara y Juan, quiero agradecerles todo lo que han hecho por mí y siguen haciendo. Y por supuesto, a Andrea. No encuentro palabras que expresen la profundidad de mi agradecimiento. Gracias por compartirlo todo conmigo y estar siempre en los momentos en los que te necesito.

Resumen

Este trabajo considera un sistema *Broadcast Channel* (BC) que consiste en un transmisor equipado con múltiples antenas y varios usuarios con una o más antenas. Dependiendo del número de antenas en el lado receptor, tales sistemas son conocidos como *Multiple-User Multiple-Input Single-Output* (MU-MISO), para usuarios con una única antena, o *Multiple-User Multiple-Input Multiple-Output* (MU-MIMO), para usuarios con varias antenas.

Este modelo es adecuado para sistemas actuales de comunicaciones inalámbricas. Respecto a la dirección del flujo de datos, diferenciamos entre el canal *downlink* o BC, y canal *uplink* o *Multiple Access Channel* (MAC). En el BC las señales se envían desde la estación base a los usuarios, mientras que la información perteneciente a los usuarios es transmitida a la estación base en el MAC.

En este trabajo nos centramos en el BC donde la estación base aplica precodificación lineal aprovechando las múltiples antenas. La información sobre el estado del canal se asume perfecta en todos los usuarios. Sin embargo, los usuarios no cooperan, y la estación base solo tiene información de canal parcial obtenida a través de un canal de realimentación en los sistemas *Frequency-Division Duplex* (FDD), que tiene un ancho de banda limitado. Esta limitación fuerza a los usuarios a aplicar algunos métodos, como quantización, para reducir la cantidad de datos a enviar a la estación base. La combinación de la información proporcionada por los usuarios es interpretada en la estación base como información de canal estocástica, y constituye un factor crítico en el diseño de los precodificadores.

En la literatura se han considerado varios métodos para evaluar el rendimiento del BC, a saber, *Signal to Interference-plus-Noise Ratio* (SINR), *Minimum Mean Square Error* (MMSE), y tasa. Algunos trabajos calculan las medidas correspondientes para cada usuario mientras que otros consideran la suma de todos ellos como la métrica de interés. En nuestro caso, nos centramos en la tasa como figura de mérito. En particular, estamos interesados en garantizar ciertas tasas por usuario. De esta manera, evitamos situaciones injustas que surgen de utilizar la tasa suma como criterio, en las que a los usuarios con canales pobres se les asignan tasas bajas, o incluso cero. Además, reducir la cantidad de potencia necesaria para satisfacer las restricciones de calidad de servicio mencionadas es una característica deseable en los sistemas de comunicaciones inalámbricas. Así, abordamos el problema de optimización consistente en minimizar la potencia total en el transmisor empleada para cumplir un conjunto de restricciones de calidad de servicio, expresadas como tasas por usuario.

Durante los últimos años el problema de minimización de potencia ha sido estudiado ampliamente para información tanto perfecta como imperfecta de canal, en los escenarios BC. Asumir conocimiento de canal perfecto es poco realista y, por tanto, consideramos

que los usuarios envían la información de canal a la estación base por medio de un canal de realimentación, normalmente disponible en los estándares de comunicación recientes. Aunque algunos autores han empleado modelos de incertidumbre limitada para el conocimiento de canal tales como rectangular, elipsoidal, o esférico, y han aprovechado esa asunción para resolver el problema de minimización de potencia, no asumimos una forma particular para esa incertidumbre sino un modelo de error estocástico.

En el modelo de sistema considerado, MU-MIMO, el número de antenas en la estación base es mayor que el número de antenas en cada usuario, e.g. MU-MISO. Además, los usuarios no cooperan para separar las señales recibidas. Debido a ésto y a la falta de grados de libertad en los usuarios, es necesario el uso de filtros transmisores, también llamados precodificadores, para eliminar las interferencias entre usuarios. De este modo, en este trabajo diseñamos conjuntamente los precodificadores lineales y los filtros receptores minimizando la potencia total en el transmisor sujeta a restricciones de tasa por usuario. Esta formulación del problema no es convexa y, por tanto, es complicada de manejar. Por este motivo, aplicamos la desigualdad de Jensen a las restricciones de tasa para obtener otras basadas en el MMSE. Como consecuencia, nuestro objetivo es diseñar los precodificadores y filtros que minimizan el MMSE para todos los usuarios. Para ello, distintos tipos de dualidades basadas en SINR, *Mean Square Error* (MSE), o tasa, han sido empleadas para el diseño de los filtros como fórmulas para intercambiar entre el BC y el MAC por conveniencia. En particular, empleamos la dualidad de MSE con conocimiento de canal imperfecto. Además, para la distribución de potencias, explotamos el marco teórico de las *standard Interference Function*, planteado para resolver el algoritmo de control de potencia. De esta manera, proponemos un algoritmo para solucionar el problema de minimización de potencia en el BC.

Para comprobar la factibilidad de las restricciones de calidad de servicio, proponemos un test que permite determinar si el algoritmo converge o no. Además, el algoritmo propuesto permite resolver el problema dual, ésto es, encontrar los objetivos de tasa balanceados correspondientes a una potencia total en el transmisor. Finalmente, algunas aplicaciones de la minimización de potencia surgen de diferentes escenarios y se resuelven por medio del algoritmo propuesto.

Usando el lenguaje de programación MATLAB se simulan experimentos con el objetivo de mostrar el rendimiento de los métodos propuestos.

Resumo

Este traballo considera un sistema *Broadcast Channel* (BC) que consiste nun transmisor equipado con múltiples antenas e varios usuarios cunha ou máis antenas. Dependendo do número de antenas no lado receptor, tales sistemas son coñecidos como *Multiple-User Multiple-Input Single-Output* (MU-MISO), para usuarios cunha única antena, ou *Multiple-User Multiple-Input Multiple-Output* (MU-MIMO), para usuarios con varias antenas.

Este modelo é adecuado para sistemas actuais de comunicacións sen fíos. Respecto á dirección do fluxo de datos, diferenciamos entre a canle *downlink* ou BC, e a canle *uplink* ou *Multiple Access Channel* (MAC). No BC os sinais envíanse dende a estación base aos usuarios, mentres que a información pertencente aos usuarios é transmitida á estación base no MAC.

Neste traballo centrámonos no BC onde a estación base aplica precodificación lineal aproveitando as múltiples antenas. A información sobre o estado da canle asúmese perfecta en todos os usuarios. Por contra, os usuarios non cooperan e a estación base só ten información da canle parcial obtida a través dunha canle de realimentación nos sistemas *Frequency-Division Duplex* (FDD), que ten un ancho de banda limitado. Esta limitación forza aos usuarios a aplicar algúns métodos, como quantización, para reducir a cantidade de datos que se envían á estación base. A combinación da información proporcionada polos usuarios é interpretada na estación base como información da canle estocástica, e constitúe un factor crítico no deseño dos precodificadores.

Na literatura consideráronse varios métodos para avaliar o rendemento do BC, a saber, *Signal to Interference-plus-Noise Ratio* (SINR), *Minimum Mean Square Error* (MMSE), e taxa. Algúns traballos calculan as medidas correspondentes para cada usuario mentres que outros consideran a suma de todos eles como a métrica de interese. No noso caso, centrámonos na taxa como figura de mérito. En particular, estamos interesados en garantir certas taxas por usuario. Deste xeito evitamos situación inxustas que xurdan de utilizar a taxa suma como criterio, nas que aos usuarios con canles pobres se lles asignan tasas baixas, ou incluso cero. Ademais, reducir a cantidade de potencia necesaria para satisfacer as restriccións de calidade de servizo mencionadas é unha característica desexable nos sistemas de comunicacións se fíos. Así, acometemos o problema de optimización consistente en minimizar a potencia total no transmisor empregada para cumprir un conxunto de restricións de calidade de servizo, expresadas como taxas por usuario.

Durante os últimos anos o problema de minimización de potencia foi estudado amplamente para información tanto perfecta como imperfecta de canle, nos escenarios BC. Asumir coñecemento perfecto de canle é pouco realista e, por tanto, consideramos que os usuarios envían a información de canle á estación base por medio dunha canle de

realimentación, normalmente dispoñible nos estándares de comunicación recentes. Aínda que algúns autores empregaron modelos de incerteza limitada para o coñecemento de canle tales como rectangular, elipsoidal, ou esférico, e aproveitaron esa asunción para solucionar o problema de minimización de potencia, non asumimos unha forma particular para esa incerteza senón un modelo de error estocástico.

No modelo de sistema considerado, MU-MIMO, o número de antenas na estación base é maior que o número de antenas en cada usuario, e.g. MU-MISO. Ademais, os usuarios non cooperan para separar os sinais recibidos. Debido a isto e á falta de graos de liberdade nos usuarios, é preciso o uso de filtros transmisores, tamén chamados precodificadores, para eliminar as interferencias entre usuarios. Deste xeito, neste traballo deseñamos conxuntamente os precodificadores lineais e os filtros receptores minimizando a potencia total no transmisor suxeita a restriccións de taxa por usuario. Esta formulación do problema non é convexa e, por tanto, é complicada de manexar. Por este motivo, aplicamos a desigualdade de Jensen ás restriccións de taxa para obter outras baseadas no MMSE. Como consecuencia, o noso obxectivo é deseñar os precodificadores e filtros que minimizan o MMSE para todos os usuarios. Para iso, distintos tipos de dualidades baseadas en SINR, *Mean Square Error* (MSE), ou taxa, foron empregadas para o deseño dos filtros coma fórmulas para intercambiar entre o BC e o MAC por conveniencia. En particular, empregamos a dualidade de MSE con coñecemento de canal imperfecto. Ademais, para a distribución de potencias, explotamos o marco teórico das *standard Interference Function*, formulado para resolver o algoritmo de control de potencia. Desta maneira, propomos un algoritmo para resolver o problema de minimización de potencia no BC.

Para comprobar a factibilidade das restriccións de calidade de servizo, propomos un test que permite determinar se o algoritmo converxe ou non. Ademais, o algoritmo proposto permite resolver o problema dual, ísto é, atopar os obxectivos de taxa balanceados correspondentes a unha potencia total no transmisor. Finalmente, algunhas aplicacións da minimización de potencia xorden de diferentes escenarios e resólvense por medio do algoritmo proposto.

Usando a linguaxe de programación MATLAB simúlanse experimentos co obxectivo de mostrar o rendemento dos métodos propostos.

Summary

This work considers a *Broadcast Channel* (BC) system, where the transmitter is equipped with multiple antennas and each user at the receiver side could have one or more antennas. Depending on the number of antennas at the receiver side, such a system is known as *Multiple-User Multiple-Input Single-Output* (MU-MISO), for single antenna users, or *Multiple-User Multiple-Input Multiple-Output* (MU-MIMO), for several antenna users.

This model is suitable for current wireless communication systems. Regarding the direction of the data flow, we differentiate between downlink channel or BC, and uplink channel or *Multiple Access Channel* (MAC). In the BC the signals are sent from the *Base Station* (BS) to the users, whereas the information from the users is sent to the BS in the MAC.

In this work we focus on the BC where the BS applies linear precoding taking advantage of multiple antennas. The *Channel State Information* (CSI) is assumed to be perfectly known at each user. However, the users do not cooperate, and the BS only has partial CSI obtained via a feedback link in *Frequency-Division Duplex* (FDD) systems, which is bandwidth limited. This limitation forces the users to apply some methods, like quantization, to reduce the amount of data to be sent to the BS. The combination of the information provided by the users is interpreted as stochastic CSI at the BS, so that the partial CSI is critical for the design of the precoders.

Several criteria have been considered to evaluate the BC performance in the literature, namely, *Signal to Interference-plus-Noise Ratio* (SINR), *Minimum Mean Square Error* (MMSE), and rate. While some works compute the corresponding metric for each of the users, others consider the sum of all of them as the value of interest. In our case, we concentrate on rate as figure of merit. In particular, we are interested in guarantying certain per-user rates. That way, we avoid unfair situations of the sum rate criterion arising when the channels for some of the users are poor with assigned low, even zero, rates. Moreover, reducing the amount of power required to fulfill the mentioned *Quality-of-Service* (QoS) restrictions is a desirable feature for a wireless communication system. Thus, we address the optimization problem consisting on minimizing the total transmit power employed at the BS to fulfill a set of given QoS constraints, expressed as per-user rates.

The power minimization problem has been widely studied during the last years for both perfect and imperfect CSI at the BS scenarios. The assumption of perfect CSI is rather unrealistic so that, as we mentioned previously, we consider that the users send the channel information to the BS by means of the feedback channel, usually available in recent wireless communication standards. Although some authors have employed

bounded uncertainty models for the CSI such as rectangular, ellipsoidal, or spherical, and have taken advantage of that assumption to solve the power minimization problem, we do not assume a particular shape for that uncertainty, but is modeled as a stochastic error.

In the considered MU-MIMO system model the number of antennas at the BS is larger than the number of antennas at each user, e.g. MU-MISO. Moreover, the users do not cooperate to separate the received signals. Due to that and to the lack of degrees of freedom at the users, it makes necessary the use of transmit filters, also denoted as precoders, to remove inter-user interference. Thus, in this work we jointly design the linear precoders and receive filters minimizing the total transmit power subject to per-user rate constraints. This problem formulation is non-convex. As a consequence, it is difficult to deal with. For such a reason, we apply the Jensen's inequality to the rate constraints to obtain a MMSE based restrictions. Consequently, our aim is to find the precoders and the filters that minimize the MMSE for all the users. To that end, several types of dualities based on SINR, *Mean Square Error* (MSE), or rate have been employed for the design of the filters as conversion formulas that allow to switch between the BC and the MAC for convenience. We employ the MSE BC/MAC duality for imperfect *Channel State Information at the Transmitter* (CSIT). Furthermore, for the power allocation design, we take advantage of the *standard Interference Function* (IF) framework, proposed to solve the power control algorithm. In such a way, an algorithm is proposed to solve the power minimization problem in the BC.

To check the feasibility of the QoS constraints, we propose a test that allows to determine the convergence of the algorithm. Additionally, the proposed algorithm can be employed to solve the dual problem, i.e., find the balanced targets for given total transmit power. Finally, some applications of the power minimization problem arising from different scenarios are studied and solved by means of the proposed algorithm.

Simulation experiments are carried out using the technical programming language MATLAB in order to show the performance of the proposed methods.

Index

List of figures

List of tables

Chapter 1

Introduction

1.1. Problem Overview

The *Broadcast Channel* (BC) is a communication system model in which a centralized transmitter sends information to several decentralized receivers. The BC arises when modeling a large number of practical situations in wireless communications, typically in the downlink of cellular systems. For this reason, the transmitter in a BC is usually referred to as the *Base Station* (BS) and the receivers are referred as users. The BC is the dual of the *Multiple Access Channel* (MAC) where several decentralized transmitters (users) send information to a centralized receiver (BS). The MAC is a model that typically arises when considering the uplink in a cellular system.

Both the BC and the MAC are examples of what in the literature of information theory is known as Multiuser Communication systems. In wireless communications, the BC and the MAC can be classified according to the number of antennas used by the transmitter and the receiver. When all terminals employ a single antenna, the term *Single-Input Single-Output* (SISO) is used to label both the BC and the MAC.

In wireless communications, performance is drastically improved if several antennas are deployed. This is particularly feasible at the BSs which are typically terminals with larger resources in terms of power supply and size. In such case, the BC is labeled *Multiple-Input Single-Output* (MISO) and the dual MAC is labeled *Single-Input Multiple-Output* (SIMO). Finally, if users are also equipped with several antennas, both the BC and the MAC are referred to as *Multiple-Input Multiple-Output* (MIMO). Along this work we assume that data sent to all the users (or received from them) is statistically independent. In addition, users do not cooperate to mitigate the inter-user interference nor share information about the channel.

The performance of a wireless communication system, like the BC considered in this work, is given by its capacity. The channel capacity is the limiting information rate,

expressed in terms of *bits per second* (bps), that can be achieved with arbitrarily small error probability [1].

Most work in the literature of BC assumes perfect *Channel State Information at the Transmitter* (CSIT) and *Channel State Information at the Receiver* (CSIR). In a practical system, receivers estimate the channel response from the incoming signal. Hence, it is reasonable to assume that receivers have a perfect knowledge of their individual *Channel State Information* (CSI). The availability of CSI at the transmitter is a more intriguing issue. In a *Time-Division Duplex* (TDD) system, CSI can be estimated at the transmitter during the uplink transmissions and invoking the reciprocity principle. In a *Frequency-Division Duplex* (FDD) system, CSI can be estimated at the receivers and sent back to the transmitter over a feedback channel. The data rate in the feedback channels is usually limited and CSI must be compressed to ensure tight scheduling constraints are satisfied. Such restriction will be referred to as limited feedback along this work.

Either in TDD or FDD it is rather unrealistic to assume that perfect CSIT is available. For the limited feedback systems considered along this work, the information sent to the BS depends only on channel statistics. Thus, the channel uncertainty will be modeled by a stochastic error whose distribution is known at the transmitter, and the average rates are computed taking the conditional expectation of the rate on the available CSI.

The BC capacity region under partial CSIT knowledge has not been found yet. Therefore, obtaining the filters that minimize the total power fulfilling given average rate restrictions is a challenging problem. Moreover, not only the precoders and the receive filters have to be designed, but also the distribution of the power among the different users appears as a critical issue in our system. Since the power allocation and the filters are coupled, any solution to the proposed problem jointly optimizes both parameters. The feasibility of the average rate constraints is another important consideration since the optimization problem could not have solution. A discussion in terms of feasibility regions is a fundamental starting point to be taken into account by the system designer.

One problem commonly studied in the literature will be addressed using the method proposed in this work. In such optimization problem, the goal is to get some balancing between the per-user rates subject to a power restriction. In other words, the per-user average rates are affected by a common factor which is to be optimized employing all the available power.

Additionally, we study more complex scenarios that can be addressed using the methods proposed for the BC. We consider the system model where the users can transmit more than one stream at the same time. This extension of the original problem means an additional complexity layer and the proposed algorithm has to be adapted accordingly. Moreover, we tackle the power minimization in the *Orthogonal Frequency Division Multiplexing* (OFDM) MIMO BC, resulting into a procedure similar to the one employed in the multiple-stream scenario. Other practical implementation arises from the

joint design of the feedback and the precoders, which is solved by means of the so-called Lloyd's algorithm.

1.2. Previous Work

The capacity of a *Single-User Multiple-Input Multiple-Output* (SU-MIMO) Gaussian channel was obtained in [2]. The multiuser scenario is considered in [3–5], and it is shown that the non-linear signaling technique *Dirty Paper Coding* (DPC) [6] is able to approach the sum capacity of a BC. On the other hand, some authors consider more practical low-complexity BC with linear precoding and minimize the transmit power subject to *Quality-of-Service* (QoS) constraints, as done in this work. Optimization is carried out employing different criteria like *Signal to Interference-plus-Noise Ratio* (SINR) [7], or *Minimum Mean Square Error* (MMSE) [8].

These contributions, however, only consider the ideal case where the CSIT is perfectly known. In the more practical case, where only an estimate of the CSI is available, the BC capacity region has not yet been found.

Regarding CSI, several considerations have been made in the literature. Some authors have employed bounded uncertainty models, such as rectangular [9], ellipsoidal (e.g. [10, 11]), or spherical (e.g. [12, 13]). Establishing such assumptions it is possible to formulate the problem with convex constraints and solve it via a *SemiDefinite Program* (SDP) [14]. Other authors, however, model that uncertainty as a stochastic error, (e.g. [15–21]), as done in this work.

Various metrics can be used to evaluate the BC performance, such as SINR [9–13, 21–29], MMSE [30], sum MMSE [15–17, 31, 32], weighted sum rate [18–20], or MMSE balancing [9, 31]. Moreover, we can distinguish between works focused on minimizing the power subject to some restrictions, consisting on achieving certain per-user values for a given metric or some level via a combination of these values over all the users [9–11, 13, 18, 21, 22, 29, 30]; and works where the goal is to obtain the best performance in terms of any of the previously mentioned metrics for given transmit power [12, 15–17, 19, 20, 22, 23, 25, 28, 29, 31].

Several methods have been proposed in the literature to solve this type of problems. For example, the authors in [18] use an approximation of the SINR where the Jensen's inequality is applied in both the numerator and the denominator. This approach was previously introduced in [7] and it has been extensively employed in many works considering SINR-based metrics with imperfect CSI. However, using such an approximation the gap between the real SINR and the approximated one is hard to evaluate. A different SINR-based problem formulation is presented in [21], where the authors propose two conservative approaches using second-order-cone formulations to

satisfy the QoS constraints with certain outage probability for Gaussian and uniform channel estimation error distributions.

The sum MMSE minimization under these assumptions has been previously considered, and several methods have been also proposed to find the optimum power. In [15] the sum MMSE problem in the BC is transformed into the dual *Multiple Access Channel* (MAC), and efficiently solved using SDP methods. In [31], however, both the sum MMSE and the max weighted MMSE are minimized introducing *Mean Square Error* (MSE) dualities and using *Geometric Programing* (GP) (see [33]) and the algorithm presented in [34], respectively. Both works do not provide relationships with the ergodic rate and only MMSE-based metrics are considered as performance measure. However, the connection between the two metrics is exploited in other works where the weighted sum rate maximization via weighted sum MMSE is presented [19, 20]. The problem is formulated using the additional weighting matrices shown in [35], and solved using heuristic approaches like *Deterministic Annealing* (DA) or *Sample Averaging Approximation* (SAA).

Since maximizing the sum MMSE could lead to unfair situations where some of the users get low (or even zero) rates, which is obviously non-desirable, we focus instead on minimizing the transmit power fulfilling some QoS constraints, e.g. [9, 10, 21], expressed in our case as per-user rate requirements, as we will explain in more detail in the following section.

1.3. Main Contributions

This work focus on the minimization of the transmit power in the MIMO BC subject to per-user rate constraints with imperfect CSI. This is as difficult problem since the design of the linear filters at transmission and reception is not jointly convex [14].

The average rates can be lower bounded by the average MMSE by means of the Jensen's inequality [36]. In that way, we obtain new MMSE-based restrictions that allow us to tackle the optimization problem in a more manageable way. The design of the filters in this new power minimization problem subject to average per-user MMSE constraints can be addressed with an *Alternate Optimization* (AO) process. Thus, the optimal receive filters are computed for fixed transmit filters as the minimum MSE receivers and then, keeping these receive filters fixed, the transmit filters are updated. Such an iterative process will lead us to the joint optimal solution for the filters design. This AO iteration has been frequently used in the literature combined with some sort of duality between the BC and the MAC, so this work is not an exception. In particular, we employ the average MSE duality proposed in [37] for imperfect CSIT and perfect CSIR. In this work the authors find the factors which allow to switch from the average MSE obtained in the BC

to the average MSE reached in the dual MAC, preserving both the per-user average MSEs and the total transmit power.

As mentioned before, the optimal filters are designed using the AO method. However, we still have to decide how the transmit power is distributed among the users. Such decision has a direct impact on the level of intra-user interference and, as a consequence, is critical in the system performance. For the design of the power allocation, we rely on the *standard Interference Function* (IF) framework proposed by Yates in [38]. This framework, successfully employed in previous works to distribute the available power among the users (e.g. [31, 39]), solves the power control algorithm and also provides useful properties like convergence and optimum uniqueness.

The AO of the filters together with the interference functions provide us the mechanisms needed to implement the algorithm that solves the optimization problem. We also show that the algorithm converges if the QoS can be fulfilled. Nevertheless, the algorithm is meaningless if the optimization problem is not feasible. Therefore, we also provide a test for checking the feasibility of the problem restrictions. This test is a generalization of that presented in [40] for the vector BC and perfect CSI for both the transmitter and the receivers, where the polytope (or bounded polyhedron) containing the feasible MMSE targets is described. In the simulation results we show how the proposed method solves the power minimization problem via the algorithmic implementation.

Additionally, we propose an algorithm to solve the balancing problem, that is, minimizing weighted per-user MMSEs under restricted total transmit power. The proposed algorithm takes advantage of the solution elaborated for the power minimization problem and the well-know bisection search. This method divides an interval on the real axis where the optimum yields and then selects a subinterval where the optimum lies for further iterations.

Finally, we studied the application of the proposed method to address some additional situations. The *Multiple-User Multiple-Input Multiple-Output* (MU-MIMO) where each user transmits more than one stream is considered. Due to the additional spatial dimension emerging from the multiple streams per user, the per-user rates have to be distributed among the different streams. In such terms, we develop a method that optimally divides the rates between the streams to find a local minimum of the total transmit power. A different method is obtained for the OFDM MU-MIMO. In this case, the rates can be distributed between the subcarriers that constitute the OFDM signal. However, due to the particular structure of the OFDM channels, the multiple stream approximation does not apply in this scenario. Moreover, a joint design of the CSI and the transmit filters is considered in this work. To that end, the Lloyd's algorithm is employed.

1.4. Publications

The international conference and journal papers presented below exhibit the acceptation of the work proposed by the Ph.D. student in the field over the last years.

- Title: Impact of transmit impairments on multiuser MIMO nonlinear transceivers

  Authors: José P. González-Coma, Paula M. Castro, and Luis Castedo

  Conference: International ITG Workshop on Smart Antennas (WSA)

  Location: Aachen, Germany

  Date: February, 2011

- Title: Transmit Impairments Influence on the Performance of MIMO Receivers and Precoders

  Authors: José P. González-Coma, Paula M. Castro, and Luis Castedo

  Conference: European Wireless (EW)

  Location: Vienna, Austria

  Date: April, 2011

- Title: Performance Evaluation of Non-Linear MIMO Precoders Under Transmit Impairments

  Authors: José P. González-Coma, Paula M. Castro, and Luis Castedo

  Conference: 19th European Signal Processing Conference (EUSIPCO)

  Location: Barcelona, Spain

  Date: August, 2011

- Title: Performance of MIMO Systems in Measured Indoor Channels with Transmitter Noise

  Authors: Paula M. Castro, José P. González-Coma, José A. García-Naya, and Luis Castedo

  Journal: EURASIP Journal on Wireless Communications and Networking (WCN)

  Impact factor in 2011: 0.873 (Q3 T2 45/78 Telecommunications)

  ISSN: 1687-1499

  Date: 2012

- Title: Power Minimization in the Multiuser Downlink under User Rate Constraints and Imperfect Transmitter CSI

  Authors: José P. González-Coma, Michael Joham, Paula M. Castro, and Luis Castedo

  Conference: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)

  Location: Vancouver, Canada

  Date: May, 2013

- Title: Power Minimization and QoS Feasibility Region in the Multiuser MIMO Broadcast Channel with Imperfect CSI

  Authors: José P. González-Coma, Michael Joham, Paula M. Castro, and Luis Castedo

  Conference: International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)

  Location: Darmstadt, Germany

  Date: June, 2013

- Title: Feedback and Power Optimization in the Multiuser Downlink under QoS Constraints

  Authors: José P. González-Coma, Michael Joham, Paula M. Castro, and Luis Castedo

  Conference: 6th International Symposium on Communications, Control, and Signal Processing (ISCCSP)

  Location: Athens, Greece

  Date: May, 2014

- Title: Power Minimization in the Multiple Stream MIMO Broadcast Channel with Imperfect CSI

  Authors: José P. González-Coma, Michael Joham, Paula M. Castro, and Luis Castedo

  Conference: Eighth IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)

  Location: A Coruña, Spain

  Date: June, 2014

- Title: Analog Joint Source-Channel Coding for MIMO Multiple Access Channels with Imperfect CSI

  Authors: Óscar Fresnedo Arias, José P. González-Coma, Luis Castedo, and J. García-Frías

  Conference: Eighth IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)

  Location: A Coruña, Spain

  Date: June, 2014

- Title: Design of MAC Access Schemes for Analog Joint Source Channel Coding

  Authors: Óscar Fresnedo Arias, José P. González-Coma, Luis Castedo, and J. García-Frías

  Conference: International Symposium on Wireless Communication Systems (ISWCS)

  Location: Barcelona, Spain

  Date: August, 2014

- Title: Power Minimization in the Multiuser MIMO-OFDM Broadcast Channel with Imperfect CSI

  Authors: José P. González-Coma, Michael Joham, Paula M. Castro, and Luis Castedo

  Conference: 22nd European Signal Processing Conference (EUSIPCO)

  Location: Lisbon, Portugal

  Date: September, 2014

- Title: Evaluation of Analog Joint Source-Channel Coding Systems for Multiple Access Channels

  Authors: Óscar Fresnedo Arias, José P. González-Coma, Luis Castedo, and J. García-Frías

  Journal: Submitted to the IEEE Transactions on Communications

  Date: October, 2014

- Title: Average Sum MSE Minimization in the Multi-User Downlink With Multiple Power Constraints

  Authors: Andreas Gründinger, Michael Joham, José Pablo González-Coma, Luis Castedo, and Wolfgang Utschick

Conference: Asilomar Conference on Signals, Systems, and Computers (ACSSC)

Location: California, USA

Date: November, 2014

- Title: QoS Constrained Power Minimization in the MISO Broadcast Channel with Imperfect CSI

  Authors: José P. González-Coma, Michael Joham, Paula M. Castro, and Luis Castedo

  Journal: Submitted to the IEEE Transactions on Signal Processing

  Date: December 2014

1.5. Thesis Overview

This thesis is organized as follows:

In Chapter 2, we present the system model for MIMO systems, considering both Single-User and Multi-User scenarios. We introduce the parameters that allow to characterize the wireless channels and the signal model for the downlink of a multiuser system. Finally, the two major options (FDD and TDD) to perform the channel state information acquisition are introduced.

Next, in Chapter 3 we review different fundamental concepts for the understanding of this work, viz. channel capacity, MSE, and dualities between the BC and the MAC preserving different MSE layers and the total transmit power.

In Chapter 4, we study the power minimization in the MISO BC for perfect and imperfect CSIT assumptions. An algorithm solution is proposed and evaluated with simulation experiments.

In Chapter 5, we extend the model to the MIMO BC and solve the power minimization problem. Moreover, we develop a test to determine the feasibility of the QoS constraints. Finally, a different problem is studied, where the total transmit power is fixed and the average rates for all the users are balanced.

In Chapter 6, we address several additional issues, namely, allocating multiple streams per-user or employing the OFDM modulation, both of them resulting into additional complexity to design the minimum transmit power precoders. Furthermore, the design of the feedback is considered and it is optimized jointly with the precoders.

Finally, Chapter 7 is dedicated to the conclusions and future work.

1.6. Notation

In this section we introduce the notation used throughout this work.

Real scalar	$x \in \mathbb{R}$
Complex scalar	$x \in \mathbb{C}$
Complex conjugate	$(\cdot)^*$
Real part	$\Re\{\cdot\}$
Imaginary part	$\Im\{\cdot\}$
Absolute value	$\lvert \cdot \rvert$
$\sqrt{-1}$	j
Set union	$\bigcup(\cdot)$
Statistical expectation	$\mathrm{E}[\cdot]$
Conditional expectation	$\mathrm{E}[\cdot\lvert x]$
Probability of an event x	$\Pr\{x\}$
Euler number $\sum_{n=0}^{\infty} \frac{1}{n!} \approx 2.71828$	e
Minimization of $f(x)$ w.r.t. x	$\min_x f(x)$
Maximization of $f(x)$ w.r.t. x	$\max_x f(x)$
Argument x' that minimizes $f(x)$	$x' = \operatorname{argmin}_x f(x)$
Logarithmic function	$\log(\cdot)$
Sign function $f(x) = \frac{x}{\lvert x \rvert}$	$\operatorname{sgn}(\cdot)$

Table 1.1: General notation.

Matrix	$\boldsymbol{X}$
Column vector	$\boldsymbol{x}$
Canonical vector, all components equal to 0 except for the kth which is 1	$\boldsymbol{e}_k$
$M \times N$ real matrix	$\boldsymbol{X} \in \mathbb{R}^{M \times N}$
$M \times N$ complex matrix	$\boldsymbol{X} \in \mathbb{C}^{M \times N}$
Element at row j and column k	$[\boldsymbol{X}]_{j,k}$
Diagonal matrix $\boldsymbol{D}$, with $[\boldsymbol{D}]_{i,i} = d_i$	$\mathrm{diag}(d_i)$
$N \times N$ identity matrix	$\mathbf{I}_N$
All zeros vector	$\mathbf{0}$
All ones vector	$\mathbf{1}$
Transpose	$(\cdot)^{\mathrm{T}}$
Hermitian	$(\cdot)^{\mathrm{H}}$
Real part	$\Re\{\cdot\}$
Circularly Symmetric Complex Gaussian random variable $\boldsymbol{X}$ with mean $\boldsymbol{\mu_x}$ and covariance $\boldsymbol{C_x}$	$\boldsymbol{X} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu_x}, \boldsymbol{C_x})$
Block diagonal matrix	$\mathrm{blockdiag}(\boldsymbol{X}_i)$
Vectorize the $R \times N$ matrix $\boldsymbol{A}$ such that $\boldsymbol{a} = [[\boldsymbol{A}]_{1:R,1}^{\mathrm{T}}, \ldots, [\boldsymbol{A}]_{1:R,N}]^{\mathrm{T}}]^{\mathrm{T}}$	$\mathrm{vec}(\cdot)$
Matrix inverse	$\boldsymbol{A}^{-1}$
Cholesky decomposition of $\boldsymbol{A}$	$\boldsymbol{A}^{1/2}$
Rank	$\mathrm{rank}(\cdot)$
Trace	$\mathrm{tr}(\cdot)$
Determinant	$\det(\cdot)$
One norm	$\|\|\cdot\|\|_1$
Euclidean norm	$\|\|\cdot\|\|_2$
Frobenious norm	$\|\|\cdot\|\|_{\mathrm{F}}$

Table 1.2: Vector and Matrix Notation.

1.7. Assumptions

All the derivations are based on zero-mean circularly symmetric complex Gaussian random symbols. In addition, the symbols for the different users are independent. We also assume perfect knowledge of the second-order statistics of the zero-mean Gaussian noise and that all random variables are stationary. We consider that CSI is perfectly known at each user. Moreover, the users do not cooperate to reduce the interference, nor share knowledge about the CSI, so that each one sends information to the BS about the channel statistics. The feedback link between the users and the BS is bandwidth limited. Thus, the users must reduce the amount of data to be sent to the transmitter by using compression techniques, like quantization. A combination of the information obtained from all the users is the CSI available at the BS. Regarding the CSI uncertainty, we do not assume a particular shape for that, and it is modeled as a stochastic error.

Chapter 2

System Model

Contrarily to wired communications where the channel remains almost unchanged during long periods of time, in wireless communications is not possible to foresee the channel behavior. Due to that, wireless channels are modeled as a random process.

A transmitted radio signal usually propagates through several different paths before arriving at the receiver. This effect is known as multipath propagation. The multiple paths arise from different effects such as scattering, reflection, diffraction, or refraction, caused by obstacles in the propagation environment. As a result, the received signal consists of an infinite sum of attenuated, delayed, and phase-shifted replicas of the transmitted signal. This combination of signals can be constructive or destructive, depending on the phase of each wave, and can lead to severe performance degradations. The fluctuations in the received signal level are termed fading. All these effects have been widely studied in the literature, e.g. [41, 42].

In this complicated environment, the use of multiple antennas at both ends of the communication link (i.e. *Multiple-Input Multiple-Output* (MIMO)) increases the reliability due to diversity gain. Having a MIMO channel also provides an additional spatial dimension and a degree of freedom gain, resulting in channel capacity gains without any extra bandwidth or power. The first works considering MIMO techniques appeared during the 1990s decade [43]. Spatial multiplexing allows to send several data streams simultaneously, and serve more than one user at the same time. The cost of deploying MIMO technologies is the added complexity to process multidimensional signals and perform the spatial data separation.

The benefits of using MIMO are appreciated by the introduction of this technology into modern wireless communication standards such as IEEE 802.11n, IEEE 802.11ac (*WiFi*) [44], 4G, *Third Generation Partnership Project* (3GPP) *Long-Term Evolution* (LTE), *Worldwide Interoperability for Microwave Access* (WiMAX), or *High-Speed Packet Access* (HSPA)+.

Current wireless communication systems often have a transmitter that sends

independent data streams to several users. Thanks to MIMO technologies it is possible to send different signals towards different users simultaneously, and over the same bandwidth. Signals are separated at reception due to the MIMO channel spatial dimension. This communication link between a *Base Station* (BS) and more than one users is known as *Broadcast Channel* (BC). When the data flow goes in the opposite direction, several users send information at the same time to the BS and this is termed *Multiple Access Channel* (MAC). For both of them the users share the channel and the capacity has to be divided among them, that is, the resource allocation is an additional task to be considered when there exists more than one user in the system.

2.1. Single-User MIMO Systems

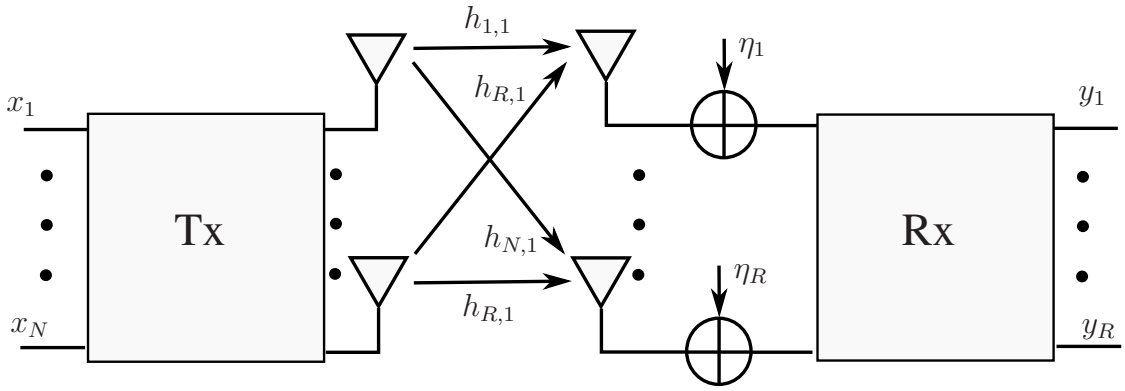


Figure 2.1: Sigle-User MIMO Model.

Figure 2.1 shows the block diagram of a general single-user MIMO communication system with N antennas at the transmitter and R at the receiver. *Single-Input Single-Output* (SISO) denotes the system with only one transmit and receive antenna i.e., $N = R = 1$. *Multiple-Input Single-Output* (MISO) indicates a scenario with only one receive antenna but several antennas at the transmitter, i.e., $N > R = 1$. And finally, *Single-Input Multiple-Output* (SIMO) is used to denote those systems with one transmit antenna but more than one receive antenna i.e., $R > N = 1$.

The MIMO channel is represented by an $R \times N$ matrix whose elements correspond to the equivalent channel impulse responses for every combination of pairs between transmit and receive antennas. More specifically, MIMO channels are represented by the channel

matrix $\boldsymbol{H}(\tau, t)$ defined as

$$\boldsymbol{H}(\tau,t) = \begin{pmatrix} h_{1,1}(\tau,t) & h_{1,2}(\tau,t) & \cdots & h_{1,N}(\tau,t) \\ h_{2,1}(\tau,t) & h_{2,2}(\tau,t) & \cdots & h_{2,N}(\tau,t) \\ \vdots & \vdots & \ddots & \vdots \\ h_{R,1}(\tau,t) & h_{R,2}(\tau,t) & \cdots & h_{R,N}(\tau,t) \end{pmatrix} \tag{2.1}$$

where t and τ stands for the time and the tap, respectively, and $[\boldsymbol{H}(\tau,t)]_{i,j} = h_{i,j}(\tau,t)$ is the channel impulse response between the jth transmit antenna and the ith receive antenna.

Using MIMO schemes allows increasing the reliability of a communication link thanks to diversity. It allows reducing the probability of a fade by increasing the number of antennas. That way, the probability of having poor gains in all the independent paths at the same time is low. On the other hand, the multiple antennas provide additional degrees of freedom from the spatial dimension, allowing to multiplex several independent data streams into the channel. As a consequence, the spatial multiplexing leads to a higher channel capacity.

When we restrict ourselves to a narrowband (flat fading) channel, the parameter τ is removed. Moreover, if the MIMO channel is time-invariant, the coefficients are the same for every t, and obviously the time variable is no longer needed. Hence, the MIMO channel reduces to

$$\boldsymbol{H} = \begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,N} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{R,1} & h_{R,2} & \cdots & h_{R,N} \end{pmatrix}. \tag{2.2}$$

Under the assumptions of flat fading and time invariance, the received signal in a single-user MIMO channel (see Fig. 2.1) is given by

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\eta}, \tag{2.3}$$

where $\boldsymbol{x} \in \mathbb{C}^N$ and $\boldsymbol{y} \in \mathbb{C}^R$ are the vectors of transmitted and received symbols, respectively, and $\boldsymbol{\eta} \in \mathbb{C}^R$ is the noise vector that represents the additive noise at the receiver, e.g. $\boldsymbol{\eta} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma^2 \mathbf{I}_R)$. This notation represents a zero-mean Gaussian noise. Note that such noise is white since the vector components are statistically independent (the covariance matrix is diagonal).

2.2. Multi-User Systems

We now consider multi-user systems, where a centralized transmitter (receiver) having multiple antennas transmits (receives) independent data streams to different users possibly having multiple antennas themselves.

This type of scenarios is more complex to model than the former ones. To establish a communication link with more than one user means that the channel has to be shared between all of them. This fact constitutes a new challenge since the resources are divided among the users following some policy, making the multiuser systems difficult to deal with. In the ensuing sections, we will distinguish between two types of multi-user communications.

2.2.1. Broadcast Channels

Here we consider the BC, where a single transmitter (BS) serves to several users. One example of BC is the WiFi router, which sends information to some devices; or a cellular system where a BS serves several mobile terminals in a certain area. When the data is transmitted from the BS to the users, this communication is also commonly named downlink. Fig. 2.2 shows the system model of a MISO BC. The BS is equipped with

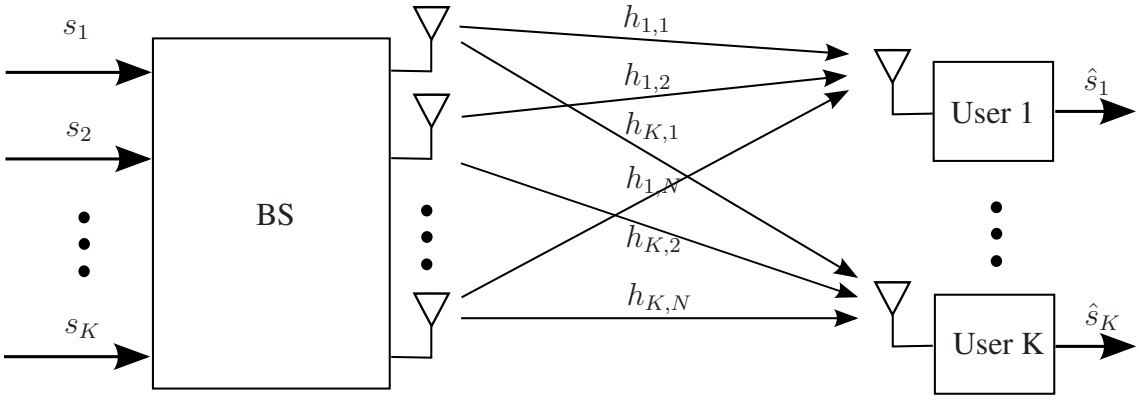


Figure 2.2: MISO Broadcast Channel.

N transmit antennas and sends K independent data signal $s_k \in \mathbb{C}$, $k \in \{1, \ldots, K\}$. Before the transmission, each data signal is transformed to obtain the signal $\boldsymbol{x}_k \in \mathbb{C}^N$. The transformation (either linear or non-linear) that produces the transmitted symbols $\boldsymbol{x}_k$ from the input data symbols s_k is often referred to as precoding. Then, $\boldsymbol{x} = \sum_{i=1}^{K} \boldsymbol{x}_i$ propagates over the MISO channels $\boldsymbol{h}_k \in \mathbb{C}^N$, also known as vector channels. Such MISO channels $\boldsymbol{h}_k$ are assumed to be a flat fading (see Section 2.3), for all k. At the user k, the received signal y_k perturbed with the noise $\eta_k \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_{\eta_k}^2)$ is obtained as follows

$$y_k = \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{x} + \eta_k. \tag{2.4}$$

Finally, the received signal is processed to get $\hat{s}_k$, the estimated data at the kth user.

Note that the transmitted signal $\boldsymbol{x}$ is the linear combination of the signals containing the data for the K users. Since the users are equipped with one antenna each and the

receivers are very simple, the complexity is concentrated at the BS where the precoders make the spatial separation possible. This scheme is useful when the users have low computational capacity.

Along this work we restrict ourselves to the case where the signal processing at the transmission and reception is carried out by means of linear filtering. More specifically, we assume that the transmitted signal $\boldsymbol{x}_k$ is obtained by linear precoding the data signal s_k, i.e. $\boldsymbol{x}_k = \boldsymbol{p}_k s_k$ where $\boldsymbol{p}_k \in \mathbb{C}^N$ represents the response of the kth user linear precoder. Hence, the signal transmitted over the BS is $\boldsymbol{x} = \sum_{i=1}^{K} \boldsymbol{p}_i s_i$. At the users, the received signal y_k (cf. (2.4)) is also linearly filtered. In a MISO BC, the receiver linear filter response reduces to a scalar value $f_k \in \mathbb{C}$. Elaborating the signal model, the estimated data $\hat{s}_k = f_k^* y_k$ is given by

$$\hat{s}_k = f_k^* \boldsymbol{h}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i s_i + f_k^* \eta_k. \tag{2.5}$$

2.2.2. Multiple Access Channel

In this subsection we introduce a different type of multiuser communication system termed MAC. In the MAC, multiple users share a common communication channel to transmit information to a single receiver. In this scenario, the users may have one or more antennas while the BS is often equipped with more than one. The data flow is the opposite to that in the BC, and it is also commonly referred as the uplink. MAC is the typical scheme in cellular systems when the mobile users communicate with the BS.

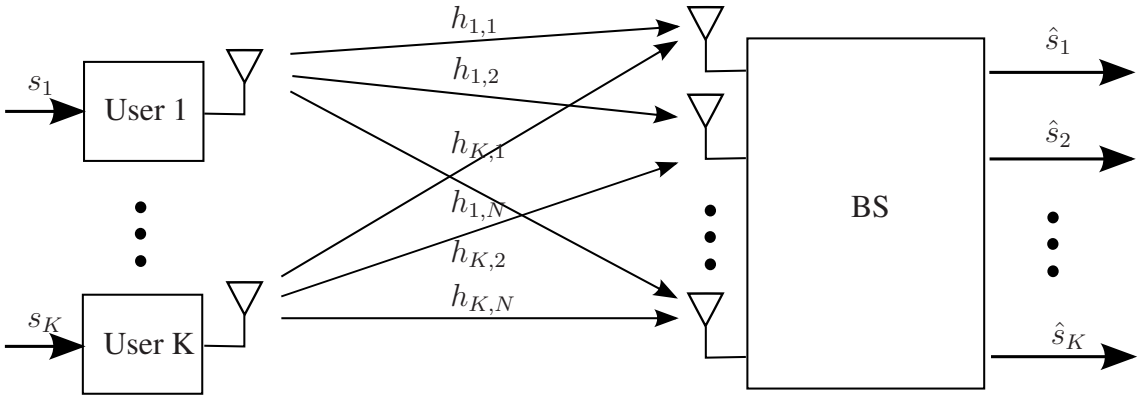


Figure 2.3: MISO Multiple Access Channel.

Figure 2.3 depicts the block diagram of a MISO MAC. The data sent by the user k, $s_k \in \mathbb{C}$, are transformed to get the transmitted signal x_k. It then propagates over the

vector channel $\boldsymbol{h}_k$. The signal at the centralized receiver, $\boldsymbol{y}$, is given by

$$\boldsymbol{y} = \sum_{i=1}^{K} \boldsymbol{h}_i x_i + \boldsymbol{\eta}, \tag{2.6}$$

where $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma_{\boldsymbol{\eta}}^2 \mathbf{I})$ represents the thermal noise. The received signal $\boldsymbol{y}$ is then processed to extract the information from each user.

Observe that the received signal contains the linear combination of the channels multiplied by the transmitted signals for all the users, together with the noise. Just like in the BC, the complexity is located at the BS, where the multiple antennas provide degrees of freedom enough to separate the data.

As for the BC, along this work we assume that signals are processed at either the transmitter and the receiver with linear filters. Hence the users send the transmit signal $x_k = t_k s_k$, where $t_k \in \mathbb{C}$ is the linear precoder. After the propagation over the channel, the BS filters the vector $\boldsymbol{y}$ with the linear receive filter of each user, $\boldsymbol{g}_k \in \mathbb{C}^N$, to obtain the estimated data $\hat{s}_k$

$$\hat{s}_k = \boldsymbol{g}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{h}_i t_i s_i + \boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{\eta}. \tag{2.7}$$

The natural extension of the BC and MAC systems above presented is to incorporate additional antennas at the users. That way, the dimensions of receive filters in the BC, precoders in the MAC, and channels of each user increase leading to the MIMO BC and MIMO MAC, respectively.

2.3. Channel Model

Wireless communications consist of electromagnetic radiation from the transmitter to the receiver. The channel models the propagation of the waves and takes into account obstructions caused by ground, buildings, mountains, etc, as well as other effects. When the wave reflects on certain objects, we have multiple versions of the transmitted signal that will arrive at the receiver from different paths either constructively or destructively. Then, for a point-to-point SISO communication, the received signal can be modeled as a *Linear Time-Invariant* (LTI) system [45, 46]

$$y\left(t\right) = \sum_{i=1}^{P} a_i(t) x(t - \tau_i(t)) + \eta(t), \tag{2.8}$$

where $a_i(t)$ and $\tau_i(t)$ are the attenuation and delay of the ith path, respectively, P is the number of paths, $\eta(t)$ is the noise, and $x(t)$ is the sine wave signal. Therefore, the base-

band equivalent discrete-time model reads as

$$y\left[m\right] = \sum_{i=1}^{P} h_i\left[m\right] x\left[m-i\right] + \eta[m]. \tag{2.9}$$

In the following, we will introduce some parameters that characterize the time-variant channel [2, 46].

The delay spread, denoted as T_d, is the time difference between the first and the last paths. It is important because it determines the channel coherence bandwidth, $W_c = \frac{1}{2T_d}$ i.e., the range of frequencies where the channel is considered as constant. In other words, the channel changes significantly when we move more than W_c *Hertz* (Hz) from a given frequency.

Another channel parameter is the coherence time, T_c. The definition is analogous to that of the frequency coherence, and it represents the interval where the time domain channel, $h_i(t)$, does not significantly change.

A large number of applications exists that transmit information over wireless channels. Depending on the application, there exists an allowed delay for which the application is working properly. This delay is translated into a bandwidth W in the frequency domain. According to the relationship between the coherence time T_c and the symbol period $T_s = \frac{1}{W}$, the channels are classified as slow or fast fading in the literature. That is, when the coherence time is shorter than the time requirement of each symbol, i.e., $T_c \ll T_s$, we have a fast fading channel and a single symbol is transmitted over several channel fades. Conversely, when the coherence time is larger than the symbol period we have slow fading, which means that the symbol is transmitted over a single channel fade.

Regarding the symbol bandwidth W and the channel coherence bandwidth, we can distinguish between flat fading and frequency-selective fading channels. The first case occurs when the bandwidth of the input signal is smaller than the coherence bandwidth, i.e., $W \ll W_c$, and implies that the channel is represented with a single tap. However, in the opposite case, the frequency-selectivity has to be expressed with multiple taps.

The following table summarizes the relationships between the parameters explained above and the channel characterizations [45].

Parameter specification	Channel type
$T_c \ll T_s$	Fast fading
$T_c \gg T_s$	Slow fading
$W \ll W_c$	Flat fading
$W \gg W_c$	Frequency-selective fading

Table 2.1: Channel Characterization.

When considering MIMO systems, the location of the antennas also plays an important role in the resulting channel. The angle spread refers to the spread in angles of arrival (or departure) of the multipath components at the antenna array. It causes space selective fading, which means that the received signal amplitudes depend on the antennas' spatial location. Space selective fading is characterized by the coherence distance, which is inversely proportional to the angle spread.

2.3.1. Rayleigh Channels

Wireless channels are typically modeled by random processes. More specifically, a general assumption is to model wireless channels as ergodic random processes. When ergodicity holds, the time average for all the moments of the process equals the statistical average regardless the moment that is chosen. In other words, the statistics can be observed.

In this section we focus on a statistical channel model. Particularly, we present a flat fading channel. That is to say, the channel is represented with a single tap $h[m]$, where m is the sample of the equivalent discrete-time base-band model. In practice, the different paths $h_i[m]$ present in (2.9) are modeled as statistically independent circularly symmetric random variables. Moreover, in *Non Line-of-Sight* (NLOS) scenarios, where no direct vision between the transmitter and the receiver exists, all contributions have the same importance. Due to the fact that the number of paths, P, is large, the summation of all these contributions follows a zero-mean Gaussian distribution due to the Central Limit Theorem, i.e.

$$h \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2). \tag{2.10}$$

Note that we dropped the sample index $[m]$ for the sake of notational brevity. Furthermore, together with the uniform distribution of the phase of the random variables, it leads to the absolute value of the tap $|h|$ being Rayleigh distributed, i.e., the *probability density function* (pdf) is

$$f_{|h|}\left(|h|\right) = \frac{|h|}{\sigma^2}\, \mathrm{e}^{\frac{-|h|^2}{2\sigma^2}}, \quad |h| \geq 0, \tag{2.11}$$

where σ^2 is the variance of the random variable h.

We now extend these concepts to the *Multi-User* (MU) MISO channel model, where a transmitter with N antennas, $N > 1$, sends data to a single antenna receiver corresponding to user k, i.e. $\boldsymbol{h}_k = [h_{k,1}, \ldots, h_{k,N}]^{\mathrm{T}}$. We assume each entry in $\boldsymbol{h}_k$ is Gaussian distributed $h_{k,i} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_{k,i}^2)$, $\forall k, i$. Moreover, the antenna array is assumed to be constructed in a way that the channel response between two different transmit antennas is statistically independent, i.e., $\mathrm{E}[h_{k,i} h_{k,j}^*] = 0$ for $j \neq i$ and $\mathrm{E}[h_{k,i} h_{k,j}^*] = \sigma_{k,i}^2$ for $j = i$, respectively. In summary, the kth user's channel vector response is modeled as a stationary zero-mean

circularly symmetric complex Gaussian random vector $\boldsymbol{h}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{h}_k})$, where $\boldsymbol{C}_{\boldsymbol{h}_k}$ is the covariance matrix, given by

$$\boldsymbol{C}_{\boldsymbol{h}_k} = \mathrm{E}[\boldsymbol{h}_k \boldsymbol{h}_k^{\mathrm{H}}] = \mathrm{diag}\left(\sigma_{k,1}^2, \ldots, \sigma_{k,N}^2\right). \tag{2.12}$$

Finally, we assume that the channels corresponding to the different users are statistically independent.

Observe that the previous channel model can be obtained from a circularly symmetric white Gaussian random vector, $\boldsymbol{h}_w \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I})$. Then, the kth channel vector can be represented as

$$\boldsymbol{h}_k = \boldsymbol{C}_{\boldsymbol{h}_k}^{1/2} \boldsymbol{h}_w, \tag{2.13}$$

where $(\cdot)^{1/2}$ is the Cholesky decomposition (see Section C.5 of Appendix C). The matrix $\boldsymbol{C}_{\boldsymbol{h}_k}^{1/2}$ represents the spatial correlation between the antennas, and it is realistic to assume that it is constant since it changes very slowly. Notice that the expectation of the right side of (2.13) is $\mathrm{E}[\boldsymbol{C}_{\boldsymbol{h}_k}^{1/2} \boldsymbol{h}_w] = \boldsymbol{C}_{\boldsymbol{h}_k}^{1/2} \mathrm{E}[\boldsymbol{h}_w] = \mathbf{0}$, whereas the covariance reads as

$$\mathrm{E}[\boldsymbol{C}_{\boldsymbol{h}_k}^{1/2} \boldsymbol{h}_w \boldsymbol{h}_w^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{h}_k}^{1/2,\mathrm{H}}] = \boldsymbol{C}_{\boldsymbol{h}_k}^{1/2} \mathrm{E}[\boldsymbol{h}_w \boldsymbol{h}_w^{\mathrm{H}}] \boldsymbol{C}_{\boldsymbol{h}_k}^{1/2,\mathrm{H}} = \boldsymbol{C}_{\boldsymbol{h}_k}, \tag{2.14}$$

since $\mathrm{E}[\boldsymbol{h}_w \boldsymbol{h}_w^{\mathrm{H}}] = \mathbf{I}$.

2.3.2. OFDM Channels

Orthogonal Frequency Division Multiplexing (OFDM) [47, 48] is a signaling technique widely exploited in multiple communication standards, like WiMAX, LTE 4G mobile communications, digital television, or audio broadcasting. It is employed in wideband communications due to its ability to cope with the channel frequency selectivity (see Table 2.1).

OFDM divides the wideband frequency-selective fading channel into multiple flat fading narrowband channels. One of the OFDM strengths with respect to other multicarrier modulations lies on the partial frequency overlapping of the subcarriers that allows to efficiently take advantage of the wideband channel.

Let us consider the baseband equivalent channel model presented in (2.9). In the OFDM modulator block diagram shown in Fig. 2.4, a block of N data symbols $\boldsymbol{x} \in \mathbb{C}^N$ is transmitted. The N symbols are transformed prior to transmission by means of the *Inverse Discrete Fourier Transform* (IDFT) i.e., $\overline{\boldsymbol{s}} = \boldsymbol{F}^{\mathrm{H}} \boldsymbol{x}$, where $\boldsymbol{F}$ is the *Discrete Fourier Transform* (DFT) matrix. The k, i entry of $\boldsymbol{F}$ is given by

$$[\boldsymbol{F}]_{k,i} = \frac{1}{\sqrt{N}} \mathrm{e}^{\frac{-j2\pi ki}{N}} \quad k, i = 1, ..., N. \tag{2.15}$$

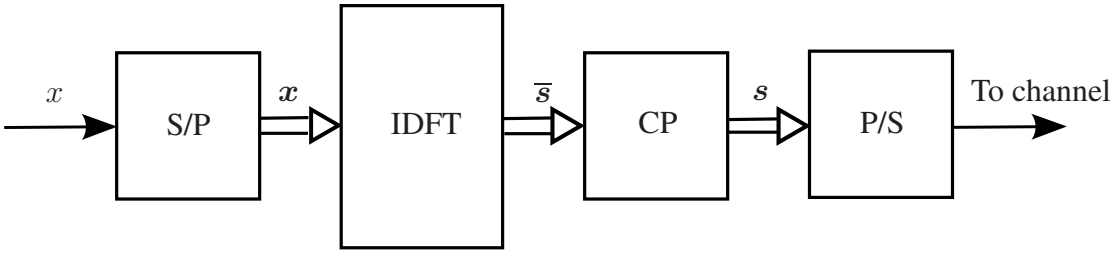


Figure 2.4: OFDM Modulator.

In the next step a cyclic prefix of length $\overline{P}$, where $\overline{P}$ is greater or equal than the number of channel taps minus one (i.e., $\overline{P} \geq P-1$), is added in order to remove the channel *Inter-Symbol Interference* (ISI). This results in an OFDM symbol of length $L = N + \overline{P}$ that can be represented by $\boldsymbol{s} = [\overline{s}_{N-\overline{P}+1}, \ldots, \overline{s}_N, \overline{s}_1, \ldots, \overline{s}_N]^{\mathrm{T}}$. When the OFDM symbol is sent over the channel, the received signal is

$$y[l] = \sum_{i=1}^{P} h_i s[l-i+1] + \eta[l], \quad l = 1, \ldots, L, \tag{2.16}$$

where $\eta[l] \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$. When the L samples of the OFDM symbol are received i.e., the vector $\boldsymbol{y}$ is received, we remove the cyclic prefix to get $\overline{\boldsymbol{y}} = \boldsymbol{H}\overline{\boldsymbol{s}} + \boldsymbol{\eta}$. That way, the $N \times N$ matrix $\boldsymbol{H}$ becomes circular and is given as follows

$$\boldsymbol{H} = \begin{pmatrix} h_1 & h_N & \ldots & h_2 \\ h_2 & h_1 & \ldots & h_3 \\ \vdots & \vdots & \ddots & \vdots \\ h_N & h_{N-1} & \ldots & h_1 \end{pmatrix}, \tag{2.17}$$

where $h_l = 0, \forall l > P$.

At the receiver, the DFT is applied to the received signal, i.e. $\boldsymbol{F}\overline{\boldsymbol{y}} = \boldsymbol{F}\boldsymbol{H}\overline{\boldsymbol{s}} + \boldsymbol{F}\boldsymbol{\eta} = \boldsymbol{F}\boldsymbol{H}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{x} + \tilde{\boldsymbol{\eta}}$. Then, the equivalent channel is that given by $\boldsymbol{H}_{\text{eq}} = \boldsymbol{F}\boldsymbol{H}\boldsymbol{F}^{\mathrm{H}}$, and the white noise is preserved since $\mathrm{E}[\tilde{\boldsymbol{\eta}}\tilde{\boldsymbol{\eta}}^{\mathrm{H}}] = \mathrm{E}[\boldsymbol{F}\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}] = \sigma^2\boldsymbol{I}_N$. Notice that the normalized eigenvector of the ith column of the circular matrix is $\boldsymbol{u}_i = 1/\sqrt{N}[1, \mathrm{e}^{j2\pi i/N}, \ldots, \mathrm{e}^{j2\pi(N-1)i/N}]^{\mathrm{T}}$ [49]. On the other hand, the eigenvalue decomposition (see Section C.1 of Appendix C) of the matrix $\boldsymbol{H}$ is $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}$. As a consequence, the DFT matrix $\boldsymbol{F}$ contains the eigenvectors of the matrix $\boldsymbol{H}$, so that the left product times the DFT and the right product times the IDFT matrices gives the following diagonal matrix,

$$\boldsymbol{H}_{\text{eq}} = \boldsymbol{F}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}} = \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Lambda}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N), \tag{2.18}$$

where $\lambda_i = \sum_{p=1}^{P} h_p \, \mathrm{e}^{-j2\pi i(p-1)/N}$ (see [49]). Note that the eigenvectors of circular matrices are independent of the matrix coefficients. Therefore, the equivalent channel employing the OFDM modulation is the diagonal channel $\boldsymbol{H}_{\text{eq}}$,

$$\boldsymbol{H}_{\text{eq}} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_N \end{pmatrix}. \tag{2.19}$$

In summary, the effect produced by the channel over the transmitted OFDM symbols are N multiplications by different scalars, which correspond to the eigenvalues of the equivalent channel $\boldsymbol{H}_{\text{eq}}$. It is important to note that our frequency-selective fading channel has been transformed into N flat fading parallel channels employing the matrices $\boldsymbol{F}^{\text{H}}$ and $\boldsymbol{F}$ at the transmitter and the receiver, respectively, and adding a proper cyclic prefix. In other words, the N symbols are transmitted independently in the frequency domain.

We can naturally extend the OFDM channel to MIMO scenarios. Recall the equivalent channel matrix in (2.19) for single-user SISO communications. When the transmitter has more than one antenna there exist interferences between the subcarriers $n \in \{1, \dots, N\}$, and $m \in \{1, \dots, N\}$ from other antennas only if $m = n$. However, if the cyclic prefix is long enough, the interference with other subcarriers can be totally removed during the demodulation process i.e., with the DFT at the receiver, leading to the following equivalent channel,

$$\boldsymbol{H}_{\text{eq}} = \begin{pmatrix} \boldsymbol{H}_1 & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}_2 & \dots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{H}_N \end{pmatrix}, \tag{2.20}$$

where the matrices $\boldsymbol{H}_i \in \mathbb{C}^{R \times T}$, $\forall i$, with T and R being the number of transmit and receive antennas, respectively. In conclusion, the equivalent MIMO-OFDM channel is represented by a block-diagonal matrix when the cyclic prefix is long enough.

2.4. Channel State Information

Channel State Information (CSI) is fundamental in the design of wireless communication systems. However, the channel knowledge by the two sides of the communication link, transmitter(s) and receiver(s), is usually different in practical systems. We denote the channel information available at transmission as *Channel State Information at the Transmitter* (CSIT), whereas *Channel State Information at the Receiver* (CSIR) stands for the channel knowledge at reception.

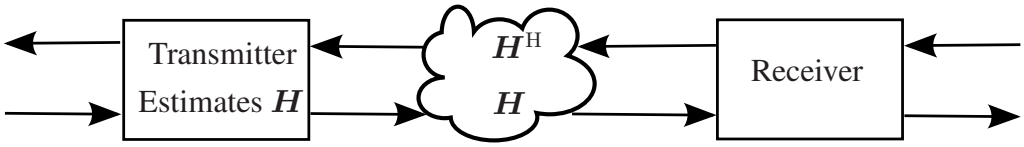


Figure 2.5: Obtaining CSIT using Reciprocity.

Since the channel is not known prior to transmission in practical systems, the assumption of perfect CSIT knowledge is unrealistic. As a consequence, it is important to explain how this information is acquired.

Bearing in mind that most communication systems are bidirectional, the uplink and downlink channels must be separated into orthogonal signaling dimensions. This separation is called *duplexing*. We consider two types of systems, *Time-Division Duplex* (TDD) and *Frequency-Division Duplex* (FDD), where different solutions are applied [50].

2.4.1. Channel State Information in TDD Systems

Figure 2.5 shows the way of obtaining CSIT in TDD systems. The transmitter only acquires the CSI indirectly since the signal goes into the channel only after leaving the transmitter [50]. Therefore, the CSI can be obtained by using the *reciprocity principle*.

The reciprocity of the wireless channel implies that the channel from the transmitter to the receiver is estimated during the transmission in the opposite direction i.e., from the receiver to the transmitter, since the relationship between both of them is only the Hermitian (e.g. [51, 52]). Pilot symbols are often used for channel estimation. The reciprocity holds if both forward and reverse links are located at the same frequency, the same time, and the same antenna locations. In practical systems, however, the forward and reverse links cannot use identical frequency, time, and spatial locations, but the reciprocity still holds if the lags between both links are respectively much smaller than the channel coherence time, the channel coherence bandwidth, and the channel coherence distance [53].

Practical channel acquisition based on reciprocity may be applicable in TDD-*Time-Division Multiple Access* (TDMA) systems [50, 54–56]. TDMA consists of dividing the frame duration T_{f} into T non-overlapping subintervals, each of duration T_{f}/T. Each transmitter has to use a particular subinterval within each frame. In TDD systems, orthogonal time slots are assigned to each user to transmit to the base station and to receive from the base station. While TDD-TDMA systems have identical forward and reverse frequency bands and antennas, there is a time lag between the forward and reverse links. As mentioned above, such time lags must be negligible compared to the channel coherence time. Even in this case, reciprocity is difficult to accomplish due to the need for very good calibration (e.g. [57]).

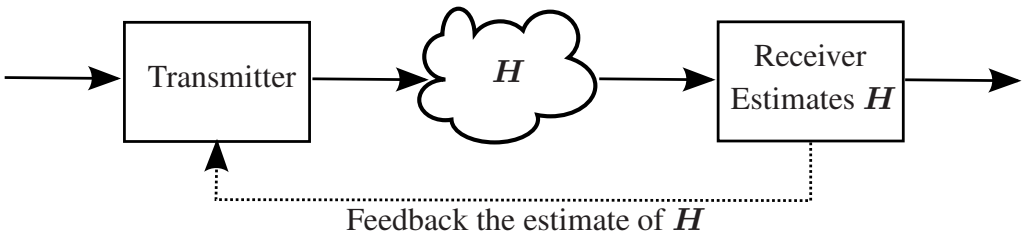


Figure 2.6: Obtaining CSIT using Feedback.

2.4.2. Channel State Information in FDD Systems

Since in FDD systems the reciprocity is usually not applicable, a feedback channel should be used to send the CSI from the transmitter to the receiver, as illustrated in Fig. 2.6. The channel response is estimated at the receiver during the transmission from the transmitter to the receiver, and the resulting estimation is sent back to the transmitter on the reverse link. Such reverse channels are actually implemented in most wireless communication standards [58–60]. In this case, calibration errors are estimated as part of the CSI and no special problems arise from calibration as for TDD. However, the time lag, D, between the channel measurement at the receivers and its use at the transmitter is a source of errors unless it is much smaller than the channel coherence time.

Moreover, the data rate of the feedback channel is highly limited. One drawback of feedback is the possible overhead of the reverse channel and the increasing consumption of transmit resources. Therefore, methods of reducing feedback overhead in a simple way, such as quantization or truncation of the feedback information, are crucial for practical implementations. As a consequence of the quantization, any system with limited rate CSI feedback suffers from erroneous CSI at the transmitter.

The same is true in a multiuser system. In *Frequency-Division Multiple Access* (FDMA) systems, the available channel bandwidth is split into a number of F frequency non-overlapping subchannels. Each subchannel is assigned to a user on demand. With FDD, separate frequency bands are assigned to each user for transmitting to or receiving from the base station. Therefore, FDD-FDMA systems often have identical temporal and spatial channel dimensions, but the frequency offset between the forward and reverse links is usually much larger than the channel coherence bandwidth. Therefore, since the channel reciprocity is not true the limited feedback channels must be used to allow the transmitter to get CSI from the users.

2.4.3. Perfect and Imperfect CSI

In this section, we distinguish between the case when there exists perfect (or full) CSI, i.e. the true channel is known, and the case where only imperfect (or partial) information

about the channel is available.

Let us introduce the channel given by the random variable h, and the corresponding pdf $f_h(h)$. Consider now a realization of h, i.e. $\tilde{h}$. When there exists perfect CSI on both the transmitter and the receiver, the available information about the channel is $\tilde{h}$. Nevertheless, the general assumption in this work is imperfect CSIT and perfect CSIR, which constitutes a more realistic approach. Then, depending on $\tilde{h}$, which is known by the receiver, any of the strategies above introduced (TDD or FDD) is implemented to decide the CSI sent to the transmitter. In particular, the receiver selects one value for the random variable v from the set of all possible values $\mathcal{V}$. Thus, the information available at the transmitter is v together with the conditional pdf $f_{h|v}(h|v)$. Hence, the channel knowledge at the transmitter is statistical and, due to the partial CSI, the channel pdf is obtained using the Bayes' rule as follows

$$f_h(h) = \int_{\mathcal{V}} f_{h|v}(h|v) f_v(v) dv. \tag{2.21}$$

2.5. Conclusions

In this chapter some fundamental concepts about wireless communication systems have been reviewed.

First, we have introduced the single-user MIMO system model, and the benefits of including the spatial dimension have been also explained. Afterwards, we considered a more complicated scenario where several users receiving (or transmitting) in a simultaneous way have been included. Such scenarios are termed BC and MAC, respectively. Moreover, the users could be equipped with more than one antenna.

A brief review of the wireless channel characterization was also presented. We introduced the some useful concepts in order to classify the channels depending on how the electromagnetic waves propagate. Additionally, two different channel models have been introduced in the chapter. In particular, we studied the flat fading Rayleigh and the frequency selective OFDM channel models, considering single and multiple antennas in both cases.

Finally, a discussion about the CSI acquisition is included. The FDD and TDD systems are described as the main methods to share information about the channel state between the transmitter and the receivers.

Chapter 3

Preliminary Concepts

In this chapter we present the main tools and concepts employed over this work. First, entropy, mutual information, and capacity concepts are introduced. Next, we determine the capacity of the *Multiple-Input Multiple-Output* (MIMO) channel for both perfect and imperfect *Channel State Information at the Transmitter* (CSIT), and end up with the MIMO *Broadcast Channel* (BC) capacity. Afterwards, we will introduce the BC/ *Multiple Access Channel* (MAC) dualities for imperfect CSIT. In particular, the per-user *Mean Square Error* (MSE) duality under imperfect CSIT is a useful tool to avoid certain difficulties arising from finding the optimal MSE filters because it enables to get rid of the dependence of the transmit filters in the BC.

3.1. Channel Capacity

In 1948 Claude Shannon presented revolutionary ideas for information theory in wireless communications. Shannon introduced the channel capacity as the maximum of the mutual information over all the possible input distributions [1]. Channel capacity is the limiting data rate that can be achieved with asymptotically small error probability, and represents a key feature of any communication channel.

Let us start introducing the concept of entropy of a random variable x, denoted as $\mathcal{H}(x)$, as the expected value of data (in bits) contained in that variable, i.e.,

$$\mathcal{H}(x) = -\int_{\mathbb{C}} f_x(x) \log_2 \left(f_x(x)\right) dx, \tag{3.1}$$

where $f_x(x)$ is the *probability density function* (pdf) of x. Let us now consider a second random variable, y, that depends on x. The conditional entropy of y with respect to x is defined as

$$\mathcal{H}(y|x) = -\int_{\mathbb{C}}\int_{\mathbb{C}} f_{x,y}(x,y) \log_2 \left(f_{y|x}(y|x)\right) dxdy, \tag{3.2}$$

where $f_{x,y}(x,y)$ is the joint distribution of x and y, and $f_{y|x}(y|x)$ is the distribution of y conditioned to x. Taking into account the mutual dependency of x and y, another insightful measure is the joint entropy, which is defined as

$$\mathcal{H}(x,y) = -\int_{\mathbb{C}}\int_{\mathbb{C}} f_{x,y}(x,y)\log_2\left(f_{x,y}(x,y)\right)dxdy. \tag{3.3}$$

Furthermore, the joint entropy of a pair of random variables is the sum of the entropy of one of them plus the conditional entropy of the other one [61], i.e.,

$$\mathcal{H}(x,y) = \mathcal{H}(x) + \mathcal{H}(y|x). \tag{3.4}$$

The mutual information between a pair of random variables x and y, denoted as $\mathcal{I}(x;y)$, is defined as the amount of uncertainty that is reduced in one of them due to the knowledge about the other one, that is,

$$\mathcal{I}(x;y) = \mathcal{H}(y) - \mathcal{H}(y|x). \tag{3.5}$$

Note that, if x and y are independent, $\mathcal{H}(y|x) = \mathcal{H}(y)$, and the mutual information is zero. Contrarily, if $y = x$, then the conditional entropy is $\mathcal{H}(y|x) = 0$, and the mutual information is maximized i.e., $\mathcal{I}(x;y) = \mathcal{H}(y)$. Considering x as the input of a communication channel and y as the output, the channel capacity is defined as

$$C = \max_{f_x(x)} \mathcal{I}(x;y). \tag{3.6}$$

We now particularize the previous concepts to the single-user MIMO system presented in Section 2.1 of Chapter 2. In such case, a data vector $\boldsymbol{x} \in \mathbb{C}^N$ is transmitted from N antennas to a receiver with R antennas. Without loss of generality, $\boldsymbol{x}$ is assumed to be zero-mean and $\mathrm{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}] = \boldsymbol{C}_x$. We consider the channel $\boldsymbol{H} \in \mathbb{C}^{R\times N}$ is deterministic, and the *Additive White Gaussian Noise* (AWGN) $\boldsymbol{\eta}$, is zero-mean circularly symmetric complex Gaussian i.e., $\boldsymbol{\eta} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{\eta}})$. Hence, the output of the system is $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\eta}$.

The capacity is the maximization of the mutual information, which can be written in terms of the entropy as $\mathcal{H}(\boldsymbol{y}) - \mathcal{H}(\boldsymbol{y}|\boldsymbol{x})$ (c.f. (3.5)). The second term in the last expression, $\mathcal{H}(\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\eta}|\boldsymbol{x})$, can be reduced to $\mathcal{H}(\boldsymbol{\eta})$ since the data and the noise are statistically independent. Therefore, maximizing the mutual information is equal to maximizing $\mathcal{H}(\boldsymbol{y})$.

It is proven that a zero-mean circularly symmetric complex Gaussian vector is the entropy maximizer [2]. As a consequence, if $\boldsymbol{x}$ is zero-mean circularly symmetric complex Gaussian, the output signal $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\eta}$ would maximize the mutual information. The pdf of a circularly symmetric complex Gaussian variable $\boldsymbol{x}$, with mean $\boldsymbol{\mu}_{\boldsymbol{x}}$ and covariance $\boldsymbol{C}_{\boldsymbol{x}}$, is given by

$$f_{\boldsymbol{x}}(\boldsymbol{x}) = \det\left(\pi\boldsymbol{C}_{\boldsymbol{x}}\right)^{-1} \mathrm{e}^{\left(-(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)}. \tag{3.7}$$

In particular, when $\boldsymbol{\mu}_{\boldsymbol{x}} = \mathbf{0}$ the entropy is given by [2]

$$\begin{aligned}\mathcal{H}(\boldsymbol{x}) &= \mathrm{E}_{\boldsymbol{x}}\left[-\log_2\left(f_{\boldsymbol{x}}\left(\boldsymbol{x}\right)\right)\right] \\ &= \log_2 \det\left(\pi \boldsymbol{C}_{\boldsymbol{x}}\right) + \log_2\left(\mathrm{e}\right) \mathrm{E}_{\boldsymbol{x}}\left[\boldsymbol{x}^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{x}}^{-1} \boldsymbol{x}\right] \\ &= \log_2 \det\left(\pi \boldsymbol{C}_{\boldsymbol{x}}\right) + \log_2\left(\mathrm{e}\right) \mathrm{E}_{\boldsymbol{x}}\left[\mathrm{tr}\left(\boldsymbol{x}^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{x}}^{-1} \boldsymbol{x}\right)\right] \\ &= \log_2 \det\left(\pi \boldsymbol{C}_{\boldsymbol{x}}\right) + \log_2\left(\mathrm{e}\right) \mathrm{E}_{\boldsymbol{x}}\left[\mathrm{tr}\left(\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{x}}^{-1}\right)\right] \\ &= \log_2 \det\left(\pi \boldsymbol{C}_{\boldsymbol{x}}\right) + \log_2\left(\mathrm{e}\right) \mathrm{tr}\left(\mathrm{E}_{\boldsymbol{x}}\left[\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}\right] \boldsymbol{C}_{\boldsymbol{x}}^{-1}\right) \\ &= \log_2 \det\left(\pi\, \mathrm{e}\, \boldsymbol{C}_{\boldsymbol{x}}\right). \end{aligned} \tag{3.8}$$

Hence, the capacity expression for a MIMO deterministic channel is [2, 61]

$$\begin{aligned}\mathcal{I}\left(x; y\right) &= \mathcal{H}\left(\boldsymbol{y}\right) - \mathcal{H}\left(\boldsymbol{y}|\boldsymbol{x}\right) \\ &= \log_2 \det\left(\pi\, \mathrm{e}\left(\boldsymbol{H}\boldsymbol{C}_x\boldsymbol{H}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}}\right)\right) - \log_2 \det\left(\pi\, \mathrm{e}\, \boldsymbol{C}_{\boldsymbol{\eta}}\right) \\ &= \log_2\left(\det\left(\pi\, \mathrm{e}\left(\boldsymbol{H}\boldsymbol{C}_x\boldsymbol{H}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}}\right)\right)\left(\det\left(\pi\, \mathrm{e}\, \boldsymbol{C}_{\boldsymbol{\eta}}\right)\right)^{-1}\right) \\ &= \log_2\left(\det\left(\pi\, \mathrm{e}\left(\boldsymbol{H}\boldsymbol{C}_x\boldsymbol{H}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}}\right)\right)\frac{1}{\det\left(\pi\, \mathrm{e}\, \boldsymbol{C}_{\boldsymbol{\eta}}\right)}\right) \\ &= \log_2\left(\det\left(\pi\, \mathrm{e}\left(\boldsymbol{H}\boldsymbol{C}_x\boldsymbol{H}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}}\right)\right)\det\left(\frac{1}{\pi\, \mathrm{e}}\boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\right)\right) \\ &= \log_2 \det\left(\left(\boldsymbol{H}\boldsymbol{C}_x\boldsymbol{H}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}}\right)\boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\right) \\ &= \log_2 \det\left(\boldsymbol{H}\boldsymbol{C}_x\boldsymbol{H}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{\eta}}^{-1} + \mathbf{I}_R\right). \end{aligned} \tag{3.9}$$

Consider now that $\boldsymbol{H}$ is a random channel generated by an ergodic process as explained in Chapter 2. We assume that the channel is known at the receiver but not at the transmitter. That is, there exists perfect *Channel State Information at the Receiver* (CSIR) and no CSIT. Therefore, the expression (3.9) is no longer valid since it assumes that the channel is perfectly known at both sides of the link. Thus, when the transmitter does not know the channel, the capacity is given by the average of the capacities obtained for each channel realization, so it depends on the channel distribution. In other words, the capacity is computed as the expectation of (3.9) over the channel $\boldsymbol{H}$, and is called ergodic capacity [2, 45, 54], i.e.

$$C = \mathrm{E}\left[\log_2 \det\left(\boldsymbol{H}\boldsymbol{C}_x\boldsymbol{H}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{\eta}}^{-1} + \mathbf{I}_R\right)\right]. \tag{3.10}$$

3.1.1. MIMO Broadcast Channel Capacity

In the previous section we introduce the concept of channel capacity for the MIMO single-user system model. The use of MIMO technologies at the *Base Station* (BS) results in an increase of degrees of freedom from having multiple antennas. Thus, it

is possible to simultaneously transmit or receive data from multiple users. Then, the previous definition of capacity makes no sense in such a scenario, where the channel has to be divided between all the users. The multiuser channel capacity is given by a capacity region [45, 54]. Such a region contains the set of vector rates that can be maintained simultaneously by all the users, providing a reliable communication. Since the users share the channel, if one of them desires to increase the rate, the others may reduce their respective rates, hence establishing a trade-off between all the users.

The achievable rate region has been found employing a technique named as *Dirty Paper Coding* (DPC) [6]. To explain how it works, let us introduce the permutation $\pi(\cdot)$ to order the users in the way that is more convenient for this technique. When we apply DPC, the user $\pi(2)$ presubstracts the interference of user $\pi(1)$. Analogously, the user $\pi(3)$ is able to cancel the interference of users $\pi(1)$ and $\pi(2)$, and so forth. Hence, the user $\pi(k)$ is affected from the interference of the users $\pi(j)$ such that $j > k$.

Let us now introduce an $N \times R$ MIMO BC system model where $\boldsymbol{H}_k$ and $\boldsymbol{C}_k$ stands for the channel and the covariance matrix of the transmitted signal for the user $k \in \{1, \cdots, K\}$, respectively. Note that the covariance matrix contains the data affected by the transformation applied prior to transmission, e.g., a multiplication times a precoding matrix. Thus, considering AWGN for all the users, we obtain the following achievable rate for the user $\pi(k)$ [5],

$$R_{\pi(k)} = \log_2 \left(\frac{\det \left(\mathbf{I}_R + \boldsymbol{H}_{\pi(k)} \sum_{i \geq k} \boldsymbol{C}_{\pi(i)} \boldsymbol{H}_{\pi(k)}^{\mathrm{H}} \right)}{\det \left(\mathbf{I}_R + \boldsymbol{H}_{\pi(k)} \sum_{i > k} \boldsymbol{C}_{\pi(i)} \boldsymbol{H}_{\pi(k)}^{\mathrm{H}} \right)} \right). \tag{3.11}$$

When we collect the rates of all the users in a vector, we obtain $\boldsymbol{R}(\pi, \boldsymbol{C})$, with $\boldsymbol{C} = \mathrm{blockdiag}(\boldsymbol{C}_1, \ldots, \boldsymbol{C}_K)$. Notice that the achievable rates depend on both the permutation and the covariance matrices contained in $\boldsymbol{C}$. Hence, the achievable capacity region $C_{\text{DPC}}(P_{tx}, \boldsymbol{H})$ is given by the convex hull of the union of the previously mentioned rate vectors over all the possible permutations and covariance matrices $\boldsymbol{C}$ fulfilling the power constraint $\mathrm{tr}(\boldsymbol{C}) \leq P_{tx}$, i.e.,

$$C_{\text{DPC}}(P_{tx}, \boldsymbol{H}) = Co \left(\bigcup_{\pi, \boldsymbol{C}} \left(\boldsymbol{R}\left(\pi, \boldsymbol{C}\right) \right) \right). \tag{3.12}$$

The dirty paper region has been shown to be equal to the capacity region for the MIMO BC [62]. Moreover, the dirty paper region is equivalent to the capacity region corresponding to the dual MIMO MAC, as shown in [5].

However, the DPC technique is a non-linear scheme and is only possible when assuming perfect knowledge of the channel at the transmitter to presubstract the inter-user interference. On the contrary, if only linear filtering is applied, the user k experiences

interference from all the other users $i \neq k$ and, therefore, the MIMO BC capacity with linear filtering is obtained as follows,

$$
\begin{aligned}
R_k &= \log_2 \left(\frac{\det\left(\mathbf{I}_R + \boldsymbol{H}_k \sum_{i=1}^{K} \boldsymbol{C}_i \boldsymbol{H}_k^{\mathrm{H}}\right)}{\det\left(\mathbf{I}_R + \boldsymbol{H}_k \sum_{i\neq k} \boldsymbol{C}_i \boldsymbol{H}_k^{\mathrm{H}}\right)} \right) \\
&= \log_2 \det \left(\left(\mathbf{I}_R + \boldsymbol{H}_k \sum_{i=1}^{K} \boldsymbol{C}_i \boldsymbol{H}_k^{\mathrm{H}}\right) \left(\mathbf{I}_R + \boldsymbol{H}_k \sum_{i\neq k} \boldsymbol{C}_i \boldsymbol{H}_k^{\mathrm{H}}\right)^{-1} \right) \\
&= \log_2 \det \left(\left(\mathbf{I}_R + \boldsymbol{H}_k \sum_{i=1}^{K} \boldsymbol{C}_i \boldsymbol{H}_k^{\mathrm{H}}\right) \left(\mathbf{I}_R + \boldsymbol{H}_k \sum_{i\neq k} \boldsymbol{C}_i \boldsymbol{H}_k^{\mathrm{H}}\right)^{-1} \right) \\
&= \log_2 \det \left(\left(\left(\boldsymbol{H}_k \boldsymbol{C}_k \boldsymbol{H}_k^{\mathrm{H}}\right) + \left(\mathbf{I}_R + \boldsymbol{H}_k \sum_{i\neq k} \boldsymbol{C}_i \boldsymbol{H}_k^{\mathrm{H}}\right) \right) \right. \\
&\quad \left. \times \left(\mathbf{I}_R + \boldsymbol{H}_k \sum_{i\neq k} \boldsymbol{C}_i \boldsymbol{H}_k^{\mathrm{H}}\right)^{-1} \right) \\
&= \log_2 \det \left(\mathbf{I}_R + \boldsymbol{H}_k \boldsymbol{C}_k \boldsymbol{H}_k^{\mathrm{H}} \left(\mathbf{I}_R + \boldsymbol{H}_k \sum_{i\neq k} \boldsymbol{C}_i \boldsymbol{H}_k^{\mathrm{H}}\right)^{-1} \right). \qquad (3.13)
\end{aligned}
$$

Analogously to the BC/MAC duality for the DPC region, any rate vector in the BC employing linear filters can be achieved in the dual MAC, as shown in [63].

Consider now random channels for all the users. In the single-user MIMO scenario with perfect CSIR, and no CSIT, the ergodic capacity is given by (3.10). The possible data rate for the user k is then

$$
\mathrm{E}[R_k] = \int_{\mathbb{C}} R_k(\boldsymbol{H}_k) f_{\boldsymbol{H}_k}(\boldsymbol{H}_k) d\boldsymbol{H}_k. \qquad (3.14)
$$

When there exist partial CSIT, the channel pdf is given by the following expression using Bayes' rule (see Section 2.4 of Chapter 2)

$$
f_{\boldsymbol{H}_k}(\boldsymbol{H}_k) = \int_{\mathcal{V}} f_{\boldsymbol{H}_k|v}(\boldsymbol{H}_k|v) f_v(v) dv, \qquad (3.15)
$$

where $v \in \mathcal{V}$ is the scalar random variable representing the information about the channel

known at the transmitter. Consequently, we rewrite the ergodic rate (3.14) as

$$\begin{aligned}\mathrm{E}[R_k] &= \int_{\mathbb{C}} R_k(\boldsymbol{H}_k) \int_{\mathcal{V}} f_{\boldsymbol{H}_k|v}(\boldsymbol{H}_k|v) f_v(v) dv d\boldsymbol{H}_k \\ &= \int_{\mathcal{V}} f_v(v) \underbrace{\left[R_k(\boldsymbol{H}_k) \int_{\mathbb{C}} f_{\boldsymbol{H}_k|v}(\boldsymbol{H}_k|v) d\boldsymbol{H}_k \right]}_{\mathrm{E}[R_k|v]} dv \\ &= \mathrm{E}[\mathrm{E}[R_k|v]]. \end{aligned} \tag{3.16}$$

Therefore, due to imperfect CSIT the ergodic rate depends not only on the channel statistics but also on the information that is fed back to the transmitter. Hence, there exist an average rate corresponding to every realization of the random variable v, and the ergodic rate is computed as the expectation over all the possible values of v.

3.2. Mean Square Error (MSE)

In statistics, the MSE of an estimator is a measure of how close is the estimator to the data. The difference between data and the corresponding estimator is referred to as the error. Hence, the MSE is the average of the squared errors. We study the BC system model that assumes linear filtering presented in Section 2.2.1 of Chapter 2, where $\boldsymbol{h}_k \in \mathbb{C}^N$, $\boldsymbol{p}_k \in \mathbb{C}^N$, $f_k \in \mathbb{C}$, $\eta_k \in \mathbb{C}$, and s_k are the *Multiple-Input Single-Output* (MISO) channel, the precoder, the receive filter, the noise, and the data for the user k, respectively, with N the number of transmit antennas. In such a scenario, the estimated symbol at the receiver k is given by $\hat{s}_k$, that is

$$\hat{s}_k = f_k^* \boldsymbol{h}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i s_i + f_k^* \eta_k. \tag{3.17}$$

Accordingly, the MSE for the user k is computed as follows

$$\mathrm{MSE}_k = \mathrm{E}\left[|s_k - \hat{s}_k|^2 \right]. \tag{3.18}$$

Substituting the estimated symbols $\hat{s}_k$, we obtain

$$\begin{aligned}
\mathrm{MSE}_k^{\mathrm{BC}} &= \mathrm{E}\left[\left|s_k - f_k^* \boldsymbol{h}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i s_i - f_k^* \eta_k\right|^2\right] \\
&= \mathrm{E}\left[\left(s_k - f_k^* \boldsymbol{h}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i s_i - f_k^* \eta_k\right)\left(s_k^* - \sum_{i=1}^{K} s_i^* \boldsymbol{p}_i^{\mathrm{H}} \boldsymbol{h}_k f_k - \eta_k^* f_k\right)\right] \\
&= \mathrm{E}\left[|s_k|^2\right] - \mathrm{E}\left[\sum_{i=1}^{K} s_k s_i^* \boldsymbol{p}_i^{\mathrm{H}} \boldsymbol{h}_k f_k\right] - \mathrm{E}\left[s_k \eta_k^* f_k\right] - \mathrm{E}\left[f_k^* \boldsymbol{h}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i s_i s_k^*\right] \\
&+ \mathrm{E}\left[|f_k|^2 \boldsymbol{h}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i s_i \sum_{j=1}^{K} s_j^* \boldsymbol{p}_j^{\mathrm{H}} \boldsymbol{h}_k\right] + \mathrm{E}\left[|f_k|^2 \boldsymbol{h}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i s_i \eta_k^*\right] - \mathrm{E}\left[f_k^* \eta_k s_k^*\right] \\
&+ \mathrm{E}\left[|f_k|^2 \sum_{i=1}^{K} \eta_k s_i^* \boldsymbol{p}_i^{\mathrm{H}} \boldsymbol{h}_k\right] + \mathrm{E}\left[|f_k|^2 |\eta_k|^2\right]. \qquad (3.19)
\end{aligned}$$

Considering a unit variance for the transmitted signal and $\sigma_{\eta_k}^2$ for the noise, respectively, and bearing in mind that the data for different users and the noise are statistically independent, i.e., $\mathrm{E}[s_k s_i^*] = 0 \, \forall k \neq i$ and $\mathrm{E}[s_k \eta_i^*] = 0$, the previous equation reduces to

$$\mathrm{MSE}_k^{\mathrm{BC}} = 1 - 2\Re\left\{f_k^* \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{p}_k\right\} + |f_k|^2 \boldsymbol{h}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i \boldsymbol{p}_i^{\mathrm{H}} \boldsymbol{h}_k + |f_k|^2 \sigma_{\eta_k}^2. \qquad (3.20)$$

Analogously, the previous derivation can be obtained for the MAC system model shown in Section 2.2 of Chapter 2, where $t_k \in \mathbb{C}$, $\boldsymbol{h}_k \in \mathbb{C}^N$, $\boldsymbol{g}_k \in \mathbb{C}^N$, and $\boldsymbol{\eta} \in \mathbb{C}^N$ are the MAC precoders, the channel, the receive filter, and the noise, respectively. Therefore, using the data estimates

$$\hat{s}_k^{\mathrm{MAC}} = \boldsymbol{g}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{h}_i t_i s_i + \boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{\eta}, \qquad (3.21)$$

we calculate the MSE for every user in the MAC as

$$\mathrm{MSE}_k^{\mathrm{MAC}} = 1 - 2\Re\left\{\boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{h}_k t_k\right\} + \boldsymbol{g}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{h}_i |t_i|^2 \boldsymbol{h}_i^{\mathrm{H}} \boldsymbol{g}_k + \boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{C}_\eta \boldsymbol{g}_k, \qquad (3.22)$$

with $\boldsymbol{C}_\eta$ being the covariance of the noise. Note that $\mathrm{MSE}^{\mathrm{BC}}$ and $\mathrm{MSE}^{\mathrm{MAC}}$ are used to denote the MSEs in the BC and the MAC, respectively.

So far, we have considered perfect CSIR and CSIT in the BC to compute the MSE. Consider now the random channels presented in Chapter 2. In addition, we assume perfect

CSIR but imperfect CSIT. Then, the MSE has to be computed as the average over the channel, i.e., $\mathrm{E}[\mathrm{MSE}_k]$.

In Section 3.1.1 the capacity for the MIMO BC with perfect CSIR and imperfect CSIT has been studied. In the case of MSE, the reasoning is analogous. We apply the Bayes' rule to obtain the average MSE as (cf. (3.16))

$$\mathrm{E}\left[\mathrm{MSE}_k\right] = \mathrm{E}\left[\mathrm{E}\left[\mathrm{MSE}_k|v\right]\right]. \tag{3.23}$$

In other words, optimizing the average MSE conditioned to v for every possible realization of the random variable results into the average MSE optimization.

3.3. Broadcast Channel / Multiple Access Channel Dualities

Over the last years, a large family of works have employed different dualities between the BC and the MAC in recent years. These dualities allow for a reformulation of the BC problem into the dual MAC, where some advantageous properties can be applied. Moreover, the figures of merit (e.g. rate, MSE, *Signal to Interference-plus-Noise Ratio* (SINR)) are equal in both the BC and the dual MAC, and the total transmit power is also the same in both domains.

Dualities with respect to the SINR [4, 7, 64], rate [5, 63], and MSE [8, 32] were presented in many previous works. Additionally, different types of dualities could be used depending on the problem formulation. In particular we focus on MSE dualities as done in [8], where three kinds of dualities are presented. The first kind preserves both the sum MSE and the total power constraint. For the second and third kinds before mentioned, every users' or stream-wise's MSEs remain unchanged while the same total power is achieved.

All these dualities are based not only on the assumption of perfect CSIR but also on perfect CSIT in the BC. Since we are interested in the BC, where the BS has partial knowledge of the *Channel State Information* (CSI), we focus on [37] instead. In this work we find formulas that allow to switch between the BC and the dual MAC keeping the per-user average MSEs unchanged, where the average is determined by the stochastic CSI available at the BS. Other dualities considering imperfect CSIT can be found in [15, 17, 31].

In the ensuing sections we will focus on the MSE dualities for both perfect and imperfect CSI.

3.3.1. MSE Duality with Perfect CSIR and CSIT

As previously mentioned, different kinds of dualities have been studied in [8]. For arbitrary filters, the equivalent ones in the dual domain are derived with different precision levels. For all of them, both the MSE of interest and the total power are preserved.

Sum-MSE Duality

We start with the less restrictive duality, in which the system sum-MSE remains the same. Simultaneously, it preserves the total power at transmission. In order to do that, only one degree of freedom is needed. The scalar $\alpha \in \mathbb{R}^+$ defines the relationship between the filters in the following way

$$\boldsymbol{p}_k = \alpha \boldsymbol{g}_k \quad \text{and} \quad f_k = \alpha^{-1} t_k, \tag{3.24}$$

where the filters correspond to the BC and MAC system models presented in Chapter 2. Moreover, the noise covariance in the BC is set to $\mathrm{E}[\eta_k \eta_k^*] = \sigma_\eta^2$, whereas the noise covariance in the MAC reads as $\boldsymbol{C}_{\boldsymbol{\eta}} = \mathrm{E}[\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{H}}] = \sigma_{\boldsymbol{\eta}}^2 \mathbf{I}_N$. Introducing (3.24) into (3.20) and equating to (3.22), we get

$$\sum_{i=1}^{K} \left| t_k^* \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{g}_i \right|^2 + \frac{\sigma_\eta^2}{\alpha^2} |t_k|^2 = \sum_{i=1}^{K} \left| \boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{h}_i t_i \right|^2 + \sigma_{\boldsymbol{\eta}}^2 \|\boldsymbol{g}_k\|_2^2 . \tag{3.25}$$

Since the equality holds for the sum-MSE, the last expression reduces to $\sum_{i=1}^{K} 1/\alpha^2 |t_k|^2 = \sum_{i=1}^{K} \|\boldsymbol{g}_k\|_2^2$. Thus, the value of α that satisfies the desired properties is found as

$$\alpha^2 = \frac{\sum_{i=1}^{K} |f_i|^2}{\sum_{i=1}^{K} \|\boldsymbol{g}_i\|_2^2}. \tag{3.26}$$

Per-user MSE duality

If we are interested in leaving the individual MSEs for all the users unchanged, we need different scalars α_k for each of them. This way the relationship between the filters is characterized as

$$\boldsymbol{p}_k = \alpha_k \boldsymbol{g}_k \quad \text{and} \quad f_k = \alpha_k^{-1} t_k. \tag{3.27}$$

Substituting $\boldsymbol{p}_k$ and $\boldsymbol{f}_k$ in the expression of the MSE for the BC, $\mathrm{MSE}_k^{\mathrm{BC}}$ of (3.20), and equating to (3.22), we obtain

$$\sum_{i=1}^{K} \left| \boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{h}_i t_i \right|^2 + \sigma_{\boldsymbol{\eta}}^2 \|\boldsymbol{g}_k\|_2^2 = \sum_{i=1}^{K} \frac{\alpha_i^2}{\alpha_k^2} \left| \boldsymbol{g}_i^{\mathrm{H}} \boldsymbol{h}_k t_k \right|^2 + \frac{\sigma_\eta^2}{\alpha_k^2} |t_k|^2 . \tag{3.28}$$

Multiplying by α_k^2 and moving to the right side all the terms except the one independent of α_j, $\forall\, j \in \{1, \ldots, K\}$, we get

$$\alpha_k^2 \left(\sum_{i \neq k} \left| \boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{h}_i t_i \right|^2 + \sigma_{\boldsymbol{\eta}}^2 \left\| \boldsymbol{g}_k \right\|_2^2 \right) - \sum_{i \neq k} \alpha_i^2 \left| \boldsymbol{g}_i^{\mathrm{H}} \boldsymbol{h}_k t_k \right|^2 = \sigma_\eta^2 \left| t_k \right|^2, \tag{3.29}$$

which is rewritten in matrix notation as

$$\boldsymbol{Z} \left[\alpha_1^2, \ldots, \alpha_K^2 \right]^{\mathrm{T}} = \sigma_\eta^2 \left[\left| t_1 \right|^2, \ldots, \left| t_K \right|^2 \right]^{\mathrm{T}}, \tag{3.30}$$

where the matrix $\boldsymbol{Z}$ is

$$[\boldsymbol{Z}]_{k,j} = \begin{cases} \sum_{i \neq k} \left| \boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{h}_i t_i \right|^2 + \sigma_{\boldsymbol{\eta}}^2 \| \boldsymbol{g}_k \|_2^2 & j = k, \\ - \left| \boldsymbol{g}_j^{\mathrm{H}} \boldsymbol{h}_k t_k \right|^2 & j \neq k. \end{cases} \tag{3.31}$$

The matrix $\boldsymbol{Z}$ is non-singular [8] (see also Section 3.3.2) and, as a consequence, the positive real-valued scalars α_k can always be found as

$$\left[\alpha_1^2, \ldots, \alpha_K^2 \right]^{\mathrm{T}} = \boldsymbol{Z}^{-1} \sigma_\eta^2 \left[\left| t_1 \right|^2, \ldots, \left| t_K \right|^2 \right]^{\mathrm{T}}. \tag{3.32}$$

Similarly, the conversion from BC to MAC can be derived. For given BC filters, the same average MSEs can be achieved using the following MAC relationship between the filters in both domains

$$\boldsymbol{g}_k = \beta_k^{-1} \boldsymbol{p}_k \quad t_k = \beta_k f_k. \tag{3.33}$$

After substituting (3.33) in the expression of the MSE for the MAC, i.e. $\mathrm{MSE}_k^{\mathrm{MAC}}$ from (3.22), we equate to (3.20) to get

$$\beta_k^2 \left(\sum_{i \neq k} \left| f_k^* \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{p}_i \right|^2 + \sigma_\eta^2 \left| f_k \right|^2 \right) = \sum_{i \neq k} \beta_i^2 \left| \boldsymbol{p}_k^{\mathrm{H}} \boldsymbol{h}_i f_i \right|^2 + \sigma_{\boldsymbol{\eta}}^2 \left\| \boldsymbol{p}_k \right\|_2^2, \tag{3.34}$$

after multiplying by β_k^2 both sides of the equality. Then, we follow the same steps as for the MAC to BC conversion (when the matrix $\boldsymbol{Z}$ was obtained), to find the matrix $\boldsymbol{W}$ that allows us to rewrite the equalities in matrix notation

$$\boldsymbol{W} \left[\beta_1^2, \ldots, \beta_K^2 \right]^{\mathrm{T}} = \sigma_{\boldsymbol{\eta}}^2 \left[\| \boldsymbol{p}_1 \|_2^2, \ldots, \| \boldsymbol{p}_K \|_2^2 \right]^{\mathrm{T}}. \tag{3.35}$$

Hence, such matrix $\boldsymbol{W}$ is computed as follows

$$[\boldsymbol{W}]_{k,j} = \begin{cases} \sum_{i \neq k} \left| f_k \right|^2 \left| \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{p}_i \right|^2 + \sigma_\eta^2 \left| f_k \right|^2 & j = k, \\ - \left| f_j \right|^2 \left| \boldsymbol{h}_j^{\mathrm{H}} \boldsymbol{p}_k \right|^2 & j \neq k. \end{cases} \tag{3.36}$$

Observe that $\boldsymbol{W}$ satisfies the same properties as $\boldsymbol{Z}$ from MAC to BC derivation. Thus, we have guaranteed that we are able to find positive real valued $\beta_k, \forall k$, such that $\forall k$: $\mathrm{MSE}_k^{\mathrm{MAC}} = \mathrm{MSE}_k^{\mathrm{BC}}$.

When the receivers have more than one antenna, which corresponds to MIMO scenarios, multiple streams can be allocated among the users taking advantage of the spatial multiplexing. In [8], another kind of duality is considered when the equivalence between the BC/MAC MSEs has to be fulfilled not only per-user but also per-stream. However, we will consider that case as a particular MIMO scenario, where each stream can be treated as a virtual user, bearing in mind that every stream of a real user propagates over the same channel.

Note that the per-stream duality implies that the per-user and the sum-MSE dualities have to be fulfilled. Therefore, the multiple-stream scenario could be seen as a more general case, whereas less degrees of freedom are needed for the first and second kinds.

3.3.2. MSE Duality with Perfect CSIR and Imperfect CSIT

In the previous section, some kinds of MSE dualities have been presented. However, all of them considered perfect CSI for both sides of the communication link. This assumption is rather unrealistic, so, for the BC system model, we consider the case where only partial information about the channel state is available at the BS, whereas the receivers perfectly know their own channel. Remember that in Section 3.2 the imperfect CSIT has been already considered (see (3.23)), and it has been characterized by the random variable v. However, in this section we will focus on the average MSE conditioned to v, i.e. $\overline{\mathrm{MSE}}_k^{\mathrm{BC}} = \mathrm{E}[\mathrm{MSE}_k^{\mathrm{BC}}|v]$.

Multiple Access Channel to Broadcast Channel

Here, the necessary coefficients for switching from the BC to the dual MAC for single-antenna receivers are computed. Contrarily to Section 3.3.1, where the noise variance is equal for all the users, we deal with the more general scenario where the noise in the dual MAC is set to $\boldsymbol{\eta} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_N)$ and the dual channel is defined as $\sigma_{\eta_k}^{-1}\boldsymbol{h}_k$, with $\sigma_{\eta_k}^2$ being the noise variance for the user k in the BC. Hence, considering the system models presented in Fig. 2.2 and Fig. 2.3 of Chapter 2, we define the relationship between the BC and the dual MAC filters as in [37], i.e.

$$\boldsymbol{p}_k = \alpha_k \boldsymbol{g}_k \quad \text{and} \quad f_k = \sigma_{\eta_k}^{-1}\alpha_k^{-1} t_k, \tag{3.37}$$

with $\alpha_k \in \mathbb{R}^+$. Then, we rewrite the average MSE in the BC conditioned to the imperfect CSIT v

$$\overline{\mathrm{MSE}}_k^{\mathrm{BC}} = \mathrm{E}\left[1 - 2\Re\{f_k^* \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{p}_k\} + |f_k|^2 \boldsymbol{h}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i \boldsymbol{p}_i^{\mathrm{H}} \boldsymbol{h}_k + |f_k|^2 \sigma_{\eta_k}^2 \,|\, v\right]. \tag{3.38}$$

Accordingly, this leads us to

$$\overline{\mathrm{MSE}}_k^{\mathrm{BC}} = \mathrm{E}\left[1 - 2\Re\left\{\sigma_{\eta_k}^{-1} t_k^* \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{g}_k\right\} + \alpha_k^{-2} |t_k|^2 + \sum_{i=1}^{K} \frac{\alpha_i^2}{\alpha_k^2} \sigma_{\eta_k}^{-2} \left|\boldsymbol{g}_i^{\mathrm{H}} \boldsymbol{h}_k t_k\right|^2 \,|\, v\right]. \tag{3.39}$$

By equating the last expression to the average MSE in the dual MAC conditioned to v, i.e.

$$\overline{\mathrm{MSE}}_k^{\mathrm{MAC}} = \mathrm{E}\left[1 - 2\Re\{\sigma_{\eta_k}^{-1} \boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{h}_k t_k\} + \boldsymbol{g}_k^{\mathrm{H}} \sum_{i=1}^{K} \sigma_{\eta_i}^{-2} \boldsymbol{h}_i |t_i|^2 \boldsymbol{h}_i^{\mathrm{H}} \boldsymbol{g}_k + \|\boldsymbol{g}_k\|_2^2 \,|\, v\right], \tag{3.40}$$

we get

$$\mathrm{E}\left[|t_k|^2 \,|\, v\right] + \mathrm{E}\left[\sum_{i=1}^{K} \alpha_i^2 \sigma_{\eta_k}^{-2} \left|\boldsymbol{g}_i^{\mathrm{H}} \boldsymbol{h}_k t_k\right|^2 \,|\, v\right] = \alpha_k^2 \|\boldsymbol{g}_k\|_2^2 + \alpha_k^2 \,\mathrm{E}\left[\sum_{i=1}^{K} \sigma_{\eta_i}^{-2} |\boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{h}_i t_i|^2 |\, v\right], \tag{3.41}$$

after multiplying by α_k^2 both equation sides. Now we simplify and rewrite (3.41) in matrix notation. To that end, we collect the α_k factors in the vector

$$\boldsymbol{a} = \left[\alpha_1^2, \ldots, \alpha_K^2\right]^{\mathrm{T}}, \tag{3.42}$$

and the MAC average powers $\mathrm{E}[|t_k|^2|\, v]$ in

$$\boldsymbol{\varsigma} = \left[\mathrm{E}\left[|t_1|^2 \,|\, v\right], \cdots, \mathrm{E}\left[|t_K|^2 \,|\, v\right]\right]. \tag{3.43}$$

Finally, defining $\boldsymbol{\Gamma} \in \mathbb{R}^{K \times K}$ as follows

$$[\boldsymbol{\Gamma}]_{k,j} = \begin{cases} \sum_{i \neq k} \sigma_{\eta_i}^{-2} \,\mathrm{E}[|\boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{h}_i t_i|^2 \,|\, v] + \|\boldsymbol{g}_k\|_2^2 & j = k, \\ -\sigma_{\eta_k}^{-2} \,\mathrm{E}[|\boldsymbol{g}_j^{\mathrm{H}} \boldsymbol{h}_k t_k|^2 \,|\, v] & j \neq k, \end{cases} \tag{3.44}$$

we get the following K equalities expressed in matrix notation

$$\boldsymbol{\Gamma} \boldsymbol{a} = \boldsymbol{\varsigma}. \tag{3.45}$$

Since $\boldsymbol{\Gamma}$ is diagonally dominant, i.e. $|[\boldsymbol{\Gamma}]_{i,i}| \geq \sum_{j \neq i} |[\boldsymbol{\Gamma}]_{i,j}|, \forall i \in \{1, \ldots, K\}$, it is non-singular. Additionally, $\boldsymbol{\Gamma}$ has positive diagonal and non-positive off-diagonal entries. Thus, $\boldsymbol{\Gamma}^{-1}$ has non-negative entries [8, 65] and the resulting α_k^2 are non-negative. In other words, we can always find $\alpha_k \in \mathbb{R}^+$ such that $\overline{\mathrm{MSE}}_k^{\mathrm{BC}} = \overline{\mathrm{MSE}}_k^{\mathrm{MAC}}$, $\forall k$. Note that $\sum_{i=1}^{K} \|\boldsymbol{g}_i\|_2^2 \alpha_i^2 = \sum_{i=1}^{K} \mathrm{E}[|t_i|^2 \,|\, v]$, which results from left multiplying (3.45) by the all-ones vector $\mathbf{1}^{\mathrm{T}}$. Therefore, we can infer that the same average transmit power is used in the BC as in the dual MAC, using the relationship in (3.37).

Broadcast Channel to Multiple Access Channel

We now focus on the conversion from MAC to BC. For given BC filters, MAC filters achieving the same average MSEs with the same transmit power can be found [37]. The duality can be obtained in the same way as for the MAC to BC conversion. First of all, we define the relationships

$$\boldsymbol{g}_k = \beta_k^{-1}\boldsymbol{p}_k \quad \text{and} \quad t_k = \sigma_{\eta_k}\beta_k f_k. \tag{3.46}$$

After substituting (3.46) in the average MSE expression for the dual MAC, i.e. $\overline{\text{MSE}}_k^{\text{MAC}}$ from (3.40), we obtain

$$\overline{\text{MSE}}_k^{\text{MAC}} = \text{E}\left[1 - 2\Re\{\boldsymbol{p}_k^{\text{H}}\boldsymbol{h}_k f_k\} + \sum_{i=1}^{K}\frac{\beta_i^2}{\beta_k^2}\left|\boldsymbol{p}_k^{\text{H}}\boldsymbol{h}_i f_i\right|^2 + \beta_k^{-2}\left\|\boldsymbol{p}_k\right\|_2^2 \mid v\right]. \tag{3.47}$$

Then, equating to $\overline{\text{MSE}}_k^{\text{BC}}$ and multiplying by β_k^2, we get

$$\beta_k^2\,\text{E}\left[\sum_{i=1}^{K}|f_k^*\boldsymbol{h}_k^{\text{H}}\boldsymbol{p}_i|^2 \mid v\right] + \beta_k^2\sigma_{\eta_k}^2\,\text{E}\left[|f_k|^2 \mid v\right] = \text{E}\left[\sum_{i=1}^{K}\beta_i^2\left|\boldsymbol{p}_k^{\text{H}}\boldsymbol{h}_i f_i\right|^2 \mid v\right] + \left\|\boldsymbol{p}_k\right\|_2^2. \tag{3.48}$$

Following the procedure presented for the opposite conversion, we obtain the following equalities in matrix notation

$$\boldsymbol{\Omega}\boldsymbol{b} = \boldsymbol{\tau}, \tag{3.49}$$

where

$$\boldsymbol{b} = \left[\beta_1^2, \ldots, \beta_K^2\right]^{\text{T}}, \tag{3.50}$$

and

$$\boldsymbol{\tau} = \left[\left\|\boldsymbol{p}_1\right\|_2^2, \ldots, \left\|\boldsymbol{p}_K\right\|_2^2\right]^{\text{T}}. \tag{3.51}$$

Thus, the entries of $\boldsymbol{\Omega}$ are given by

$$[\boldsymbol{\Omega}]_{k,j} = \begin{cases} \sum_{i\neq k}\text{E}\left[|f_k|^2\left|\boldsymbol{h}_k^{\text{H}}\boldsymbol{p}_i\right|^2 \mid v\right] + \sigma_{\eta_k}^2\,\text{E}\left[|f_k|^2 \mid v\right] & j = k, \\ -\,\text{E}\left[|f_j|^2\left|\boldsymbol{h}_j^{\text{H}}\boldsymbol{p}_k\right|^2 \mid v\right] & j \neq k. \end{cases} \tag{3.52}$$

Note that $\boldsymbol{\Omega}$ satisfies the same properties as $\boldsymbol{\Gamma}$ from previous derivation. As a consequence, it is always possible to find positive real valued $\beta_k, \forall k$, such that $\overline{\text{MSE}}_k^{\text{MAC}} = \overline{\text{MSE}}_k^{\text{BC}}, \forall k$, with the same total transmit power for both the BC and the MAC.

3.4. Conclusions

In this chapter we have presented some concepts fundamental for the understanding of the rest of work presented in this dissertation. The channel capacity was introduced as the theoretical limit for the rate transmission that can be achieved with asymptotically small error probability. Moreover, the capacity of the multiuser systems was discussed considering both linear and non-linear transmit strategies in the BC. Another key point is the CSI available at the transmitter. If only partial CSIT is acquired, the ergodic rate turns to be the adequate metric.

A different performance measure is also introduced in this chapter, that is, the MSE. It consists on evaluating the error incurred in the estimated signal with respect to the original one. Such metric is widely used in the literature of wireless communications. Again, we distinguish between scenarios where the CSIT is perfectly known and the ones where only imperfect information is available.

The MSE duality has been also presented in this chapter as a powerful tool to circumvent the difficulty of jointly determining the optimal precoders in the BC. Instead, the corresponding MAC receive filters can be easily obtained due to the proposed dualities. Furthermore, several kinds of dualities have been studied, from those that preserve the per-user MSE, or the per-stream MSE, to those oriented to the sum-MSE. Finally, imperfect CSIT and perfect CSIR assumptions have been included in the per-user MSE duality.

Chapter 4

Transmit Power Minimization in MISO Broadcast Channels

In this chapter we focus in the *Multiple-Input Single-Output* (MISO) *Broadcast Channel* (BC) where data streams are transmitted from a multi-antenna base station to several non-cooperative single-antenna receivers. At the base station, data is precoded with linear filters that are designed to accomplish individual rate requirements while minimizing the total transmit power. Perfect *Channel State Information* (CSI) is typically assumed when designing such linear filters. In practical settings, however, CSI is obtained via a limited feedback channel or estimation in the reverse link. Hence, only imperfect CSI is actually available at the base station (cf. Section 2.4 of Chapter 2).

The minimization of the transmit power subject to *Quality-of-Service* (QoS) constraints for imperfect CSI scenarios is a difficult problem. Exploiting the relationship between *Minimum Mean Square Error* (MMSE) and rate, we approximate the original formulation to be conveniently addressed. Based on the appropriate duality between the *Multiple Access Channel* (MAC) and the BC for the *Mean Square Error* (MSE), the BC filters design problem is reformulated in the dual MAC. Due to the assumption of erroneous CSI, however, the dualities presented in [5, 7, 8, 32, 63, 66] cannot be applied. Instead, we have to resort to the duality shown in [37] which allows different levels of CSI at both the transmitter and the receivers.

We employ a stochastic error model, e.g., resulting from estimation in the reverse link or feedback, and a formulation based on ergodic rates as in [27], where bounds to the achievable rates for linear zero-forcing precoders based on imperfect CSI were presented. In [10, 22, 29, 67], however, the precoder design was based on a model with bounded errors, which is well suited in feedback systems. For a stochastic error model, the average sum MSE has been minimized in several works, e.g. [15, 37]. The precoder design under probabilistic constraints has been also considered in [9, 21, 68]. Additionally, we do not apply the assumption that the CSI errors are identical at both sides of the communication

link, as done in [15].

Our contribution is an algorithmic solution to the transmit power minimization problem under the assumption of imperfect *Channel State Information at the Transmitter* (CSIT) exploiting the duality result of [37]. In particular, we highlight the possibility of using a standard *Interference Function* (IF) [38] based on the MMSE resulting from applying scalar receivers in the MISO MAC, which leads to a low complexity of the fixed-point iteration used to allocate the power.

In the ensuing sections both perfect and imperfect CSIT scenarios are studied. An algorithmic solution is proposed for the perfect CSIT case, and it is then extended for the more realistic scenario where the assumption of partial CSIT is available.

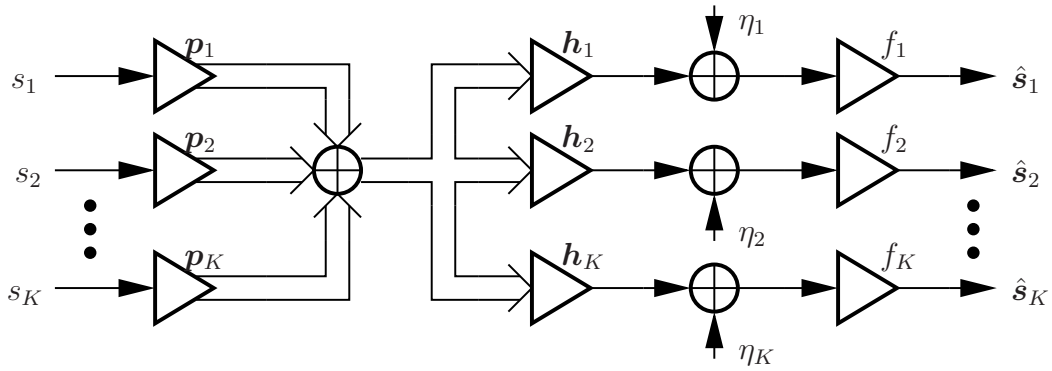


Figure 4.1: MISO Broadcast Channel.

4.1. Scenario 1: Perfect Channel State Information at the Transmitter

Figure 2.2 in Chapter 2 depicts the BC model to be considered along this work. The BC channel in our particular scenario is that shown in Fig. 4.1. We assume a Gaussian zero-mean data signal $s_k \in \mathbb{C}$ for user k, with $1 \leq k \leq K$ and $\mathrm{E}[|s_k|^2] = 1$, is precoded by $\boldsymbol{p}_k \in \mathbb{C}^N$, where K and N are the number of users and transmit antennas, respectively. The transmit signal propagates over a MISO channel $\boldsymbol{h}_k \in \mathbb{C}^N$. The *Additive White Gaussian Noise* (AWGN) is $\eta_k \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_{\eta_k}^2)$. The data signals are mutually independent and also independent from the noise. The linear receive filter $f_k \in \mathbb{C}$ provides the following estimates of the data symbols

$$\hat{s}_k = f_k^* \boldsymbol{h}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i s_i + f_k^* \eta_k. \tag{4.1}$$

In this section, we consider perfect CSIT and perfect *Channel State Information at the Receiver* (CSIR). The rate for the user k is particularized from the *Multiple-Input Multiple-Output* (MIMO) BC in (3.13), with the number of receive antennas R equal to one and the covariance of the transmitted signal as $\boldsymbol{C}_k = \mathrm{E}[\boldsymbol{p}_k s_k s_k^* \boldsymbol{p}_k^{\mathrm{H}}] = \boldsymbol{p}_k \mathrm{E}[s_k s_k^*] \boldsymbol{p}_k^{\mathrm{H}} = \boldsymbol{p}_k \boldsymbol{p}_k^{\mathrm{H}}$ i.e.,

$$R_k = \log_2 \left(1 + \boldsymbol{p}_k^{\mathrm{H}} \boldsymbol{h}_k \left(\sigma_k^2 + \boldsymbol{h}_k^{\mathrm{H}} \sum_{i \neq k} \boldsymbol{p}_i \boldsymbol{p}_i^{\mathrm{H}} \boldsymbol{h}_k \right)^{-1} \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{p}_k \right). \tag{4.2}$$

Our aim is to minimize the total transmit power $P_T = \sum_{i=1}^{K} ||\boldsymbol{p}_i||_2^2$ necessary to guarantee certain QoS constraints, that is, a given rate for each user, denoted by ρ_k, $k \in \{1, \ldots, K\}$. That leads us to the following optimization problem

$$\min_{\{\boldsymbol{p}_k\}_{k=1}^K} P_T = \sum_{i=1}^{K} \|\boldsymbol{p}_i\|_2^2 \quad \text{subject to} \quad R_k \geq \rho_k, \forall k. \tag{4.3}$$

4.1.1. Problem Formulation

Consider now the performance measure given by $\mathrm{MSE}_k = \mathrm{E}\left[|s_k - \hat{s}_k|^2\right]$, as presented in (3.20) of Chapter 3

$$\mathrm{MSE}_k = 1 - 2\Re\left\{ f_k^* \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{p}_k \right\} + |f_k|^2 \boldsymbol{h}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i \boldsymbol{p}_i^{\mathrm{H}} \boldsymbol{h}_k + |f_k|^2 \sigma_{\eta_k}^2. \tag{4.4}$$

Recall that the transmitter and the receiver have full CSI, i.e., perfect CSIT and CSIR, respectively. Hence, any meaningful MSE receivers f_k are functions of the channel state

$$\begin{aligned} f_k^{\mathrm{MMSE}} &= \mathrm{argmin}_{f_k} \mathrm{E}\left[|s_k - \hat{s}_k|^2\right] \\ &= \mathrm{argmin}_{f_k} \mathrm{E}\left[\left| s_k - f_k^* \boldsymbol{h}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i s_i + f_k^* \eta_k \right|^2 \right]. \end{aligned} \tag{4.5}$$

To compute the MMSE receivers f_k^{MMSE}, we calculate the derivative of MSE_k in (4.4) with respect to f_k^* i.e.,

$$\frac{\partial \mathrm{MSE}_k}{\partial f_k^*} = -\boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{p}_k + f_k \left(\boldsymbol{h}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i \boldsymbol{p}_i^{\mathrm{H}} \boldsymbol{h}_k + \sigma_{\eta_k}^2 \right). \tag{4.6}$$

Equating the last expression to zero, we get

$$f_k^{\mathrm{MMSE}} = \left(\boldsymbol{h}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i \boldsymbol{p}_i^{\mathrm{H}} \boldsymbol{h}_k + \sigma_{\eta_k}^2 \right)^{-1} \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{p}_k. \tag{4.7}$$

Now, substituting (4.7) into the MSE expression of (4.4) we get the following MMSE for user k

$$\begin{aligned}\text{MMSE}_k &= 1 - 2\boldsymbol{p}_k^{\text{H}}\boldsymbol{h}_k \left(\boldsymbol{h}_k^{\text{H}} \sum_{i=1}^{K} \boldsymbol{p}_i \boldsymbol{p}_i^{\text{H}} \boldsymbol{h}_k + \sigma_{\eta_k}^2\right)^{-1} \boldsymbol{h}_k^{\text{H}}\boldsymbol{p}_k \\ &\quad + \boldsymbol{p}_k^{\text{H}}\boldsymbol{h}_k \left(\boldsymbol{h}_k^{\text{H}} \sum_{i=1}^{K} \boldsymbol{p}_i \boldsymbol{p}_i^{\text{H}} \boldsymbol{h}_k + \sigma_{\eta_k}^2\right)^{-1} \left(\boldsymbol{h}_k^{\text{H}} \sum_{i=1}^{K} \boldsymbol{p}_i \boldsymbol{p}_i^{\text{H}} \boldsymbol{h}_k + \sigma_{\eta_k}^2\right) \\ &\quad \left(\boldsymbol{h}_k^{\text{H}} \sum_{i=1}^{K} \boldsymbol{p}_i \boldsymbol{p}_i^{\text{H}} \boldsymbol{h}_k + \sigma_{\eta_k}^2\right)^{-1} \boldsymbol{h}_k^{\text{H}}\boldsymbol{p}_k \\ &= 1 - f_k^{*,\text{MMSE}} \boldsymbol{h}_k^{\text{H}}\boldsymbol{p}_k \\ &= 1 - \boldsymbol{p}_k^{\text{H}}\boldsymbol{h}_k \left(\boldsymbol{h}_k^{\text{H}} \sum_{i=1}^{K} \boldsymbol{p}_i \boldsymbol{p}_i^{\text{H}} \boldsymbol{h}_k + \sigma_{\eta_k}^2\right)^{-1} \boldsymbol{h}_k^{\text{H}}\boldsymbol{p}_k. \end{aligned} \tag{4.8}$$

Notice that it is possible to find a relationship between the rate in (4.2) and the MMSE in (4.8). Rewriting (4.8) using the equality $1 - \frac{a}{a+b} = (1 + \frac{a}{b})^{-1}$, we get

$$\text{MMSE}_k = \left(1 + \boldsymbol{p}_k^{\text{H}}\boldsymbol{h}_k \left(\boldsymbol{h}_k^{\text{H}} \sum_{i \neq k} \boldsymbol{p}_i \boldsymbol{p}_i^{\text{H}} \boldsymbol{h}_k + \sigma_{\eta_k}^2\right)^{-1} \boldsymbol{h}_k^{\text{H}}\boldsymbol{p}_k\right)^{-1}, \tag{4.9}$$

and then the relationship with the rate given by (4.2) can be easily obtained as

$$R_k = \log_2\left(\text{MMSE}_k^{-1}\right). \tag{4.10}$$

Since $\log_2(\cdot)$ is a monotonic increasing function, reducing the MMSE_k of the user k leads to increasing the corresponding rate R_k. Contrarily, any increase in the MMSE_k causes a reduction in the rate R_k of the corresponding user. Due to that, we focus on MMSE instead of rate in our problem formulation: the optimization problem presented in (4.3) can be reformulated based on the MMSE equivalent constraints, such that $\text{MMSE}_k \leq \varepsilon_k, \forall k$. To fulfill the original QoS constraints, we compute the corresponding MMSE restrictions $\varepsilon_k = 2^{-\rho_k}$. In that way, our new problem formulation equivalent to that presented in (4.3) reads as

$$\min_{\{\boldsymbol{p}_k, f_k\}_{k=1}^{K}} P_T = \sum_{i=1}^{K} \|\boldsymbol{p}_i\|_2^2 \quad \text{subject to} \quad \text{MMSE}_k \geq 2^{-\rho_k}, \forall k. \tag{4.11}$$

4.1.2. Exploiting the MSE Duality

In (4.7) we have derived the receivers that minimize the MMSE for the user k. Note that these receivers can be individually optimized since f_k does not have any impact in

the MSE of the other users. Contrary to that, the precoder $\boldsymbol{p}_k$ has influence on the MSE_i, when $i \neq k$.

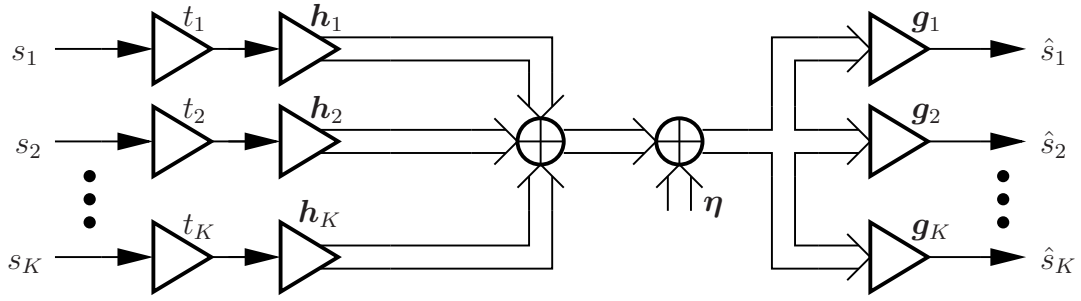


Figure 4.2: MISO Multiple Access Channel.

Consider now the MAC system model shown in Fig. 2.3, which is presented in more detail in Fig. 4.2. In such a scenario, we set the dual MAC of the BC with $t_k \in \mathbb{C}$, $\sigma_{\eta_k}^{-1}\boldsymbol{h_k} \in \mathbb{C}^N$, $\boldsymbol{\eta} \in \mathbb{C}^N$, and $\boldsymbol{g}_k \in \mathbb{C}^N$ being the precoders, the dual channel, the AWGN, and the receivers, respectively. The MSE in the MAC has been computed in (3.22), and it reads in the dual MAC as

$$\text{MSE}_k^{\text{MAC}} = 1 - 2\Re\left\{\boldsymbol{g}_k^{\text{H}}\boldsymbol{\theta}_k t_k\right\} + \boldsymbol{g}_k^{\text{H}} \sum_{i=1}^{K} \boldsymbol{\theta}_i \left|t_i\right|^2 \boldsymbol{\theta}_i^{\text{H}} \boldsymbol{g}_k + \boldsymbol{g}_k^{\text{H}} \boldsymbol{g}_k, \tag{4.12}$$

where $\boldsymbol{\theta}_k$ is $\sigma_{\eta_k}^{-1}\boldsymbol{h}_k$ and $\eta \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_N)$. It is important to note that in (4.12) the receive filters $\boldsymbol{g}_k$ can be individually optimized, whereas the precoders t_k are coupled and any change in one of them has impact on the MSE of all the users. Therefore, we employ the MSE dualities presented in Section 3.3 of Chapter 3 to find the optimal BC receive filters f_k in the BC domain, whereas the BC precoders are computed in the dual MAC as the optimum $\boldsymbol{g}_k$. In that way, we avoid the difficult problem of finding the transmit and the receive filters in the domain where they are coupled.

In particular we are interested in preserving the per-user MSEs. Therefore, we employ the second kind of duality explained on Section 3.3.1 of Chapter 3, which was based on [8], but including the noise variance, i.e.

$$\boldsymbol{p}_k = \alpha_k \boldsymbol{g}_k \quad \text{and} \quad f_k = \sigma_{\eta_k}^{-1} \alpha_k^{-1} t_k. \tag{4.13}$$

Recall that it is always possible to find $\alpha_k \in \mathbb{R}^+$ such that $\text{MSE}_k^{\text{MAC}} = \text{MSE}_k^{\text{BC}}$ employing the same total average transmit power. Likewise, it is possible to obtain the MAC filters from the BC ones in such a way that the per-user MSE and the total transmit power remains unchanged. This relationship is given by

$$\boldsymbol{g}_k = \beta_k^{-1} \boldsymbol{p}_k \quad t_k = \sigma_k \beta_{\eta_k} f_k, \tag{4.14}$$

as we previously showed in Section 3.3.1 of Chapter 3. Thus, for the computation of the optimal precoders, a reformulation in the dual MAC is necessary. The MSE duality guarantees that the solution in one domain can be achieved in the dual one. Therefore, we equivalently study the optimization problem of (4.11) in the dual MAC

$$\min_{\{t_k,\,\boldsymbol{g}_k\}_{k=1}^{K}} P_T^{\text{MAC}} = \sum_{i=1}^{K} |t_i|^2 \qquad \text{subject to} \quad \text{MMSE}_k^{\text{MAC}} \leq 2^{-\rho_k}. \tag{4.15}$$

From duality it is straightforward to see that the same total transmit power is reached in both domains, i.e., $P_T = P_T^{\text{MAC}}$.

Hence, the optimal MAC receivers, that is, the BC precoders, are computed as the MMSE receive filters $\boldsymbol{g}_k^{\text{MMSE}}$ deriving (4.12) with respect to $\boldsymbol{g}_k^*$ as follows (cf. matrix derivatives in Section C.6 of Appendix C)

$$\frac{\partial \text{MSE}_k^{\text{MAC}}}{\partial \boldsymbol{g}_k^*} = -\boldsymbol{\theta}_k t_k + \left(\sum_{i=1}^{K} \boldsymbol{\theta}_i \left| t_i \right|^2 \boldsymbol{\theta}_i^{\text{H}} + \mathbf{I}_N \right) \boldsymbol{g}_k. \tag{4.16}$$

Observe that since the $\text{MSE}_k^{\text{MAC}}$ is scalar, we can equivalently calculate the derivative of $\text{tr}(\text{MSE}_k^{\text{MAC}})$. Now, equating the derivative to zero we get the optimal MAC receive filter

$$\boldsymbol{g}_k^{\text{MMSE}} = \left(\sum_{i=1}^{K} \boldsymbol{\theta}_i \left| t_i \right|^2 \boldsymbol{\theta}_i^{\text{H}} + \mathbf{I}_N \right)^{-1} \boldsymbol{\theta}_k t_k. \tag{4.17}$$

Taking advantage of the mentioned duality, we compute the BC precoders as the MAC receivers using (4.17) and the BC receivers are those from (4.7). Observe that when an update of the BC precoders $\boldsymbol{p}_k$ is performed, the receiver filters f_k are modified accordingly by means of (4.7). Likewise, updating the MAC precoders t_k affects the MAC filters $\boldsymbol{g}_k$ (cf. (4.17)). Thus, we propose to find the optimal transmit and receive filters following an iterative process where both the transmit and receive filters are updated in an alternate manner, which is named as *Alternate Optimization* (AO) approach, as we will see in Section 4.1.4.

4.1.3. Power Allocation using Interference Functions

So far we have shown how to compute the optimal transmit and receive filters using the AO procedure aforementioned, taking advantage of the MSE BC to MAC and MAC to BC dualities. However, an additional issue arises in the Multiuser systems, the power allocation. That is, how we distribute a certain amount of power between the users. Note that such decision has a major impact in the system performance, since the interference suffered for each user depends on the power assigned to the rest of them. Hence, the

power allocation has to be updated to fulfill the QoS constraints. Therefore, we split off the power allocation in the MAC $\xi_k = |t_k|^2$, i.e., $t_k = \sqrt{\xi_k}\tau_k$, with $|\tau_k|^2 = 1$. Accordingly, the MAC MSE reads as (cf. (4.12))

$$\mathrm{MSE}_k^{\mathrm{MAC}} = 1 - 2\sqrt{\xi_k}\Re\{\boldsymbol{g}_k^{\mathrm{H}}\boldsymbol{\theta}_k\tau_k\} + \sum_{i=1}^{K}\xi_i\left|\tau_i\right|^2\boldsymbol{g}_k^{\mathrm{H}}\boldsymbol{\theta}_i\boldsymbol{\theta}_i^{\mathrm{H}}\boldsymbol{g}_k + \|\boldsymbol{g}_k\|_2^2\,. \tag{4.18}$$

The optimal receivers $\boldsymbol{g}_k^{\mathrm{MMSE}}$ still have the form given by (4.17) but using the power normalization previously introduced in this section the can be rewritten as

$$\boldsymbol{g}_k^{\mathrm{MMSE}} = \left(\sum_{i=1}^{K}\xi_i\boldsymbol{\theta}_i\left|\tau_i\right|^2\boldsymbol{\theta}_i^{\mathrm{H}} + \mathbf{I}_N\right)^{-1}\boldsymbol{\theta}_k\tau_k\sqrt{\xi_k}. \tag{4.19}$$

Substituting (4.19) into (4.18) provides the following MMSE expression

$$\begin{aligned}
\mathrm{MMSE}_k^{\mathrm{MAC}} = 1 &- 2\xi_k\left|\tau_k\right|^2\boldsymbol{\theta}_k^{\mathrm{H}}\left(\sum_{i=1}^{K}\xi_i\boldsymbol{\theta}_i\left|\tau_i\right|^2\boldsymbol{\theta}_i^{\mathrm{H}} + \mathbf{I}_N\right)^{-1}\boldsymbol{\theta}_k + \\
&\xi_k\left|\tau_k\right|^2\boldsymbol{\theta}_k^{\mathrm{H}}\left(\sum_{i=1}^{K}\xi_i\boldsymbol{\theta}_i\left|\tau_i\right|^2\boldsymbol{\theta}_i^{\mathrm{H}} + \mathbf{I}_N\right)^{-1}\left(\sum_{i=1}^{K}\xi_i\boldsymbol{\theta}_i\left|\tau_i\right|^2\boldsymbol{\theta}_i^{\mathrm{H}} + \mathbf{I}_N\right) \\
&\left(\sum_{i=1}^{K}\xi_i\boldsymbol{\theta}_i\left|\tau_i\right|^2\boldsymbol{\theta}_i^{\mathrm{H}} + \mathbf{I}_N\right)^{-1}\boldsymbol{\theta}_k \\
&= 1 - \boldsymbol{g}_k^{\mathrm{MMSE,H}}\boldsymbol{\theta}_k\tau_k\sqrt{\xi_k} \\
&= 1 - \xi_k\left|\tau_k\right|^2\boldsymbol{\theta}_k^{\mathrm{H}}\left(\sum_{i=1}^{K}\xi_i\boldsymbol{\theta}_i\left|\tau_i\right|^2\boldsymbol{\theta}_i^{\mathrm{H}} + \mathbf{I}_N\right)^{-1}\boldsymbol{\theta}_k.
\end{aligned} \tag{4.20}$$

Note that this expression depends on the normalized precoders τ_k and the power allocation $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_K]^{\mathrm{T}} \geq \mathbf{0}$.

To find the minimum power necessary to fulfill the QoS constraints, we resort to the *standard* interference function framework proposed by [38] (see Appendix B). This framework is useful to provide a solution to power allocation problems where there exist interference between the different users.

Using $\boldsymbol{\varphi}_k = \sqrt{\xi}_k \tau_k \boldsymbol{\theta}_k$ and $\boldsymbol{X}_k = \sum_{i \neq k} \xi_i \boldsymbol{\theta}_i |\tau_i|^2 \boldsymbol{\theta}_i^{\mathrm{H}} + \mathbf{I}_N$ we rewrite (4.20) as follows

$$\begin{aligned}
\mathrm{MMSE}_k^{\mathrm{MAC}} &= 1 - \boldsymbol{\varphi}_k^{\mathrm{H}} \left(\boldsymbol{X}_k + \boldsymbol{\varphi}_k \boldsymbol{\varphi}_k^{\mathrm{H}} \right)^{-1} \boldsymbol{\varphi}_k && (4.21)\\
&= 1 - \boldsymbol{\varphi}_k^{\mathrm{H}} \left(\boldsymbol{X}_k^{-1} - \boldsymbol{X}_k^{-1} \boldsymbol{\varphi}_k \left(1 + \boldsymbol{\varphi}_k^{\mathrm{H}} \boldsymbol{X}_k^{-1} \boldsymbol{\varphi}_k \right)^{-1} \boldsymbol{\varphi}_k^{\mathrm{H}} \boldsymbol{X}_k^{-1} \right) \boldsymbol{\varphi}_k && (4.22)\\
&= 1 - \left(\boldsymbol{\varphi}_k^{\mathrm{H}} \boldsymbol{X}_k^{-1} \boldsymbol{\varphi}_k - \left(\boldsymbol{\varphi}_k^{\mathrm{H}} \boldsymbol{X}_k^{-1} \boldsymbol{\varphi}_k \right)^2 \left(1 + \boldsymbol{\varphi}_k^{\mathrm{H}} \boldsymbol{X}_k^{-1} \boldsymbol{\varphi}_k \right)^{-1} \right) \\
&= 1 - \left(\boldsymbol{\varphi}_k^{\mathrm{H}} \boldsymbol{X}_k^{-1} \boldsymbol{\varphi}_k \left(1 + \boldsymbol{\varphi}_k^{\mathrm{H}} \boldsymbol{X}_k^{-1} \boldsymbol{\varphi}_k \right)^{-1} \right) \\
&= \left(1 + \boldsymbol{\varphi}_k^{\mathrm{H}} \boldsymbol{X}_k^{-1} \boldsymbol{\varphi}_k \right)^{-1} \\
&= \frac{1}{\xi_k} \left(\frac{1}{\xi}_k + |\tau_k|^2 \boldsymbol{\theta}_k^{\mathrm{H}} \left(\sum_{i \neq k} \xi_i \boldsymbol{\theta}_i |\tau_i|^2 \boldsymbol{\theta}_i^{\mathrm{H}} + \mathbf{I}_N \right)^{-1} \boldsymbol{\theta}_k \right)^{-1}. && (4.23)
\end{aligned}$$

Equation (4.20) can be rewritten as (4.21). Then, applying the matrix inversion lemma (see Section C.4 in Appendix C) we get (4.22). Thus, we define the interference function for the user k as $I_k(\boldsymbol{\xi}) = \xi_k \mathrm{MMSE}_k^{\mathrm{MAC}}$, as follows

$$I_k(\boldsymbol{\xi}) = \left(\frac{1}{\xi}_k + |\tau_k|^2 \boldsymbol{\theta}_k^{\mathrm{H}} \left(\sum_{i \neq k} \xi_i \boldsymbol{\theta}_i |\tau_i|^2 \boldsymbol{\theta}_i^{\mathrm{H}} + \mathbf{I}_N \right)^{-1} \boldsymbol{\theta}_k \right)^{-1}. \qquad (4.24)$$

To use the special properties of the *standard* interference functions we have to proof that $I_k(\boldsymbol{\xi})$ satisfies the conditions presented in Appendix B, i.e., positivity, monotonicity, and scalability. Taking into account that $\boldsymbol{\xi} \geq \mathbf{0}$, and $(\sum_{i \neq k} \xi_i \boldsymbol{\theta}_i |\tau_i|^2 \boldsymbol{\theta}_i^{\mathrm{H}} + \mathbf{I}_N)^{-1}$ is a positive-definite matrix (see Section C.5 of Appendix C), $I_k(\boldsymbol{\xi}) \geq 0$, and positivity is fulfilled. Moreover, when we consider the power allocation $\boldsymbol{\xi}' \geq \boldsymbol{\xi}$, $(\sum_{i \neq k} \xi_i' \boldsymbol{\theta}_i |\tau_i|^2 \boldsymbol{\theta}_i^{\mathrm{H}} + \mathbf{I}_N)^{-1} \leq (\sum_{i \neq k} \xi_i \boldsymbol{\theta}_i |\tau_i|^2 \boldsymbol{\theta}_i^{\mathrm{H}} + \mathbf{I}_N)^{-1}$, and $1/\xi_k' \leq 1/\xi_k$. Thus, $I_k(\boldsymbol{\xi}') \geq I_k(\boldsymbol{\xi})$, and the monotonicity condition is proved. Consider now the power allocation $\alpha\boldsymbol{\xi}$, with $\alpha > 1$. Then, using that $\alpha > 1$, we have

$$\begin{aligned}
I_k(\alpha\boldsymbol{\xi}) &= \left(\frac{1}{\alpha \xi}_k + \frac{|\tau_k|^2}{\alpha} \boldsymbol{\theta}_k^{\mathrm{H}} \left(\sum_{i \neq k} \xi_i \boldsymbol{\theta}_i |\tau_i|^2 \boldsymbol{\theta}_i^{\mathrm{H}} + \frac{1}{\alpha} \mathbf{I}_N \right)^{-1} \boldsymbol{\theta}_k \right)^{-1} \\
&< \left(\frac{1}{\alpha \xi}_k + \frac{|\tau_k|^2}{\alpha} \boldsymbol{\theta}_k^{\mathrm{H}} \left(\sum_{i \neq k} \xi_i \boldsymbol{\theta}_i |\tau_i|^2 \boldsymbol{\theta}_i^{\mathrm{H}} + \mathbf{I}_N \right)^{-1} \boldsymbol{\theta}_k \right)^{-1} = \alpha I_k(\boldsymbol{\xi}), \qquad (4.25)
\end{aligned}$$

and the scalability property is also satisfied. Now we can reformulate the optimization problem in the dual MAC of (4.15) employing the interference function $I_k(\boldsymbol{\xi})$, that is

$$\min_{\boldsymbol{\xi}} P_T^{\mathrm{MAC}} = \sum_{i=1}^{K} \xi_i \qquad \text{s.t.:} \quad I_k(\boldsymbol{\xi}) / \xi_k \leq 2^{-\rho_k}, \ \forall k. \qquad (4.26)$$

We have shown that $I_k(\boldsymbol{\xi})$ satisfies the conditions for the *standard* interference functions. These properties imply that $\boldsymbol{I}(\boldsymbol{\xi}) = [I_1(\boldsymbol{\xi}), \ldots, I_K(\boldsymbol{\xi})]$ is also *standard*. The QoS requirements can be described by the inequality $\boldsymbol{\xi} \geq \boldsymbol{Q}^{-1}\boldsymbol{I}(\boldsymbol{\xi})$, with $\boldsymbol{Q} = \mathrm{diag}(2^{-\rho_1}, \ldots, 2^{-\rho_K})$. From the properties of $\boldsymbol{I}(\boldsymbol{\xi})$, it can be concluded that the fixed point iteration $\boldsymbol{\xi}^{(\ell)} = \boldsymbol{Q}^{-1}\boldsymbol{I}(\boldsymbol{\xi}^{(\ell-1)})$ converges to the global optimum of (4.26) for given MAC transmit filters t_k, as it is proven in [38] (see also Appendix B). Since $I_k(\boldsymbol{\xi})$ is a function of $|t_k|^2$, $\forall k$, and the receive filters $\boldsymbol{g}_k$ are implicitly updated when calculating the interference function, the power minimization reduces to find the optimum $\boldsymbol{\xi}$ and compute the corresponding filters afterwards, as it can be seen in Algorithm 4.1: PM.MISO.PCSI in the ensuing section.

Finally, note that the solution for the power allocation is unique, $\boldsymbol{\xi}^{\mathrm{opt}}$ (c.f. Appendix B).

4.1.4. Algorithmic Solution

Algorithm 4.1: PM.MISO.PCSI. Power Min. in the MISO BC with Perfect CSI

Initialize: $\ell \leftarrow 0$, random initialization for $\boldsymbol{p}_k^{(0)} \forall k$
for $k = 1$ to K **do**
  $f_k \leftarrow$ update BC receiver using (4.7)
end for
$t_k, \boldsymbol{g}_k, \forall k \leftarrow$ BC-to-MAC conversion (see Section 4.1.2)
$\xi_k^{(0)} \leftarrow |t_k|^2, \forall k$
repeat
  $\ell \leftarrow \ell + 1$
  $\boldsymbol{I}(\boldsymbol{\xi}^{(\ell-1)}) \leftarrow$ update the interference function using (4.24)
  for $k = 1$ to K **do**
    $\xi_k^{(\ell)} \leftarrow 2^{\rho_k} I_k(\boldsymbol{\xi}^{(\ell-1)})$
  end for
until $\left\|\boldsymbol{\xi}^{(\ell)} - \boldsymbol{\xi}^{(\ell-1)}\right\|_1 \leq \delta$
for $k = 1$ to K **do**
  $t_k^{\mathrm{opt}} \leftarrow \sqrt{\xi_k^{(\ell)}}$
end for
for $k = 1$ to K **do**
  $\boldsymbol{g}_k^{\mathrm{opt}} \leftarrow$ update MAC receiver using (4.19)
end for
$\boldsymbol{p}_k^{\mathrm{opt}}, f_k^{\mathrm{opt}}, \forall k \leftarrow$ MAC-to-BC conversion (see Section 4.1.2)

Algorithm 4.1: PM.MISO.PCSI shows the pseudocode that solves the power

minimization problem (4.11). First, the BC precoders $\boldsymbol{p}_k$ are randomly initialized. Next, the corresponding BC receivers are computed in the line 3. Then, the conversion to the dual MAC is performed by means of the duality that allows to transform the filters from the BC domain to the dual MAC via the relationships presented in Section 4.1.2 (line 5). The initial power allocation is calculated in the line 6 prior to the loop. At every iteration, the line 9 allows to calculate the interference function for all the users. Note that the MAC receivers $\boldsymbol{g}_k$ are also implicitly updated. After that, the new power allocation is found in the line 11, and the process is repeated until a fixed precision is reached. The threshold δ in the line 13 determines when the loop is finished. Once the optimal power allocation is found, the corresponding optimal MAC transmitters are computed in the line 14, and using them we obtain the optimal MAC receivers from the step 18. With the dual MAC filters we eventually get the optimal BC filters performing the conversion from the MAC to the BC, as shown in Section 4.1.2 (line 20).

Notice that every iteration step in the algorithm reduces the total transmit power or remains unchanged when the solution is feasible. Due to the existence of a unique minimum of (4.26), this property implies that the power converges [38] (see also Appendix B). However, it is easy to see that the solution for $\boldsymbol{g}_k$ and t_k is not unique since the product of the filter by a complex number of the form $\mathrm{e}^{j\phi}$ does not have any impact on the final result (cf. (4.12)).

4.1.5. Simulation Results

In this section we present the results of some experiments carried out to show the performance of the proposed algorithm, referred to as Algorithm 4.1: PM.MISO.PCSI in the previous section. In such experiments, we consider a scenario where the *Base Station* (BS) equipped with $N = 4$ transmit antennas sends data to $K = 4$ single antenna users. The rate targets for all the users are set to $\rho_1 = 1, \rho_2 = 2, \rho_3 = 2.5$ and $\rho_4 = 1.5$ bits per channel use, respectively. Equivalently, the MMSE targets are $\varepsilon_1 = 0.5000, \varepsilon_2 = 0.2500, \varepsilon_3 = 0.1768$, and $\varepsilon_4 = 0.3536$. The channel for the user k, $\boldsymbol{h}_k$, is modeled using the Rayleigh channel model described in Section 2.3.1 of Chapter 2, i.e., $\boldsymbol{h}_k \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{I}_N)$, channels are statistically independent among users. The stop threshold δ of Algorithm 4.1: PM.MISO.PCSI is set to 10^{-4} and the noise variance is equal to 1 for all the users, i.e. $\sigma_k^2 = 1, \forall k$.

Figure 4.3 shows the evolution of the rates for all the users throughout the execution of the Algorithm 4.1: PM.MISO.PCSI. As it can be seen from the figure, the rates converge to the values given by the QoS constraints after only few iterations. Figure 4.4 shows the corresponding MMSE achieved after some iterations. Since the Algorithm 4.1: PM.MISO.PCSI solves the problem reformulated so that the QoS restrictions expressed as rates are transformed into new equivalent MMSE ones, this figure shows

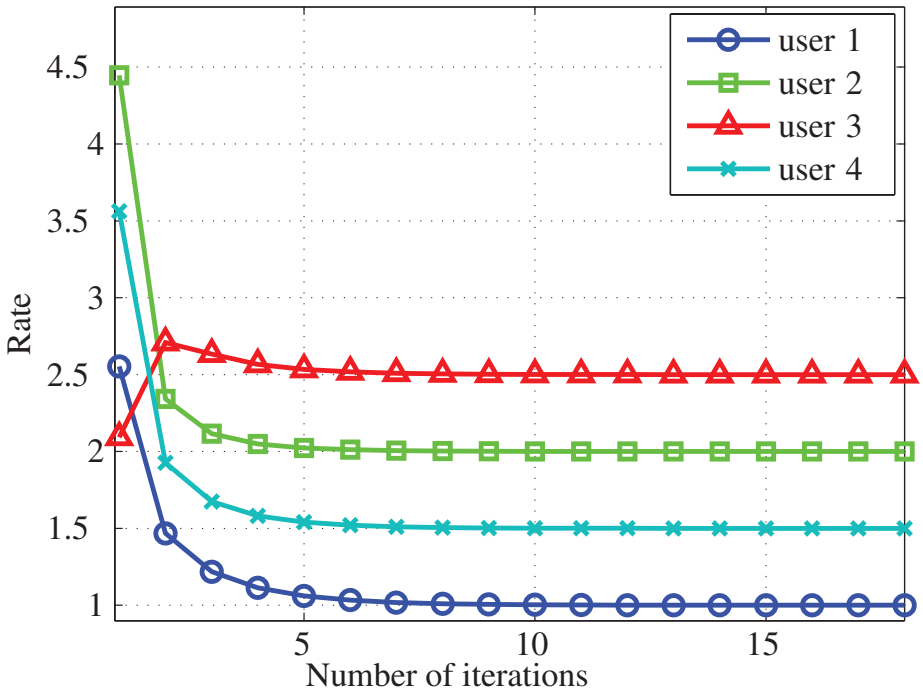


Figure 4.3: Power Minimization in the MISO BC with Perfect CSI: Rate vs. Number of Iterations.

how the performance of the measure of interest evolves by using the Algorithm 4.1: PM.MISO.PCSI. Observe that rates converge to the desired values. Finally, Fig. 4.5 depicts the total transmit power, i.e. the sum of the powers employed for all the users, P_T, to get the MMSE targets. This total power is initially $15\,dB$ and it is reduced to $8.5\,dB$ after 6 iterations. The behavior observed in this figure is the expected evolution of the total transmit power bearing in mind that both the rate targets and the corresponding MMSE targets are feasible.

4.2. Scenario 2: Imperfect Channel State Information at the Transmitter

In this section we consider the power minimization in the MISO BC subject to QoS constraints with perfect CSIR and imperfect CSIT [69].

Let us consider the system model of a MISO BC depicted in Fig. 4.1. We assume that the BS is equipped with N transmit antennas. This BS sends the data signal $s_k \in \mathbb{C}$ to the

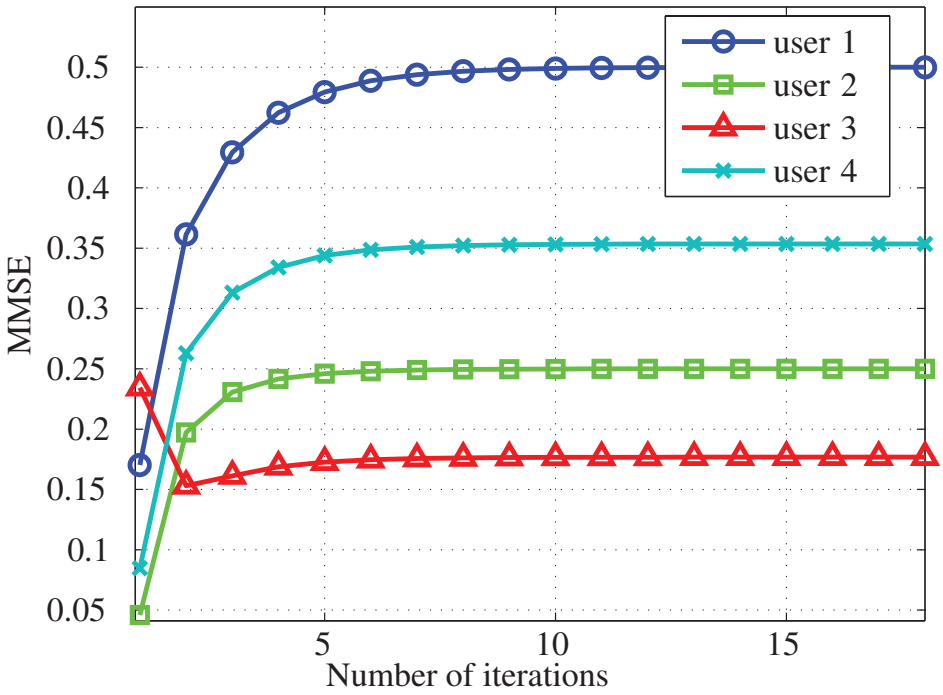


Figure 4.4: Power Minimization in the MISO BC with Perfect CSI: MMSE vs. Number of Iterations.

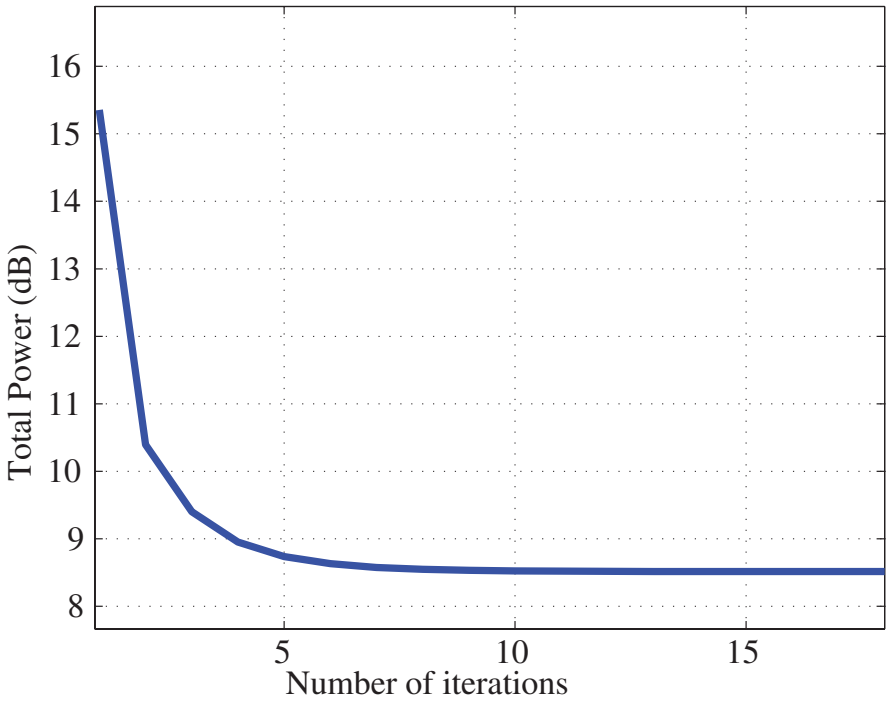


Figure 4.5: Power Minimization in the MISO BC with Perfect CSI: Total Power vs. Number of Iterations.

user $k \in \{1, \ldots, K\}$. Furthermore, data signals are zero-mean unit-variance Gaussian (i.e., $s_k \sim \mathcal{N}_{\mathbb{C}}(0,1)\ \forall k$), and uncorrelated (i.e., $\mathrm{E}[s_k^* s_j] = 0$ for $j \neq k$). Such signals are then precoded with the linear filters $\boldsymbol{p}_k \in \mathbb{C}^N$ at the BS and propagate over the vector channels $\boldsymbol{h}_k \in \mathbb{C}^N$. At the users, the received signals are linear filtered with $f_k \in \mathbb{C}$ to produce an estimate of the kth user data signal, which is expressed as

$$\hat{s}_k = f_k^* \boldsymbol{h}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i s_i + f_k^* \eta_k, \tag{4.27}$$

where $\eta_k \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_{\eta_k}^2)$ represents the AWGN which is independent from the data signals. According to (4.27), $\hat{s}_k$ is a noisy version of the data signal s_k, and the achievable data rate in such situation is given by

$$R_k = \log_2(1 + \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{p}_k \boldsymbol{p}_k^{\mathrm{H}} \boldsymbol{h}_k x_k^{-1}), \tag{4.28}$$

where we introduced the scalar

$$x_k = \boldsymbol{h}_k^{\mathrm{H}} \left(\sum_{i \neq k} \boldsymbol{p}_i \boldsymbol{p}_i^{\mathrm{H}} \right) \boldsymbol{h}_k + \sigma_{\eta_k}^2. \tag{4.29}$$

Note that this expression can be obtained particularizing the rate in (3.13) to the scenario where the users have only one antenna, and setting the covariance of the transmitted signal accordingly to $\boldsymbol{C}_k = \mathrm{E}[\boldsymbol{p}_k s_k s_k^* \boldsymbol{p}_k^{\mathrm{H}}] = \boldsymbol{p}_k \, \mathrm{E}[s_k s_k^*] \boldsymbol{p}_k^{\mathrm{H}} = \boldsymbol{p}_k \boldsymbol{p}_k^{\mathrm{H}}$.

We assume that the receiver k has perfect knowledge of all the linear precoding vectors $\{\boldsymbol{p}_k\}_{k=1}^K$ and the following CSI: its own channel, $\boldsymbol{h}_k$, and the variance of its own AWGN, $\sigma_{\eta_k}^2$. Contrarily, the BS has only access to partial CSI modeled by the random variable v and the conditional *probability density function* (pdf)s $f_{\boldsymbol{h}_k|v}(\boldsymbol{h}_k|v)$.

The partial CSI model is particularized as follows. It is reasonable to assume that the receiver k perfectly knows its corresponding true channel $\boldsymbol{h}_k$. In *Frequency-Division Duplex* (FDD) systems, all the users can feed back some information regarding their CSI to the transmitter, which is combined to get v. Alternatively, the channel can be estimated in the uplink of *Time-Division Duplex* (TDD) systems, assuming channel reciprocity between the uplink and the downlink (see Section 2.4 of Chapter 2 for further information about CSI acquisition). In both cases, the transmitter only knows $\boldsymbol{h}_k$ statistically via v. Note that no additional assumptions are made, e.g. no bounded CSI error models, nor CSI errors with a particular shape. However, the following model for the imperfect CSIT is made

$$\boldsymbol{h}_k = \bar{\boldsymbol{h}}_k + \tilde{\boldsymbol{h}}_k, \tag{4.30}$$

with $\bar{\boldsymbol{h}}_k = \mathrm{E}[\boldsymbol{h}_k | \, v]$ and $\tilde{\boldsymbol{h}}_k$ being the error due to imperfect CSI. Note that the true channel $\boldsymbol{h}_k$ is unknown at the transmitter but $\bar{\boldsymbol{h}}_k$ can be recovered from v. We assume

$\tilde{\boldsymbol{h}}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\tilde{\boldsymbol{h}}_k})$, where $\boldsymbol{C}_{\tilde{\boldsymbol{h}}_k} = \mathrm{E}[(\boldsymbol{h}_k - \bar{\boldsymbol{h}}_k)(\boldsymbol{h}_k - \bar{\boldsymbol{h}}_k)^{\mathrm{H}} | \, v]$. The error covariance matrix of the kth user, $\boldsymbol{C}_{\tilde{\boldsymbol{h}}_k}$, represents the quality of the CSIT from user k. In the limit case where $\boldsymbol{C}_{\tilde{\boldsymbol{h}}_k} = \mathbf{0}$, the BS has perfect knowledge of the channel $\boldsymbol{h}_k$ and the power minimization is solved by means of the Algorithm 4.1: PM.MISO.PCSI presented in the previous section.

As it can be seen in Section 3.1.1 of Chapter 3, when there exists partial CSIT and the channel is given by an ergodic process, the rate to be possibly achieved turns out to be

$$\mathrm{E}[R_k] = \mathrm{E}[\log_2(1 + \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{p}_k \boldsymbol{p}_k^{\mathrm{H}} \boldsymbol{h}_k x_k^{-1})]. \tag{4.31}$$

Hence, let $\rho_k, \, \forall k$, be the per-user average rate to be accomplished by the system. Remember that our ultimate objective is to design the precoders $\boldsymbol{p}_k$ that minimize the average transmit power, P_T, therefore fulfilling the inequality constraints $\mathrm{E}[R_k] \geq \rho_k, \, \forall k$. Such constrained minimization problem, however, is difficult to solve. Notice, nevertheless, that by means of Bayes' rule the average rate $\mathrm{E}[R_k]$ can be rewritten as $\mathrm{E}[\mathrm{E}[R_k | \, v]]$, where the outer expectation is over v, while the inner one is over $\boldsymbol{h}_k | \, v$. Thus, finding the optimum $\boldsymbol{p}_k$ that minimize the transmit power, P_T, for all the possible values of v is equivalent to minimizing the overall transmit power.

Therefore, for given imperfect CSI realization v we seek to determine the optimal precoders $\boldsymbol{p}_k$ that minimize the average transmit power $P_T = \sum_{k=1}^{K} ||\boldsymbol{p}_k||_2^2$ subject to the per-user conditioned average rate constraints $\mathrm{E}\left[R_k | \, v\right] \geq \rho_k, \, \forall k$, i.e.,

$$\min_{\{\boldsymbol{p}_k\}_{k=1}^K} \; P_T = \sum_{k=1}^{K} \|\boldsymbol{p}_k\|_2^2 \quad \text{subject to} \quad \mathrm{E}\left[R_k | \, v\right] \geq \rho_k, \, \forall k. \tag{4.32}$$

In the ensuing section we exploit the relationship between the average rate and the average MMSE to reformulate the optimization problem (4.32) in a more manageable way.

4.2.1. Problem Formulation

Let us start considering the kth user MSE in the BC, denoted by $\mathrm{MSE}_k^{\mathrm{BC}}$. Assuming perfect CSI at both ends of the BC, such MSE is given by $\mathrm{E}[|s_k - \hat{s}_k|^2]$, i.e.

$$\mathrm{MSE}_k^{\mathrm{BC}} = 1 - 2\Re\left\{f_k^* \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{p}_k\right\} + |f_k|^2 \left|\boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{p}_k\right|^2 + |f_k|^2 \, x_k, \tag{4.33}$$

where x_k is given by (4.29). The derivation of this performance measure can be found in Section 3.2 of Chapter 3. Correspondingly, the filter minimizing the MSE is the so-called MMSE receive filter, which is given by

$$f_k^{\text{MMSE}} = \text{argmin}_{f_k} \, \text{E}\left[|s_k - \hat{s}_k|^2\right]$$
$$= \text{argmin}_{f_k} \, \text{E}\left[\left|s_k - f_k^* \boldsymbol{h}_k^{\text{H}} \sum_{i=1}^{K} \boldsymbol{p}_i s_i + f_k^* \eta_k\right|^2\right]. \quad (4.34)$$

The MMSE receive filters are easily found computing the derivative of MSE_k with respect to f_k^*, i.e.,

$$\frac{\partial \text{MSE}_k}{\partial f_k^*} = -\boldsymbol{h}_k^{\text{H}} \boldsymbol{p}_k + f_k \left(\boldsymbol{h}_k^{\text{H}} \sum_{i=1}^{K} \boldsymbol{p}_i \boldsymbol{p}_i^{\text{H}} \boldsymbol{h}_k + \sigma_{\eta_k}^2 \right). \quad (4.35)$$

Now, by equating the last derivative to zero, we obtain

$$f_k^{\text{MMSE}} = \left(\boldsymbol{h}_k^{\text{H}} \sum_{i=1}^{K} \boldsymbol{p}_i \boldsymbol{p}_i^{\text{H}} \boldsymbol{h}_k + \sigma_{\eta_k}^2 \right)^{-1} \boldsymbol{h}_k^{\text{H}} \boldsymbol{p}_k. \quad (4.36)$$

The minimum MSE accomplished by the user k in the BC, denoted as $\text{MMSE}_k^{\text{BC}}$, is simply obtained substituting (4.36) into (4.33), i.e.

$$\text{MMSE}_k = 1 - f_k^{*,\text{MMSE}} \boldsymbol{h}_k^{\text{H}} \boldsymbol{p}_k$$
$$= 1 - \boldsymbol{p}_k^{\text{H}} \boldsymbol{h}_k \left(\boldsymbol{h}_k^{\text{H}} \sum_{i=1}^{K} \boldsymbol{p}_i \boldsymbol{p}_i^{\text{H}} \boldsymbol{h}_k + \sigma_{\eta_k}^2 \right)^{-1} \boldsymbol{h}_k^{\text{H}} \boldsymbol{p}_k. \quad (4.37)$$

Finally, having in mind (4.28) and applying the matrix inversion lemma (see Section C.4 of Appendix C) to rewrite the $\text{MMSE}_k^{\text{BC}}$ given by (4.37), the kth user rate can be expressed as

$$R_k = -\log_2(\text{MMSE}_k^{\text{BC}}). \quad (4.38)$$

Equations (4.33), (4.36), and (4.37) are suitable for the BC design when perfect CSI is available at both ends. Notice, however, that this is not the scenario studied in this section, in which the specific channel realizations $\boldsymbol{h}_k$, although available at the receivers, are unknown at the transmitter. Instead, the transmitter obtains realizations of the acquired partial CSI, v.

For this reason, we now define the average MSE at the BC as $\overline{\text{MSE}}_k^{\text{BC}} = \text{E}[\text{MSE}_k^{\text{BC}} | \, v]$ as follows

$$\overline{\text{MSE}}_k^{\text{BC}} = \text{E}\left[1 - 2\Re\left\{f_k^* \boldsymbol{h}_k^{\text{H}} \boldsymbol{p}_k\right\} + |f_k|^2 \left|\boldsymbol{h}_k^{\text{H}} \boldsymbol{p}_k\right|^2 + |f_k|^2 x_k | \, v\right], \quad (4.39)$$

where the expectation is taken over $\boldsymbol{h}_k | \, v$, i.e. over all possible channel realizations $\boldsymbol{h}_k$ for given v. The derivation of such MSE can be found in Section 3.3.2 of Chapter 3.

Notice that such expectation can be calculated at the transmitter since the pdf $f_{\boldsymbol{h}_k|v}(\boldsymbol{h}_k|v)$ is assumed to be known. Correspondingly, and having in mind (4.36), the average MMSE in the BC is given by $\overline{\text{MMSE}}_k^{\text{BC}} = \text{E}[\text{MMSE}_k^{\text{BC}}|\, v]$, i.e.

$$\begin{aligned}\overline{\text{MMSE}}_k^{\text{BC}} &= \text{E}\left[1 - f_k^{\text{MMSE},*}\boldsymbol{h}_k^{\text{H}}\boldsymbol{p}_k|\, v\right] \\ &= \text{E}\left[1 - \boldsymbol{p}_k^{\text{H}}\boldsymbol{h}_k\left(\boldsymbol{h}_k^{\text{H}}\sum_{i=1}^{K}\boldsymbol{p}_i\boldsymbol{p}_i^{\text{H}}\boldsymbol{h}_k + \sigma_{\eta_k}^2\right)^{-1}\boldsymbol{h}_k^{\text{H}}\boldsymbol{p}_k|\, v\right]. \end{aligned} \tag{4.40}$$

We now recall that, in our power minimization problem (4.32), the users average rate conditioned to v has to satisfy the QoS constraints $\text{E}\left[R_k|\, v\right] \geq \rho_k,\ \forall k$. Contrarily to the perfect CSIT scenario, where the relationship between rate and MMSE is given by (4.38), we have to employ an approximation. Taking advantage of the concavity of the $\log_2(\cdot)$ function and employing Jensen's inequality (see Appendix A), it is possible to find a lower bound for the average rate in (4.31) as

$$\text{E}[R_k|\, v] \geq \log_2(\text{E}[1 + \boldsymbol{h}_k^{\text{H}}\boldsymbol{p}_k\boldsymbol{p}_k^{\text{H}}\boldsymbol{h}_k x_k^{-1}|\, v]). \tag{4.41}$$

Due to that, we arrive at the following MMSE lower bound for the average rate applying the matrix inversion lemma (see Section C.4 of Appendix C) in (4.40)

$$\text{E}\left[R_k|\, v\right] \geq -\log_2\left(\overline{\text{MMSE}}_k^{\text{BC}}\right). \tag{4.42}$$

Therefore, the QoS constraints in the problem formulation of (4.32) can be replaced by the more restrictive average MMSE-based constraints $-\log_2\left(\overline{\text{MMSE}}_k^{\text{BC}}\right) \geq \rho_k$. The resulting conservative average MMSE bounds read as

$$\overline{\text{MMSE}}_k^{\text{BC}} \leq 2^{-\rho_k}. \tag{4.43}$$

Note that the original average rate constraints are fulfilled using the average MMSE constraints introduced in (4.43). Thus, the optimization problem (4.32) can be reformulated as follows

$$\min_{\{\boldsymbol{p}_k, f_k\}_{k=1}^K} P_T = \sum_{k=1}^{K}\|\boldsymbol{p}_k\|_2^2 \quad \text{subject to} \quad \overline{\text{MMSE}}_k^{\text{BC}} \leq 2^{-\rho_k},\ \forall k. \tag{4.44}$$

Notice that the average transmit power resulting from this reformulation is larger than that obtained by (4.32), since the constraints in (4.44) are more restrictive than those in (4.32). From now on, we focus on solving the power minimization problem given by (4.44) to obtain the BC filters. In the following section, the optimal MSE transmit and receive filters are studied employing the duality known as BC/MAC MSE duality.

4.2.2. Exploiting BC/MAC MSE Duality

It is important to note that $\overline{\text{MSE}}_k^{\text{BC}}$ in (4.39) is independent of the receive filter f_j for $j \neq k$, but depends on the precoders $\boldsymbol{p}_j$ for $j \neq k$. This means that it is possible to optimize the kth user receive filter individually using (4.36). Recall that each user perfectly knows its own channel. However, all the precoders should be jointly optimized. We propose to avoid such dependence by exploiting the MAC/BC MSE duality described in [37] (see Section 3.3.2 of Chapter 3).

Figure 4.2 depicts the block diagram of the Gaussian *Single-Input Multiple-Output* (SIMO) MAC dual to the Gaussian MISO BC. The receive and transmit filters are represented by $\boldsymbol{g}_k \in \mathbb{C}^N$ and $t_k \in \mathbb{C}$, respectively, while $\boldsymbol{\theta}_k = \boldsymbol{h}_k \sigma_{\eta_k}^{-1} \in \mathbb{C}^N$ and $\boldsymbol{n} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_N)$ represent the channel response and the AWGN in the dual MAC. The estimated symbols at the MAC receiver are

$$\hat{s}_k^{\text{MAC}} = \boldsymbol{g}_k^{\text{H}} \sum_{i=1}^{K} \boldsymbol{\theta}_i t_i s_i + \boldsymbol{g}_k^{\text{H}} \boldsymbol{n}. \tag{4.45}$$

For given transmit and receive filters, and channel realization, the MSE in the MAC reads as (see Section 3.2 of Chapter 3)

$$\text{MSE}_k^{\text{MAC}} = 1 - 2\Re\left\{\boldsymbol{g}_k^{\text{H}} \boldsymbol{\theta}_k t_k\right\} + \sum_{i=1}^{K} |t_i|^2 \left|\boldsymbol{g}_k^{\text{H}} \boldsymbol{\theta}_i\right|^2 + \|\boldsymbol{g}_k\|_2^2 . \tag{4.46}$$

In the following, we demonstrate the MSE duality between the MISO BC and the SIMO MAC. Remember that the receivers have full CSI in the BC, whereas the transmitter only has partial CSI via v. Therefore, the obtaining of the dual MAC precoder weights t_k can be based on full CSI, while the obtaining of the dual MAC equalizers $\boldsymbol{g}_k$ is limited to the knowledge of only v, i.e. partial CSIR. The average MSE in the BC, denoted as $\overline{\text{MSE}}_k^{\text{BC}}$, is given by (4.39). Thus, the average MSE in the dual MAC, $\overline{\text{MSE}}_k^{\text{MAC}} = \text{E}[\text{MSE}_k^{\text{MAC}} | \, v]$, is given by (cf. (3.22) of Section 3.2)

$$\overline{\text{MSE}}_k^{\text{MAC}} = 1 - 2\,\text{E}\left[\Re\left\{\boldsymbol{g}_k^{\text{H}} \boldsymbol{\theta}_k t_k\right\} | \, v\right] + \|\boldsymbol{g}_k\|_2^2 + \text{E}\left[\sum_{i=1}^{K} |t_i|^2 \left|\boldsymbol{g}_k^{\text{H}} \boldsymbol{\theta}_i\right|^2 | \, v\right]. \tag{4.47}$$

We now seek for conversion formulas that enable us to switch between the MAC and the BC. Suppose that the transmit and receive filters in the MAC, i.e., t_k and $\boldsymbol{g}_k$, $\forall k$, respectively, are given. We introduce the set of real positive scalars $\{\alpha_k\}_{k=1}^K$, and the following relationships between the BC and the MAC filters

$$\boldsymbol{p}_k = \alpha_k \boldsymbol{g}_k \quad \text{and} \quad f_k = \alpha_k^{-1} \sigma_{\eta_k}^{-1} t_k. \tag{4.48}$$

In Section 3.3.2 of Chapter 3 the scalars α_k such that $\overline{\text{MSE}}_k^{\text{BC}} = \overline{\text{MSE}}_k^{\text{MAC}}$, $\forall k$, are found, and it is proven that they always exist. Note that the relationship given by (4.48) not only preserves the user-wise average MSE but also the average transmit power.

A procedure similar to the one just described can be followed to determine the set of real positive scalars that enable us to determine the dual MAC filers for given BC transmit and receive filters $\boldsymbol{p}_k$ and f_k, i.e.,

$$\boldsymbol{g}_k = \beta_k^{-1}\boldsymbol{p}_k \quad \text{and} \quad t_k = \sigma_{\eta_k}\beta_k f_k. \tag{4.49}$$

In summary, the problem based on $\overline{\text{MSE}}_k^{\text{BC}}$ in the BC can be equivalently reformulated in the dual MAC with the $\overline{\text{MSE}}_k^{\text{MAC}}$ in (4.47), and vice-versa. This duality result is exploited to reformulate the problem in the MAC domain as follows,

$$\min_{\{\boldsymbol{g}_k, t_k\}_{k=1}^K} P_T^{\text{MAC}} = \sum_{k=1}^{K} \text{E}\left[|t_k|^2 \,|\, v\right] \quad \text{subject to} \quad \overline{\text{MMSE}}_k^{\text{MAC}} \leq 2^{-\rho_k},\ \forall k. \tag{4.50}$$

Due to the BC/MAC duality, the average transmit power achieved is the same for both problem formulations, that is, $P_T = P_T^{\text{MAC}}$. Hence, the optimal MAC receivers, that is, BC precoders, are computed as follows

$$\frac{\partial \overline{\text{MSE}}_k^{\text{MAC}}}{\partial \boldsymbol{g}_k^*} = -\,\text{E}\left[\boldsymbol{\theta}_k t_k |\, v\right] + \left(\sum_{i=1}^{K} \text{E}\left[\boldsymbol{\theta}_i \,|t_i|^2\, \boldsymbol{\theta}_i^{\text{H}} |\, v\right] + \mathbf{I}_N\right) \boldsymbol{g}_k, \tag{4.51}$$

where we take into account the linear property of the expectation $\text{E}[\cdot]$ and the equality $\text{tr}(\overline{\text{MSE}}_k^{\text{MAC}}) = \overline{\text{MSE}}_k^{\text{MAC}}$, to calculate the derivative with respect to the complex vector $\boldsymbol{g}_k$, as detailed in Section C.6 of Appendix C. Finally, to obtain the optimal MAC receiver we equate the previous derivative to zero, which leads to

$$\boldsymbol{g}_k^{\text{MMSE}} = \left(\sum_{i=1}^{K} \text{E}\left[\boldsymbol{\theta}_i \,|t_i|^2\, \boldsymbol{\theta}_i^{\text{H}} |\, v\right] + \mathbf{I}_N\right)^{-1} \text{E}\left[\boldsymbol{\theta}_k t_k |\, v\right]. \tag{4.52}$$

Remember that the users perfectly know their own channel. Then, the optimum receive filters are found in the BC as shown in (4.36). For computing the MMSE BC precoders, however, we move to the dual MAC where the interdependence is removed, and it is possible to calculate them as the optimal MAC receivers, $\boldsymbol{g}_k^{\text{MMSE}}$. Finding the optimum filters is not trivial since any precoder updating has an impact over the receivers, and vice-versa. Moreover, we move from one domain to another to find individually the optimal filters. Thus, finding the optimal filters is an AO process since we iterate until achieving the convergence of both filters.

4.2.3. Power Allocation Via Interference Functions

The discussion in the previous section leads us to the expressions of the optimal MMSE transmit and receive filters. In (4.52), the MMSE receivers of the dual MAC are given. Observe, however, that there exist two expectations which have to be computed. In order to do that, our proposal is to perform Monte Carlo numerical integration with the M realizations obtained from the pdfs $f_{\boldsymbol{h}_k|v}(\boldsymbol{h}_k|v)$, assumed to be available at the transmitter. We previously established the information obtained via v as $\{\bar{\boldsymbol{h}}_k, \boldsymbol{C}_{\tilde{\boldsymbol{h}}_k}\}_{k=1}^K$, with the imperfect CSI channel model given by (4.30). Observe that no further assumptions are made. Then, the channel realizations are computed as

$$\hat{\boldsymbol{h}}_k^{(m)} = \bar{\boldsymbol{h}}_k + \tilde{\boldsymbol{h}}_k^{(m)}, \tag{4.53}$$

where $\tilde{\boldsymbol{h}}_k^{(m)} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_k)$. We collect the M channel realizations $\hat{\boldsymbol{h}}_k^{(m)}$, with $m = 1, \ldots, M$, into $\boldsymbol{\Theta}_k = \sigma_{\eta_k}^{-1}[\hat{\boldsymbol{h}}_k^{(1)}, \ldots, \hat{\boldsymbol{h}}_k^{(M)}] \in \mathbb{C}^{N\times M}$, for $k = 1, \ldots, K$. The average transmit power is accordingly defined as $\xi_k = \frac{1}{M}\sum_{m=1}^M |t_k^{(m)}|^2$. We also define the matrix $\boldsymbol{T}_k \in \mathbb{C}^{M\times M}$ containing the normalized transmit filters of user k, which for notation simplicity are expressed as

$$\boldsymbol{T}_k = \frac{1}{\sqrt{\xi_k}} \operatorname{diag}\left(t_k^{(1)}, \ldots, t_k^{(M)}\right). \tag{4.54}$$

Therefore, the expectations of (4.47) are approximated by matrix products. Introducing the all ones vector $\mathbf{1}$, the new average MSE reads as

$$\overline{\mathrm{MSE}}_k^{\mathrm{MAC}} = 1 - \frac{2}{M}\Re\left\{\boldsymbol{g}_k^{\mathrm{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}\sqrt{\xi}_k\right\} + \|\boldsymbol{g}_k\|_2^2 + \frac{1}{M}\boldsymbol{g}_k^{\mathrm{H}}\sum_{i=1}^K \xi_i\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}}\boldsymbol{g}_k, \tag{4.55}$$

and the MAC filters are rewritten accordingly, that is,

$$\boldsymbol{g}_k^{\mathrm{MMSE}} = \left(\frac{1}{M}\sum_{i=1}^K \xi_i\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}} + \mathbf{I}_N\right)^{-1}\frac{1}{M}\sqrt{\xi_k}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}. \tag{4.56}$$

Hence, the MMSE in the dual MAC is obtained by substituting (4.56) into (4.55) to get

$$\overline{\mathrm{MMSE}}_k^{\mathrm{MAC}} = 1 - \frac{1}{M^2}\xi_k\mathbf{1}^{\mathrm{T}}\boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{\Theta}_k^{\mathrm{H}}\left(\frac{1}{M}\sum_{i=1}^K \xi_i\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}} + \mathbf{I}_N\right)^{-1}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}. \tag{4.57}$$

Note that equalities (4.55) and (4.56) are true if the number of channel realizations M is sufficiently large and ergodicity holds true. In such case, the Monte Carlo numerical integrations implicit in (4.55) and (4.56) are tight approximations to the conditional

expectations in (4.47) and (4.52). Likewise, substituting (4.56) into (4.55) we obtain the expression for the average MMSE (4.57).

As described above, the optimal transmit and receive filters have been obtained although it is still necessary to allocate the powers, that is, how the power is distributed among the users. To that end, we first rewrite the expression of the MMSE including the matrix $\boldsymbol{X}_k = \frac{1}{M}\sum_{i\neq k}\xi_i\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}} + \mathbf{I}_N$ that contains the inter-user interference and the noise, as follows

$$\begin{aligned}
\overline{\mathrm{MMSE}}_k^{\mathrm{MAC}} &= 1 - \frac{1}{M^2}\xi_k\mathbf{1}^{\mathrm{T}}\boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{\Theta}_k^{\mathrm{H}}\left(\frac{\xi_k}{M}\boldsymbol{\Theta}_k\boldsymbol{T}_k\boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{\Theta}_k^{\mathrm{H}} + \boldsymbol{X}_k\right)^{-1}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1} \\
&= 1 - \frac{1}{M^2}\xi_k\mathbf{1}^{\mathrm{T}}\boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{\Theta}_k^{\mathrm{H}}\left(\boldsymbol{X}_k^{-1} - \boldsymbol{X}_k^{-1}\boldsymbol{\Theta}_k\boldsymbol{T}_k\right. \\
&\quad \left.\left(\frac{M}{\xi_k}\mathbf{I}_M + \boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{\Theta}_k^{\mathrm{H}}\boldsymbol{X}_k^{-1}\boldsymbol{\Theta}_k\boldsymbol{T}_k\right)^{-1}\boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{\Theta}_k^{\mathrm{H}}\boldsymbol{X}_k^{-1}\right)\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1} \qquad (4.58) \\
&= 1 - \frac{1}{M^2}\xi_k\mathbf{1}^{\mathrm{T}}\boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{\Theta}_k^{\mathrm{H}}\left(\boldsymbol{X}_k^{-1}\boldsymbol{\Theta}_k\boldsymbol{T}_k\right. \\
&\quad \left.\left(\mathbf{I}_M - \left(\frac{M}{\xi_k}\mathbf{I}_M + \boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{\Theta}_k^{\mathrm{H}}\boldsymbol{X}_k^{-1}\boldsymbol{\Theta}_k\boldsymbol{T}_k\right)^{-1}\boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{\Theta}_k^{\mathrm{H}}\boldsymbol{X}_k^{-1}\boldsymbol{\Theta}_k\boldsymbol{T}_k\right)\right)\mathbf{1} \\
&= \frac{1}{M}\mathbf{1}^{\mathrm{T}}\left(\mathbf{I}_M - \boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{\Theta}_k^{\mathrm{H}}\boldsymbol{X}_k^{-1}\boldsymbol{\Theta}_k\boldsymbol{T}_k\left(\frac{M}{\xi_k}\mathbf{I}_M + \boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{\Theta}_k^{\mathrm{H}}\boldsymbol{X}_k^{-1}\boldsymbol{\Theta}_k\boldsymbol{T}_k\right)^{-1}\right)\mathbf{1} \\
&= \frac{1}{\xi_k}\mathbf{1}^{\mathrm{T}}\left(\frac{M}{\xi_k}\mathbf{I}_M + \boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{\Theta}_k^{\mathrm{H}}\boldsymbol{X}_k^{-1}\boldsymbol{\Theta}_k\boldsymbol{T}_k\right)^{-1}\mathbf{1}, \qquad (4.59)
\end{aligned}$$

where we have applied the matrix inversion lemma (see Section C.4 of Appendix C) to get (4.58). Thus, the new expression for the MAC MMSE is

$$\overline{\mathrm{MMSE}}_k^{\mathrm{MAC}} = \frac{1}{\xi_k}\mathbf{1}^{\mathrm{T}}\left(\frac{M}{\xi_k}\mathbf{I}_N + \boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{\Theta}_k^{\mathrm{H}}\left(\frac{1}{M}\sum_{i\neq k}\xi_i\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}} + \mathbf{I}_N\right)^{-1}\boldsymbol{\Theta}_k\boldsymbol{T}_k\right)^{-1}\mathbf{1}. \qquad (4.60)$$

Now, we resort to the *standard* interference function framework proposed in [38] (see also Appendix B). This framework provides useful tools for functions that satisfy the required properties. We define the interference function for the user k, as $I_k(\boldsymbol{\xi}) = \xi_k\overline{\mathrm{MMSE}}_k^{\mathrm{MAC}}$, i.e.

$$I_k(\boldsymbol{\xi}) = \mathbf{1}^{\mathrm{T}}\left(\frac{M}{\xi_k}\mathbf{I}_N + \boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{\Theta}_k^{\mathrm{H}}\left(\frac{1}{M}\sum_{i\neq k}\xi_i\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}} + \mathbf{I}_N\right)^{-1}\boldsymbol{\Theta}_k\boldsymbol{T}_k\right)^{-1}\mathbf{1}. \qquad (4.61)$$

In the following we show that the proposed interference function satisfies the required properties to be *standard*, i.e., positivity, monotonicity, and scalability.

To proof positivity, we focus on the matrix

$$\boldsymbol{A}_k = \frac{1}{M}\sum_{i\neq k}\xi_i\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}} + \mathbf{I}_N, \tag{4.62}$$

which is positive definite (see Section C.5 of Appendix C). The same property holds for its inverse, $\boldsymbol{A}_k^{-1}$, and for the matrix product $\boldsymbol{B}_k = \boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{\Theta}_k^{\mathrm{H}}\boldsymbol{A}_k^{-1}\boldsymbol{\Theta}_k\boldsymbol{T}_k$. Finally, taking that into account we conclude that the inverse of the matrix $\boldsymbol{C}_k = M/\xi_k\mathbf{I}_N + \boldsymbol{B}_k$ is also positive definite, leading to $I_k(\boldsymbol{\xi}) > 0$ for $\boldsymbol{\xi} \geq \mathbf{0}$. Consider now the power allocation $\boldsymbol{\xi}' \geq \boldsymbol{\xi}$. Then, the corresponding matrix $\boldsymbol{A}_k'^{,-1} \leq \boldsymbol{A}_k^{-1}$ and the same is true for the matrices $\boldsymbol{B}_k' \leq \boldsymbol{B}_k$ and $\boldsymbol{C}_k' \leq \boldsymbol{C}_k$. Accordingly, $I_k(\boldsymbol{\xi}') \geq I_k(\boldsymbol{\xi})$ and the monotonicity property is fulfilled. To check if the scalability property is satisfied we introduce the power allocation $\alpha\boldsymbol{\xi}$, with $\alpha > 1$, to obtain

$$\begin{aligned} I_k(\alpha\boldsymbol{\xi}) &= \mathbf{1}^{\mathrm{T}}\left(\frac{M}{\alpha\xi_k}\mathbf{I}_N + \frac{1}{\alpha}\boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{\Theta}_k^{\mathrm{H}}\left(\frac{1}{M}\sum_{i\neq k}\xi_i\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}} + \frac{1}{\alpha}\mathbf{I}_N\right)^{-1}\boldsymbol{\Theta}_k\boldsymbol{T}_k\right)^{-1}\mathbf{1} \\ &< \alpha\mathbf{1}^{\mathrm{T}}\left(\frac{M}{\xi_k}\mathbf{I}_N + \boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{\Theta}_k^{\mathrm{H}}\left(\frac{1}{M}\sum_{i\neq k}\xi_i\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}} + \mathbf{I}_N\right)^{-1}\boldsymbol{\Theta}_k\boldsymbol{T}_k\right)^{-1}\mathbf{1} \\ &= \alpha I_k(\boldsymbol{\xi}). \end{aligned} \tag{4.63}$$

Then, the necessary properties are fulfilled and the vector containing the interference functions for all the users $\boldsymbol{I}(\boldsymbol{\xi}) = [I_1(\boldsymbol{\xi}), \ldots, I_K(\boldsymbol{\xi})]^{\mathrm{T}}$ is *standard*. Notice, however, that every iteration in the power control algorithm performs an update of the interference function $\boldsymbol{I}(\boldsymbol{\xi})$, which corresponds to computing the two matrix inversions of (4.61). Therefore, we propose an alternative interference function that allows to avoid such costly computational operations.

We introduce the scalar receive filter in the MAC, r_k, so that $\boldsymbol{g}_k = r_k\tilde{\boldsymbol{g}}_k$. Such scalar filters allow a simple update of the receive filters norm in the power iteration step. Considering this alternative notation, the MAC average MSE reads as

$$\begin{aligned} \overline{\mathrm{MSE}}_{k,\mathrm{scalar}}^{\mathrm{MAC}} = 1 &- \frac{2}{M}\Re\left\{r_k^*\tilde{\boldsymbol{g}}_k^{\mathrm{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}\sqrt{\xi_k}\right\} + |r_k|^2\left\|\tilde{\boldsymbol{g}}_k\right\|_2^2 \\ &+ |r_k|^2\,\tilde{\boldsymbol{g}}_k^{\mathrm{H}}\left(\frac{1}{M}\sum_{i=1}^{K}\xi_i\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}}\right)\tilde{\boldsymbol{g}}_k. \end{aligned} \tag{4.64}$$

The MMSE scalar receiver is calculated by doing the derivative of (4.64) with respect to r_k^*, that is,

$$\frac{\partial\overline{\mathrm{MSE}}_{k,\mathrm{scalar}}^{\mathrm{MAC}}}{\partial r_k^*} = \frac{\sqrt{\xi_k}}{M}\tilde{\boldsymbol{g}}_k^{\mathrm{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1} + r_k\tilde{\boldsymbol{g}}_k^{\mathrm{H}}\left(\frac{1}{M}\sum_{i=1}^{K}\xi_i\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}}\right)\tilde{\boldsymbol{g}}_k + r_k\left\|\tilde{\boldsymbol{g}}_k\right\|_2^2. \tag{4.65}$$

Equating the last expression to zero, we get the MMSE scalar filters as follows,

$$r_k^{\text{MMSE}} = \frac{\frac{1}{M}\tilde{\boldsymbol{g}}_k^{\text{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}\sqrt{\xi_k}}{\tilde{\boldsymbol{g}}_k^{\text{H}}\left(\frac{1}{M}\sum_{i=1}^{K}\xi_i\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\text{H}}\boldsymbol{\Theta}_i^{\text{H}}\right)\tilde{\boldsymbol{g}}_k + \|\tilde{\boldsymbol{g}}_k\|_2^2}. \tag{4.66}$$

Substituting the optimal scalar receiver, r_k^{MMSE}, into (4.64) we can, eventually, compute the scalar receiver MMSE as follows

$$\begin{aligned}\overline{\text{MMSE}}_{k,\text{scalar}}^{\text{MAC}} &= 1 - \frac{2\xi_k}{M^2}\left|\tilde{\boldsymbol{g}}_k^{\text{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}\right|^2 y_k^{-1}(\boldsymbol{\xi}) + \frac{\xi_k}{M^2}\left|\tilde{\boldsymbol{g}}_k^{\text{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}\right|^2 y_k^{-1}(\boldsymbol{\xi})y_k(\boldsymbol{\xi})y_k^{-1}(\boldsymbol{\xi}) \\ &= 1 - \frac{\xi_k}{M^2}\left|\tilde{\boldsymbol{g}}_k^{\text{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}\right|^2 y_k^{-1}(\boldsymbol{\xi}), \end{aligned} \tag{4.67}$$

where

$$y_k(\boldsymbol{\xi}) = \tilde{\boldsymbol{g}}_k^{\text{H}}\left(\frac{1}{M}\sum_{i=1}^{K}\xi_i\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\text{H}}\boldsymbol{\Theta}_i^{\text{H}}\right)\tilde{\boldsymbol{g}}_k + \|\tilde{\boldsymbol{g}}_k\|_2^2. \tag{4.68}$$

This new MMSE expression can be rewritten applying the equality $1 - \frac{a}{b} = (1 + \frac{a}{b-a})^{-1}$, as follows

$$\overline{\text{MMSE}}_{k,\text{scalar}}^{\text{MAC}} = \left(1 + \frac{\xi_k}{M^2}\left|\tilde{\boldsymbol{g}}_k^{\text{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}\right|^2 z_k^{-1}(\boldsymbol{\xi})\right)^{-1}, \tag{4.69}$$

with $z_k(\boldsymbol{\xi}) = y_k(\boldsymbol{\xi}) - \frac{\xi_k}{M^2}|\tilde{\boldsymbol{g}}_k^{\text{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}|^2$.

So far, we have found the optimal vector and scalar receivers in the MAC, $\tilde{\boldsymbol{g}}_k$ and r_k^{MMSE}, respectively, corresponding to the BC precoder $\boldsymbol{p}_k$. The receivers depend on the normalized MAC precoders $\boldsymbol{T}_k$, i.e., the normalized BC receivers, and the MAC power allocation $\boldsymbol{\xi}$. In the following, we will find the jointly optimal MAC power allocation, $\boldsymbol{\xi}$, and receivers, $\boldsymbol{g}_k$, for given normalized precoders, $\boldsymbol{T}_k$. To that end, we rely again on standard interference functions.

Accordingly, we define the interference function $J_k(\boldsymbol{\xi}) = \xi_k\overline{\text{MMSE}}_{k,\text{scalar}}^{\text{MAC}}$ that can be interpreted as the interference for user k,

$$J_k(\boldsymbol{\xi}) = \left(\frac{1}{\xi_k} + \frac{1}{M^2}\left|\tilde{\boldsymbol{g}}_k^{\text{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}\right|^2 z_k^{-1}(\boldsymbol{\xi})\right)^{-1}. \tag{4.70}$$

Consequently, we need to check if the required conditions for the function $J_k(\boldsymbol{\xi})$ to be *standard* are satisfied.

First, the positivity property is straightforward to be proved taking into account that $z_k(\boldsymbol{\xi}) = y_k(\boldsymbol{\xi}) - \frac{\xi_k}{M^2}|\tilde{\boldsymbol{g}}_k^{\text{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}|^2$, with $y_k(\boldsymbol{\xi})$ from (4.68), is monotonically increasing in $\boldsymbol{\xi}$ and positive. That is, if we focus on the difference for the elements of the user k in $z_k(\boldsymbol{\xi})$, we get

$$\frac{\xi_k}{M}\tilde{\boldsymbol{g}}_k^{\text{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k(\boldsymbol{I}_M - 1/M\mathbf{1}\mathbf{1}^{\text{T}})\boldsymbol{T}_k^{\text{H}}\boldsymbol{\Theta}_k^{\text{H}}\tilde{\boldsymbol{g}}_k, \tag{4.71}$$

which contains the matrix $\boldsymbol{\Pi} = \boldsymbol{I}_M - 1/M\mathbf{1}\mathbf{1}^{\mathrm{T}}$. Such a matrix is Hermitian and its eigenvalues (see Section C.1 of Appendix C for further details) are $M-1$ one elements and only one 0. Therefore, $\boldsymbol{\Pi}$ is positive-semidefinite. Then, the expression inside brackets in (4.70) is positive.

Secondly, monotonicity comes from $z_k(\boldsymbol{\xi})$ being monotonically increasing in $\boldsymbol{\xi}$, as we previously mentioned. Therefore, the second term in (4.70), $\frac{1}{M^2}|\tilde{\boldsymbol{g}}_k^{\mathrm{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}|^2 z_k^{-1}$, is decreasing in $\boldsymbol{\xi}$, as well as the first one, $1/\xi_k$, and, finally, the inverse of the summation inside brackets in (4.70) increases with $\boldsymbol{\xi}$. That is $I_k(\boldsymbol{\xi}) \geq I_k(\boldsymbol{\xi}')$, with $\boldsymbol{\xi} \geq \boldsymbol{\xi}'$, fulfilling the required property.

Finally, scalability is proved considering the scalar $\alpha > 1$. Then, we have that

$$\begin{aligned}\alpha J_k(\boldsymbol{\xi}) &= \alpha\left(\frac{1}{\xi_k} + \frac{1}{M}\left|\tilde{\boldsymbol{g}}_k^{\mathrm{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}\right|^2 z_k^{-1}(\boldsymbol{\xi})\right)^{-1} > \left(\frac{1}{\alpha\xi_k} + \frac{1}{M}\left|\tilde{\boldsymbol{g}}_k^{\mathrm{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}\right|^2\right. \\ &\left.\left(\alpha\tilde{\boldsymbol{g}}_k^{\mathrm{H}}\left(\frac{1}{M}\sum_{i=1}^{K}\xi_i\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}}\right)\tilde{\boldsymbol{g}}_k + \|\tilde{\boldsymbol{g}}_k\|_2^2 - \alpha\frac{\xi_k}{M^2}\left|\tilde{\boldsymbol{g}}_k^{\mathrm{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}\right|^2\right)^{-1}\right)^{-1} \\ &= J_k(\alpha\boldsymbol{\xi}). \end{aligned} \tag{4.72}$$

Consider now the optimal MMSE receive filter in the dual MAC given by (4.56). Let us introduce the variables $\boldsymbol{R} = \frac{1}{M}\sum_{i=1}^{K}\xi_i\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}}$ and $\boldsymbol{\mu}_k = \frac{1}{M}\sqrt{\xi_k}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}$ such that $\boldsymbol{g}_k^{\mathrm{MMSE}} = (\boldsymbol{R}+\mathbf{I}_N)^{-1}\boldsymbol{\mu}_k$. Substituting $\tilde{\boldsymbol{g}}_k = \boldsymbol{g}_k^{\mathrm{MMSE}}$ in $z_k(\boldsymbol{\xi})$ from (4.69) gives

$$z_k(\boldsymbol{\xi}) = \boldsymbol{\mu}_k^{\mathrm{H}}(\boldsymbol{R}+\mathbf{I})^{-1}\boldsymbol{\mu}_k - \left(\boldsymbol{\mu}_k^{\mathrm{H}}(\boldsymbol{R}+\mathbf{I})^{-1}\boldsymbol{\mu}_k\right)^2. \tag{4.73}$$

Employing this expression in the scalar average MMSE expression of (4.69), where the MMSE scalar receiver r_k^{MMSE} is used, we get for $\tilde{\boldsymbol{g}}_k = \boldsymbol{g}_k^{\mathrm{MMSE}}$ that

$$\begin{aligned}\overline{\mathrm{MMSE}}_{k,\mathrm{scalar}}^{\mathrm{MAC}} &= \left(1 + \frac{\left(\boldsymbol{\mu}_k^{\mathrm{H}}\left(\boldsymbol{R}+\mathbf{I}_N\right)^{-1}\boldsymbol{\mu}_k\right)^2}{\boldsymbol{\mu}_k^{\mathrm{H}}\left(\boldsymbol{R}+\mathbf{I}_N\right)^{-1}\boldsymbol{\mu}_k - \left(\boldsymbol{\mu}_k^{\mathrm{H}}\left(\boldsymbol{R}+\mathbf{I}_N\right)^{-1}\boldsymbol{\mu}_k\right)^2}\right)^{-1} \\ &= 1 - \boldsymbol{\mu}_k^{\mathrm{H}}\left(\boldsymbol{R}+\mathbf{I}_N\right)^{-1}\boldsymbol{\mu}_k \\ &= \overline{\mathrm{MMSE}}_k^{\mathrm{MAC}}. \end{aligned} \tag{4.74}$$

As can be seen, applying $(1 + a/(1-a))^{-1} = 1 - a$ leads to the conclusion that $\overline{\mathrm{MMSE}}_{k,\mathrm{scalar}}^{\mathrm{MAC}} = \overline{\mathrm{MMSE}}_k^{\mathrm{MAC}}$, with $\overline{\mathrm{MMSE}}_k^{\mathrm{MAC}}$ given by (4.57), if $\tilde{\boldsymbol{g}}_k = \boldsymbol{g}_k^{\mathrm{MMSE}}$. Thus, the two interference functions lead to the same power allocation in every step if the receive filters are updated at every step of the fixed point iteration using the scalar interference function.

Remember the average MSE duality and the relationship between the average rate and the average MMSE. Then, the QoS constraints of the original problem formulation can

equivalently be expressed as $\overline{\text{MMSE}}_k^{\text{MAC}} \leq 2^{-\rho_k}$, as previously shown in the optimization problem (4.50). Now, we introduce the interference function $\overline{\text{MMSE}}_{k,\text{scalar}}^{\text{MAC}} = J_k(\boldsymbol{\xi})/\xi_k$, to get eventually the MAC problem reformulation as follows

$$\min_{\boldsymbol{\xi},\{\tilde{\boldsymbol{g}}_k\}_{k=1}^K} \sum_{i=1}^{K} \xi_i \quad \text{subject to} \quad J_k\left(\boldsymbol{\xi}\right)/\xi_k \leq 2^{-\rho_k},\ \forall k. \tag{4.75}$$

Recall that $J_k(\boldsymbol{\xi})$ satisfies the conditions for the *standard* interference functions. These properties imply that $\boldsymbol{J}(\boldsymbol{\xi}) = [J_1(\boldsymbol{\xi}), \ldots, J_K(\boldsymbol{\xi})]$ is also *standard*. The QoS requirements can be described by the inequality $\boldsymbol{\xi} \geq \boldsymbol{Q}^{-1}\boldsymbol{J}(\boldsymbol{\xi})$, with $\boldsymbol{Q} = \text{diag}(2^{-\rho_1}, \ldots, 2^{-\rho_K})$. From the properties of $\boldsymbol{J}(\boldsymbol{\xi})$, it can be concluded that the fixed point iteration $\boldsymbol{\xi}^{(\ell)} = \boldsymbol{Q}^{-1}\boldsymbol{J}(\boldsymbol{\xi}^{(\ell-1)})$ converges to the global optimum of (4.26) for given $\boldsymbol{g}_k$, as proven in [38] (see also Appendix B). In [70], the iteration was extended to find the global optimum of (4.26), i.e., to find the optimal filter $\boldsymbol{g}_k$ as considered in that work.

We next define an auxiliary function $Z_k(\boldsymbol{\xi}, \tilde{\boldsymbol{g}}_k)$, which is equal to $J_k(\boldsymbol{\xi})$ when a fixed $\tilde{\boldsymbol{g}}_k$ is employed, and set to $\boldsymbol{Z}(\boldsymbol{\xi}, \tilde{\boldsymbol{G}}) = [Z_1(\boldsymbol{\xi}, \tilde{\boldsymbol{g}}_1), \ldots, Z_K(\boldsymbol{\xi}, \tilde{\boldsymbol{g}}_K)]^{\text{T}}$, with $\tilde{\boldsymbol{G}} = [\tilde{\boldsymbol{g}}_1, \ldots, \tilde{\boldsymbol{g}}_K]$. Since $\boldsymbol{Z}(\boldsymbol{\xi}, \tilde{\boldsymbol{G}})$ is *standard*, also $\min_{\tilde{\boldsymbol{G}}} \boldsymbol{Z}(\boldsymbol{\xi}, \tilde{\boldsymbol{G}})$ is *standard* with the element-wise minimization [38]. As shown in [70], the iteration

$$\begin{aligned} \forall k: \tilde{\boldsymbol{g}}_k^{(\ell)} &\leftarrow \arg\min_{\tilde{\boldsymbol{g}}_k} Z_k\left(\boldsymbol{\xi}^{(\ell-1)}, \tilde{\boldsymbol{g}}_k\right), \\ \boldsymbol{\xi}^{(\ell)} &\leftarrow \boldsymbol{Q}^{-1}\boldsymbol{Z}\left(\boldsymbol{\xi}^{(\ell-1)}, \tilde{\boldsymbol{G}}^{(\ell)}\right), \end{aligned} \tag{4.76}$$

converges to the global optimum of (4.75). However, it is important to note that the power allocation and the filters are optimum if and only if the following conditions are simultaneously achieved for all k

$$\begin{aligned} \tilde{\boldsymbol{g}}_k^{\text{opt}} &= \arg\min_{\tilde{\boldsymbol{g}}_k} Z_k\left(\boldsymbol{\xi}^{\text{opt}}, \tilde{\boldsymbol{g}}_k\right), \\ \xi_k^{\text{opt}} &= 2^{\rho_k} Z_k\left(\boldsymbol{\xi}^{\text{opt}}, \tilde{\boldsymbol{g}}_k^{\text{opt}}\right). \end{aligned} \tag{4.77}$$

The last line of (4.77) has to be fulfilled for the optimum power allocation, reached at the convergence of the power control algorithm.

Consider now that there exists a receiver leading into an interference lower than that achieved with $\boldsymbol{g}_k^{\text{opt}}$. Then it is possible to find a power allocation smaller than ξ_k^{opt} for which the last line of (4.77) holds. Therefore, $\boldsymbol{\xi}^{\text{opt}}$ will not be the optimum anymore. Hence, both conditions have to be satisfied at the same time.

Note that the solution for the power allocation is unique, $\boldsymbol{\xi}^{\text{opt}}$ (cf. Appendix B). Nevertheless, that is not true for the MAC receive filters $\boldsymbol{g}_k^{\text{opt}}$ since the product $\mathrm{e}^{j\phi}\,\boldsymbol{g}_k^{\text{opt}}$, for any ϕ, does not impact the MSE, e.g. (4.64). Hence, when the conditions (4.77) are reached, we optimally solve one of the steps in the AO. Such step is where the MAC

power allocation and the MAC receiver, which corresponds to the BC precoder due to the duality (see (4.48)), are found for given MAC precoders, i.e. BC receivers.

The BC receivers f_k are computed for given BC precoders $\boldsymbol{p}_k$ (see (4.36)). With the BC/MAC dual transform, the obtained BC receivers become the MAC precoders, t_k. In the dual MAC, the optimum MAC power allocation, $\boldsymbol{\xi}$, and the MAC receiver, $\boldsymbol{g}_k$, are calculated by means of the iteration of (4.76). Again, applying the BC/MAC duality, the optimum MAC receivers become the BC precoders and the iteration loop continues with the search of the optimal BC receivers for given BC precoders, as we will show in the ensuing section.

4.2.4. Algorithmic Solution

Algorithm 4.2: PM.MISO.ICSI.1. Power Min. by AO. (First Implementation)

$\ell \leftarrow 0$, initialize $\boldsymbol{p}_i^{(0)}$, $\forall i$
repeat
  $\ell \leftarrow \ell + 1$
  for $m = 1$ to M **do**
    $f_k^{(\ell-1,m)} \leftarrow$ update BC receiver using (4.36), $\forall k$
  end for
  $t_k^{(\ell-1,m)}, \tilde{\boldsymbol{g}}_k^{(l-1)}$, $\forall k, \forall m \leftarrow$ BC-to-MAC conversion (see Section 4.2.2)
  $\xi_k^{(\ell-1)} \leftarrow \frac{1}{M} \sum_{m=1}^{M} |t_k^{(\ell-1,m)}|^2$, $\forall k$
  $\boldsymbol{T}_k^{(\ell)} \leftarrow 1/\sqrt{\xi_k^{(\ell-1)}}\, \mathrm{diag}(t_k^{(\ell,1)}, \ldots, t_k^{(\ell,M)})$, $\forall k$
  $n \leftarrow 0$, $\boldsymbol{\xi}^{[0]} \leftarrow [\xi_1^{(\ell)}, \ldots, \xi_K^{(\ell)}]^{\mathrm{T}}$, $\tilde{\boldsymbol{g}}_k^{[0]} \leftarrow \tilde{\boldsymbol{g}}_k^{(\ell-1)}$, $\forall k$
  repeat
    $\boldsymbol{J}(\boldsymbol{\xi}^{[n]}) \leftarrow$ update interference function (4.70)
    $\xi_k^{[n+1]} \leftarrow 2^{\rho_k}\, J_k(\boldsymbol{\xi}^{[n]})$, $\forall k$
    $\tilde{\boldsymbol{g}}_k^{[n+1]} \leftarrow$ update MAC receiver (4.56), $\forall k$
    $n \leftarrow n + 1$
  until $||\boldsymbol{\xi}^{[n]} - \boldsymbol{\xi}^{[n-1]}||_1 \leq \delta_1$
  $\boldsymbol{\xi}^{(\ell)} \leftarrow \boldsymbol{\xi}^{[n]}$, $\tilde{\boldsymbol{g}}_k^{(\ell)} \leftarrow \tilde{\boldsymbol{g}}_k^{[n]}$, $\forall k$
  for $m = 1$ to M **do**
    $t_k^{(\ell,m)} \leftarrow \sqrt{\xi_k^{(\ell)}[\boldsymbol{T}_k^{(\ell)}]_{m,m}}$ update MAC precoder, $\forall k$
  end for
  $\boldsymbol{p}_k^{(\ell)}, f_k^{(\ell,m)}$, $\forall k, \forall m \leftarrow$ MAC-to-BC conversion (see Section 4.2.2)
until $\left\|\boldsymbol{\xi}^{(\ell)} - \boldsymbol{\xi}^{(\ell-1)}\right\|_1 \leq \delta$

Algorithm 4.3: PM.MISO.ICSI.2. Power Min. by AO. (Second Implementation)

$\ell \leftarrow 0$, initialize $\boldsymbol{p}_i^{(0)}$, $\forall i$
repeat
  $\ell \leftarrow \ell + 1$
  for $m = 1$ to M **do**
    $f_k^{(\ell-1,m)} \leftarrow$ update BC receiver using (4.36), $\forall k$
  end for
  $t_k^{(\ell-1,m)}, \tilde{\boldsymbol{g}}_k^{(\ell-1)}$, $\forall k, \forall m \leftarrow$ BC-to-MAC conversion (see Section 4.2.2)
  $\xi_k^{(\ell-1)} \leftarrow \frac{1}{M} \sum_{m=1}^{M} |t_k^{(\ell-1,m)}|^2$, $\forall k$
  $\boldsymbol{T}_k^{(\ell)} \leftarrow 1/\sqrt{\xi_k^{(\ell-1)}}\, \mathrm{diag}(t_k^{(\ell-1,1)}, \ldots, t_k^{(\ell-1,M)})$, $\forall k$
  $\boldsymbol{J}(\boldsymbol{\xi}^{(\ell-1)}) \leftarrow$ update interference function (4.70)
  $\xi_k^{(\ell)} \leftarrow 2^{\rho_k}\, J_k(\boldsymbol{\xi}^{(\ell-1)})$, $\forall k$
  $\tilde{\boldsymbol{g}}_k^{(\ell)} \leftarrow$ update MAC receiver (4.56), $\forall k$
  for $m = 1$ to M **do**
    $t_k^{(\ell,m)} \leftarrow \sqrt{\xi_k^{(\ell)}}[\boldsymbol{T}_k^{(\ell)}]_{m,m}$ update MAC precoder, $\forall k$
  end for
  $\boldsymbol{p}_k^{(\ell)}, f_k^{(\ell,m)}$, $\forall k, \forall m \leftarrow$ MAC-to-BC conversion (see Section 4.2.2)
until $||\boldsymbol{\xi}^{(\ell)} - \boldsymbol{\xi}^{(\ell-1)}||_1 \leq \delta$

In this section, we present two implementations of the algorithm for power minimization referred to as Algorithm 4.2: PM.MISO.ICSI.1, and Algorithm 4.3: PM.MISO.ICSI.2, respectively. These algorithms solve the minimization problem (4.44) in a suboptimally way via AO. The first part of the algorithms, from the lines 1 to 9, is common to both of them. After the initialization, the line 5 updates the BC receivers for every channel realization $m \in 1, \ldots, M$ and user $k \in 1, \ldots, K$, and given BC precoders. Then, we switch from the BC to the MAC to obtain the corresponding MAC transmit and receive filters in the line 7. The transmit filters are decomposed in the lines 8 and 9 to get first the power allocation and then the normalized MAC transmit matrices. Now, we focus on the Algorithm 4.2: PM.MISO.ICSI.1, where an initialization for the inner loop is performed in step 10. Such an inner loop naturally arises to perform the fixed point iteration of (4.76), i.e., the interference functions are computed in the line 12 to find the new power allocation that satisfies the QoS conditions in the line 13. Then, it follows the updating of the MAC receivers in the line 14. Recall that employing $\boldsymbol{J}(\boldsymbol{\xi})$ instead of $\boldsymbol{I}(\boldsymbol{\xi})$ to reduce the computation complexity involves to explicitly update the MAC receive filters, contrarily to what happened with $\boldsymbol{I}(\boldsymbol{\xi})$. The iteration of the inner loop is repeated until desired accuracy, δ_1, is reached, and the power allocation and the MAC receivers are updated (line 17). Nevertheless, the inner loop entails some difficulties as we will explain in the following.

In the aforementioned inner loop, the MAC receiver updating is incorporated into the interference function as a minimization. That is considered inside the *standard* interference function framework [38] (see Appendix B for further details). In other words, employing both of them the required properties for the function are fulfilled. However, inside such an inner loop only the power allocation and the receive filters are updated. Due to that, the use of the fixed transmit filters could lead to situations where the optimization problem is not feasible, that is, the QoS cannot be fulfilled for given MAC precoders t_k, $\forall k$. This behavior, although non desirable, can be avoided since the convergence of the inner loop is not necessary for the overall convergence of the algorithm, as we observed from our simulation experiments with Algorithm 4.3: PM.MISO.ICSI.2.

In Algorithm 4.3: PM.MISO.ICSI.2, we propose to prevent these unfeasible parameters without including the inner loop. In the line 10 the interference function is calculated. Afterwards, as in the Algorithm 4.2: PM.MISO.ICSI.1, the power allocation and the MAC receive filters are updated (see the lines 11 and 12, respectively), i.e, the iteration (4.76) is performed. Finally, the subsequent code is again common to both algorithms 4.2: PM.MISO.ICSI.1 and 4.3: PM.MISO.ICSI.2. The computation of the new power assignment for the MAC transmit filters is shown in the line 19 for the Algorithm 4.2: PM.MISO.ICSI.1, and the line 14 for the Algorithm 4.3: PM.MISO.ICSI.2. Next, we switch back to the BC. The overall convergence of both algorithms is checked in the last line of the code, where the threshold is set to δ. Due

to the existence of a unique minimum in (4.44), and to the fact that every step in the Algorithm 4.3: PM.MISO.ICSI.2 either reduces the average MMSE or the power, with the power lower bounded by $\mathbf{0}$, the convergence of the algorithm is guaranteed if the QoS constraints are feasible.

4.2.5. Simulation Results

Here, we present the numerical results obtained from the simulation experiments carried out to show the performance of the two algorithms proposed in the previous section, denoted as Algorithm 4.2: PM.MISO.ICSI.1 and Algorithm 4.3: PM.MISO.ICSI.2. In our scenario, the BS equipped with $N = 4$ transmit antennas sends information to $K = 4$ single-antenna users.

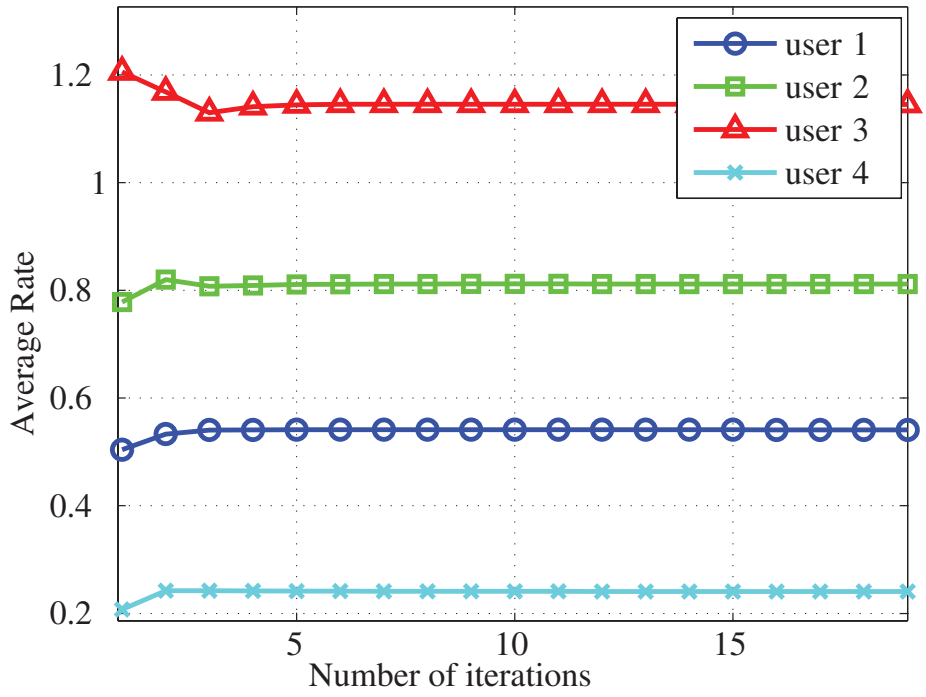


Figure 4.6: Power Minimization in the MISO BC with Imperfect CSI: Rate vs. Number of Iterations (Algorithm 4.2: PM.MISO.ICSI.1).

We employ the channel model presented in (4.30), where the CSI at the transmitter is given by v. Such a model allows to generate the channel realizations $\hat{\boldsymbol{h}}_k^{(m)} = \bar{\boldsymbol{h}}_k + \tilde{\boldsymbol{h}}_k^{(m)}$, for $k = \{1, \ldots, K\}$, and $m = \{1, \ldots, M\}$, with $\tilde{\boldsymbol{h}}_k^{(m)} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\tilde{\boldsymbol{h}}_k})$. We generated $M = 1\,000$ channel realizations considering $\boldsymbol{C}_{\tilde{\boldsymbol{h}}_k} = \mathbf{I}_N$, and $\bar{\boldsymbol{h}}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\bar{\boldsymbol{h}}_k}), \forall k$. The

AWGN is considered to be equal for all the users with a variance $\sigma^2_{\eta_k} = 1, \forall k$.

We choose different average rate requirements for all the users, viz., $\rho_1 = 0.5146, \rho_2 = 0.737, \rho_3 = 1$, and $\rho_4 = 0.2345$ bits per channel use, respectively. These requirements correspond to the following MMSE lower bounds $\varepsilon_1 = 0.7, \varepsilon_2 = 0.6, \varepsilon_3 = 0.5$, and $\varepsilon_4 = 0.85$, respectively, using $\varepsilon_k = 2^{-\rho_k}, \forall k$. The finishing thresholds for both algorithms are set to $\delta = 10^{-4}$, and $\delta_1 = 10^{-3}$ for the stop threshold of the inner loop in the Algorithm 4.2: PM.MISO.ICSI.1. The initial precoders are randomly generated since its choice does not affect the algorithm's convergence nor the final result.

The performance of the Algorithm 4.2: PM.MISO.ICSI.1 is depicted in Figs. 4.6 to 4.8. Fig. 4.6 shows the evolution of the average rates for all the users with the number of iterations. Observe that the QoS constrains are fulfilled in a conservative way, which corresponds to higher transmit total average power. Figure 4.7 represents the evolution of the corresponding average MSEs. During the first iterations, the MSE targets are not feasible for the given MAC precoders, which translates into high total average powers. The number of iterations of the outer loop is limited, otherwise for infeasible constraints the total average power increases trying to fulfill the MSE targets and the convergence of the inner loop is never reached. After the first iterations, the average MSEs remain flat due to the convergence of the inner loop with feasible conditions and updated versions of the MAC transmit filters. Finally, Fig. 4.8 shows how the power evolves during the algorithm computation. Observe that for the first iterations corresponding to non-feasible requirements the power grows to large values. Then, after few iterations the power reduces at every iteration until the desired accuracy is reached at $3\,dB$.

Figure 4.9 shows how the MMSE of each user converges to the desired target ε_k for the Algorithm 4.3: PM.MISO.ICSI.2. Since the problem is feasible, the minimum total average power will be reached when the constraints in (4.75) are fulfilled with equality. As can be seen, the first iterations go in the direction of fulfilling the requirements so the MMSEs are reduced or increased accordingly. Nevertheless, the subsequent iterations refine the MMSEs until the targets ε_k are reached for all the users. Correspondingly, as shown in Fig. 4.10, the total average power (i.e., $\sum_{i=1}^{K} \xi_i$) gradually reduces throughout the iterations until convergence is reached at about 3 dB. The total average power is dramatically reduced during the first iterations whereas the improvement is marginal for the last iterations.

Figure 4.11 shows the evolution of the average rates over the iterations. Considering the MMSE-based targets ε_k, the real average rates are lower bounded by $\mathrm{E}[R_k|\, v] \geq -\log_2(\varepsilon_k)$, as discussed in Section 4.2.1 (see also [13]). The gap between the average rates obtained with Algorithm 4.3: PM.MISO.ICSI.2 and the average rate targets corresponding to the QoS constraints can be also observed in Fig. 4.11. Moreover, we also include in this figure the rates obtained employing the *Signal to Interference-plus-Noise Ratio* (SINR) approximation utilized in [7] and widely employed afterwards (e.g. [13],

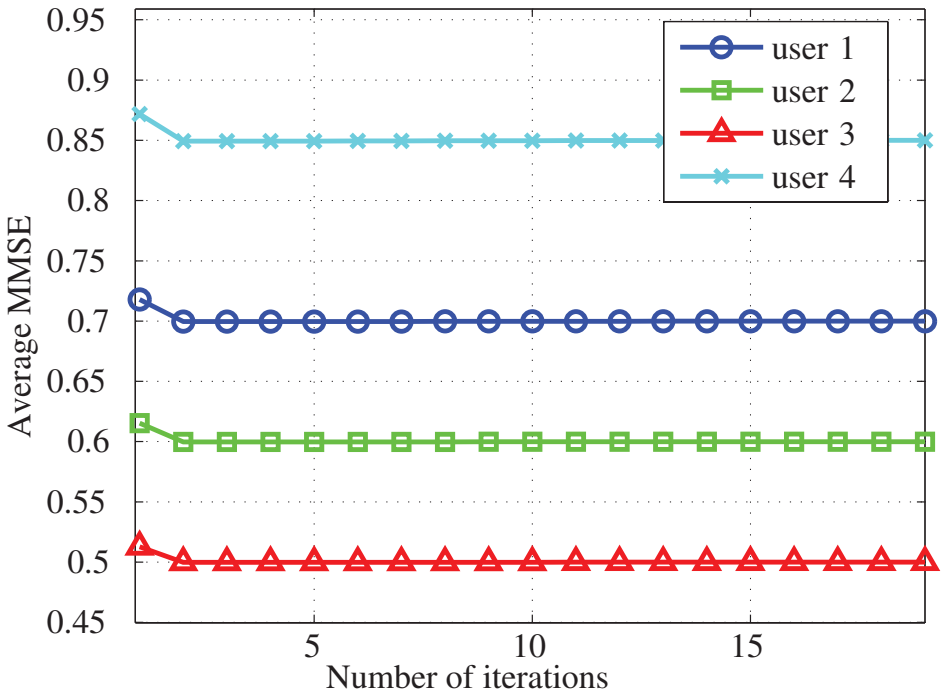


Figure 4.7: Power Minimization in the MISO BC with Imperfect CSI: MMSE vs. Number of Iterations (Algorithm 4.2: PM.MISO.ICSI.1).

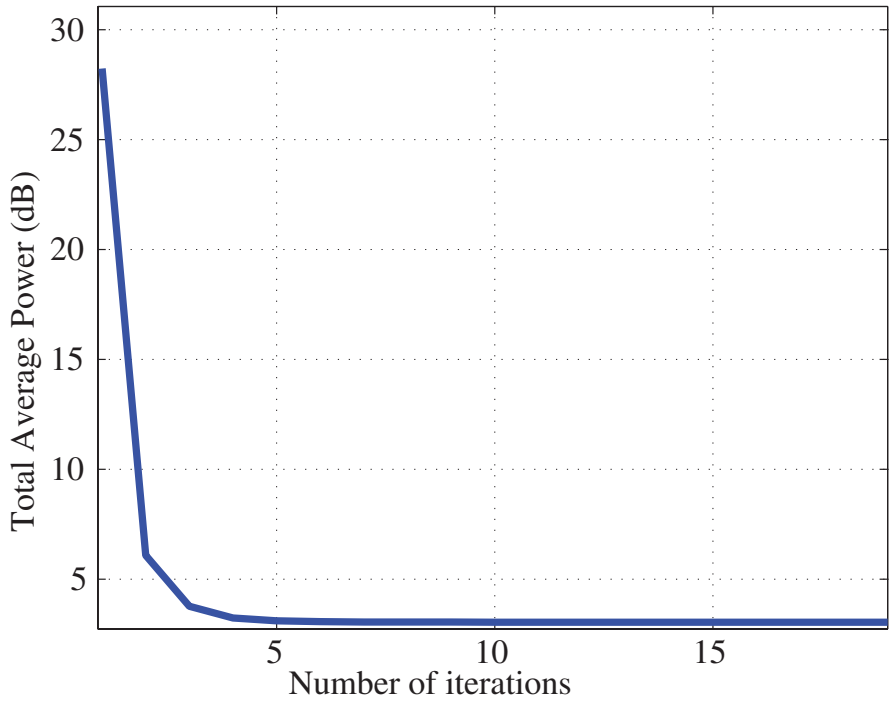


Figure 4.8: Power Minimization in the MISO BC with Imperfect CSI, Power vs. Number of Iterations (Algorithm 4.2: PM.MISO.ICSI.1).

[23], [18]). This approach determines the average rates as $\mathrm{E}[R_k|\, v] = \log_2(1 + \overline{\mathrm{SINR}}_k)$, where $\overline{\mathrm{SINR}}_k$ is obtained from applying separately the expectation operator to both the numerator and the denominator of the SINR, i.e.

$$\overline{\mathrm{SINR}}_k = \frac{\boldsymbol{p}_k^{\mathrm{H}}\, \mathrm{E}\left[\boldsymbol{h}_k \boldsymbol{h}_k^{\mathrm{H}}|\, v\right] \boldsymbol{p}_k}{\sigma_{\eta_k}^2 + \sum_{i \neq k} \boldsymbol{p}_i^{\mathrm{H}}\, \mathrm{E}\left[\boldsymbol{h}_k \boldsymbol{h}_k^{\mathrm{H}}|\, v\right] \boldsymbol{p}_i}. \tag{4.78}$$

Figure 4.11 shows the resulting values for $\log_2(1 + \overline{\mathrm{SINR}}_k)$ along the iterations in Algorithm 4.3: PM.MISO.ICSI.2. Note that the average rates for the SINR approximation are larger than the true average rates for users 2 and 3, but smaller for users 1 and 4. Hence, it is not possible to know beforehand whether the given average rate requirements are fulfilled or not because this approximation does not allow to determine whether the rates are larger or smaller than the desired ones. Contrarily, fulfilling the MMSE-based targets, as proposed in our approach, ensures average rates which are larger than the target rates.

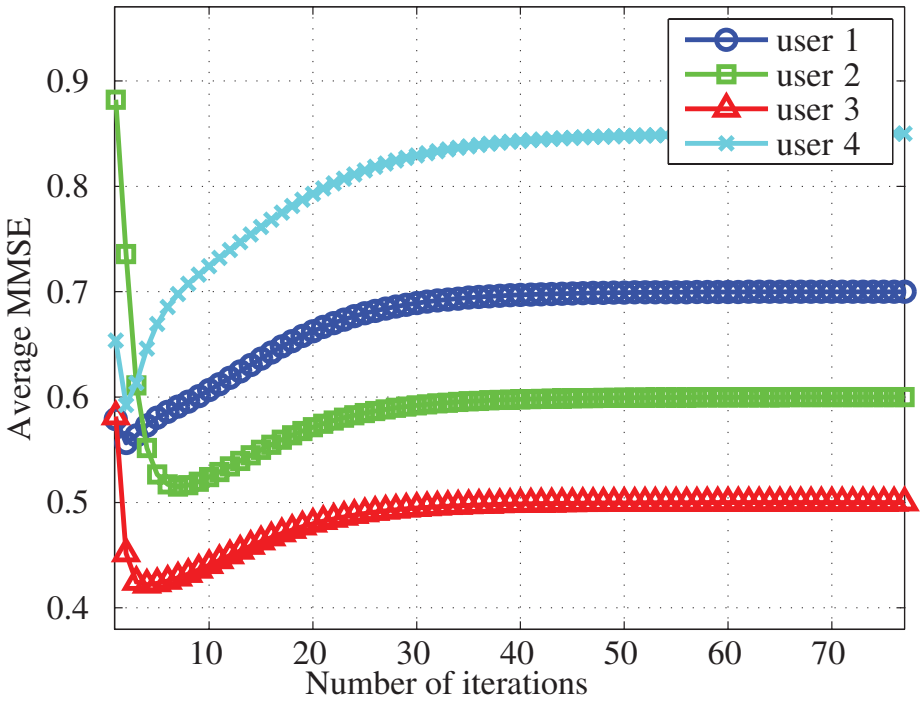


Figure 4.9: Power Minimization in the MISO BC with Imperfect CSI: MMSE vs. Number of Iterations (Algorithm 4.3: PM.MISO.ICSI.2).

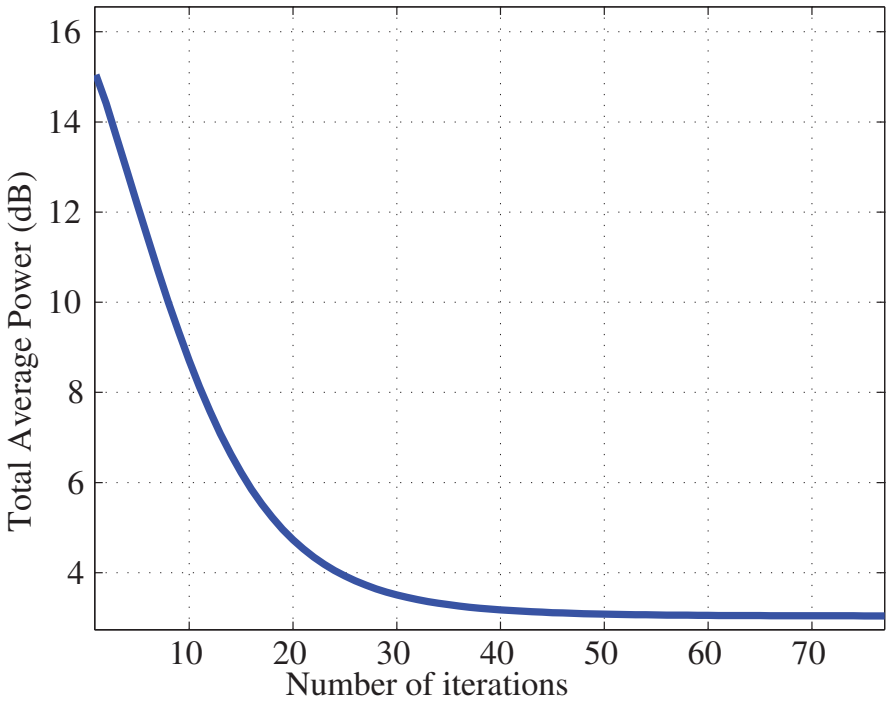


Figure 4.10: Power Minimization in the MISO BC with Imperfect CSI: Power vs. Number of Iterations (Algorithm 4.3: PM.MISO.ICSI.2).

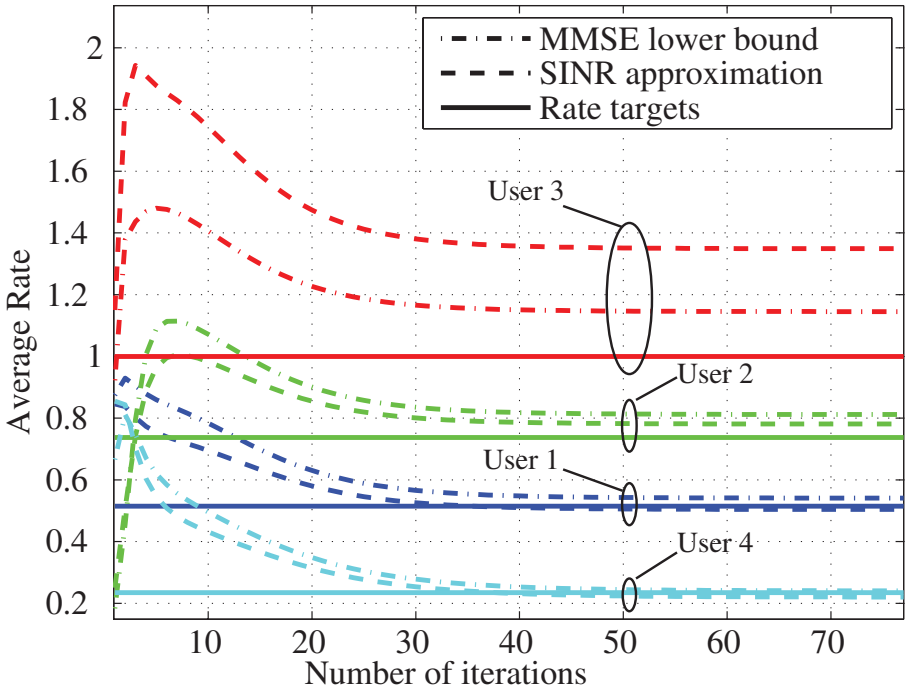


Figure 4.11: Power Minimization in the MISO BC with Imperfect CSI: Rate vs. Number of Iterations (Algorithm 4.3: PM.MISO.ICSI.2).

4.3. Conclusions

In this chapter we have addressed the problem of the power minimization in the MISO BC. First, the perfect CSIT scenario is considered as a simpler approximation to the problem of interest. Our proposal consists on exploiting the relationship between the rate and the MMSE, to allow for a problem formulation easier to deal with. Moreover, using the MSE BC/MAC duality, it is possible to perform an AO to find the MMSE transmit and receive filters. The total transmit power is minimized by means of *standard* interference functions. An algorithmic solution is finally proposed to solve the optimization problem.

In the second section, we move to the more involved problem where the CSIT is only partial. Applying the methods proposed for the perfect CSIT scenario, together with an approximation via the Jensen's inequality, we eventually find a solution. We proposed an algorithm for the power minimization in the MISO BC under minimum ergodic rate constraints via imposing conservative average MMSE constraints. Furthermore, two different *standard* interference functions are proposed, allowing to reduce the computational complexity.

Two possible implementations of the proposed algorithm are evaluated. The first one is rejected since convergence problems arise for some iterations of the algorithm even when the QoS constraints are feasible. However, with the second proposed implementation, the convergence to the minimum total transmit power is guaranteed.

We have carried out computer experiments with the purpose of comparing the results obtained with the Jensen's inequality based on the lower bound, with those resulting from the ergodic rate approximation above mentioned. The comparison demonstrates that using the lower bound the original QoS constraints are fulfilled. On the contrary, the QoS restrictions can be violated when the approximation is used. Moreover, such experiments also show that the gap between the lower bound and the true average rate is small.

Chapter 5

Transmit Power Minimization and QoS Feasibility in MIMO Broadcast Channels

In this chapter we aim to jointly achieve individual rate requirements and minimum total transmit power in a *Multiple-User Multiple-Input Multiple-Output* (MU-MIMO) *Broadcast Channel* (BC). Data streams are transmitted from a multi-antenna *Base Station* (BS) to several independent and non-cooperative multi-antenna users. Perfect *Channel State Information* (CSI) is assumed to be known at the receivers and it fed back to the transmitter, where only partial CSI is used for the design of the linear transmit filters. Note that the optimization of the linear precoders based on data rates is difficult in the case of imperfect CSI. Therefore, we rely on the average *Mean Square Error* (MSE) to end up with the optimization of the average rates lower bounds. Moreover, employing the duality between the *Multiple Access Channel* (MAC) and the BC with respect to the average MSE we design the linear filters for both the transmitter and the receivers. The duality proof conserving the average total transmit power was shown in [37] (see also Section 3.3 of Chapter 3). Thanks to that, and identifying *standard* interference functions, it is possible to find the optimal transceivers and power allocation by means of a fixed-point iteration. In such a way, we propose an algorithmic joint solution for the transmit filter design and the power allocation.

We also show that the resulting algorithm converges if the *Quality-of-Service* (QoS) constraints can be fulfilled. Nevertheless, the algorithm is meaningless if the optimization problem is not feasible. Therefore, we will provide a test to check the feasibility of the average rate restrictions. There are several works concerning feasibility (e.g. [7, 39, 40]) but they are based on the perfect CSI assumption. The mentioned test is a generalization of that presented in [40] for the vector BC and perfect CSI assumptions for both the transmitter and the receivers.

Additionally, we consider the balancing problem, that is, the maximization of the minimum of the weighted average rates under a total transmit power constraint. Again, this problem is reformulated by conservatively bounding the average rates based on the average *Minimum Mean Square Error* (MMSE)s, leading to the minimization of the maximum weighted average MMSE with a total power constraint. Note that, contrarily to the power minimization, the balancing problem is always feasible. The proposed algorithm takes advantage of the solution elaborated for the power minimization problem, and performs a bisection search to find the larger balanced average rates for given average transmit power.

5.1. Power Minimization in the MIMO BC with Imperfect CSIT

In this section we extend the solution proposed for *Multiple-Input Single-Output* (MISO) BC in Section 4.2 of Chapter 4 to the *Multiple-Input Multiple-Output* (MIMO) BC. Again, perfect *Channel State Information at the Receiver* (CSIR) and imperfect *Channel State Information at the Transmitter* (CSIT) are considered [71].

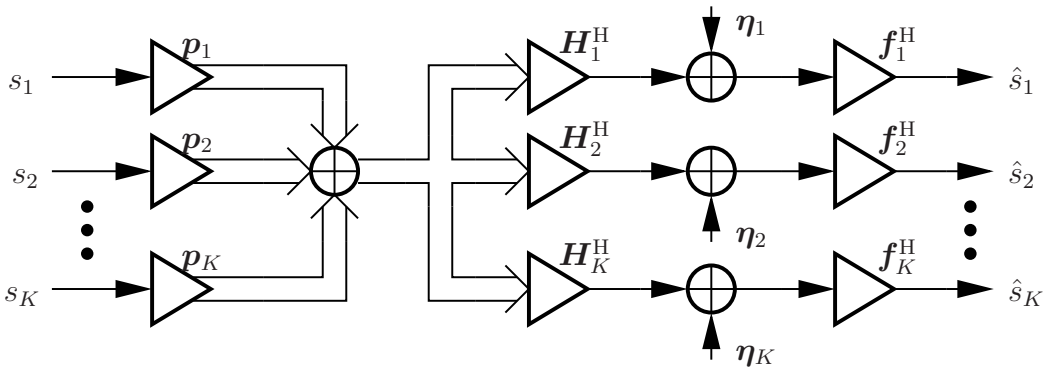


Figure 5.1: MIMO Broadcast Channel.

Figure 5.1 depicts the BC model considered in this section. The zero-mean Gaussian data signal $s_k \in \mathbb{C}$ for user k, with $1 \leq k \leq K$, and $\mathrm{E}[|s_k|^2] = 1$, is precoded by $\boldsymbol{p}_k \in \mathbb{C}^N$, where K and N are the number of users and transmit antennas, respectively. The transmit signal propagates over a MIMO channel $\boldsymbol{H}_k \in \mathbb{C}^{N \times R}$, with R being the number of receive antennas for each user. The *Additive White Gaussian Noise* (AWGN) in the MIMO channel is $\boldsymbol{\eta}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{\eta}_k})$. The data signals are mutually independent and also independent from the noise. The linear equalizer $\boldsymbol{f}_k \in \mathbb{C}^R$ provides the data

symbols estimates as follows

$$\hat{s}_k = \boldsymbol{f}_k^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i s_i + \boldsymbol{f}_k^{\mathrm{H}} \boldsymbol{\eta_k}. \tag{5.1}$$

In Section 3.1.1 of Chapter 3, it is shown that the rate of the MIMO BC is given by (3.13). Taking into account the system model for the BC described in this section, the covariance of the transmitted signal is $\boldsymbol{C}_k = \mathrm{E}[\boldsymbol{p}_k s_k s_k^* \boldsymbol{p}_k^{\mathrm{H}}] = \boldsymbol{p}_k \, \mathrm{E}[s_k s_k^*] \boldsymbol{p}_k^{\mathrm{H}} = \boldsymbol{p}_k \boldsymbol{p}_k^{\mathrm{H}}$, and (3.13) can be rewritten as

$$R_k = \log_2 \det \left(\mathbf{I}_R + \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{p}_k \boldsymbol{p}_k^{\mathrm{H}} \boldsymbol{H}_k \left(\boldsymbol{C}_{\boldsymbol{\eta}_k} + \boldsymbol{H}_k^{\mathrm{H}} \sum_{i \neq k} \boldsymbol{p}_i \boldsymbol{p}_i^{\mathrm{H}} \boldsymbol{H}_k \right)^{-1} \right). \tag{5.2}$$

We consider the transmitter does not have a perfect knowledge of the CSI but a partial one modeled through v. We assume the conditional *probability density function* (pdf)s $f_{\boldsymbol{H}_k|v}\left(\boldsymbol{H}_k|v\right)$ are available for all k. Contrarily, the receivers are assumed to know their own channel $\boldsymbol{H}_k$, i.e, we consider perfect CSIR. To model the random variable v we assume the same conditions as in Section 4.2 of Chapter 4, i.e., only statistical information is contained into the imperfect CSI v. Hence, we model the error as follows

$$\boldsymbol{H}_k = \bar{\boldsymbol{H}}_k + \tilde{\boldsymbol{H}}_k, \tag{5.3}$$

with $\bar{\boldsymbol{H}}_k = \mathrm{E}[\boldsymbol{H}_k | \, v]$ and $\tilde{\boldsymbol{H}}_k$ being the imperfect CSI error, with $\tilde{\boldsymbol{H}}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\tilde{\boldsymbol{H}}_k})$, where $\boldsymbol{C}_{\tilde{\boldsymbol{H}}_k} = \mathrm{E}[(\boldsymbol{H}_k - \bar{\boldsymbol{H}}_k)(\boldsymbol{H}_k - \bar{\boldsymbol{H}}_k)^{\mathrm{H}} | \, v]$. The CSI quality at the BS is given by this error covariance. Observe that this model is a reasonable approach for both *Frequency-Division Duplex* (FDD) and *Time-Division Duplex* (TDD) types of CSI acquisition (see Chapter 2).

As previously mentioned, we address the minimization of the total transmit power when certain QoS restrictions given as per-user rates have to be fulfilled. In (5.2) the rate for the MIMO BC was presented. However, this expression holds only when the CSIT is perfect. Due to that, and considering that the channel is given by an ergodic process, we have to employ the ergodic rate expression obtained in Section 3.1.1 of Chapter 3

$$\mathrm{E}[R_k] = \mathrm{E}\left[\log_2 \det \left(\mathbf{I}_R + \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{p}_k \boldsymbol{p}_k^{\mathrm{H}} \boldsymbol{H}_k \left(\boldsymbol{C}_{\boldsymbol{\eta}_k} + \boldsymbol{H}_k^{\mathrm{H}} \sum_{i \neq k} \boldsymbol{p}_i \boldsymbol{p}_i^{\mathrm{H}} \boldsymbol{H}_k \right)^{-1} \right) \right]. \tag{5.4}$$

As already shown in Section 3.1.1, when there exists partial CSIT the Bayes' rule can be applied to obtain the ergodic rate as the average over all possible realizations of the random variable v, for the average rates conditioned to v, i.e., $\mathrm{E}[R_k] = \mathrm{E}[\mathrm{E}[R_k | \, v]]$. In other words, if $\mathrm{E}[R_k | \, v] \geq \rho_k$, $\forall k$, with $\{\rho_k\}_{k=1}^K$ being the set of targets for all the users

for every realization of v, then also $\mathrm{E}[R_k] \geq \rho_k$, $\forall k$, is fulfilled. Therefore, we focus on $\mathrm{E}[R_k|\, v]$ from now on.

Let us define the total transmit power as $P_T = \sum_{k=1}^{K} ||\boldsymbol{p}_k||_2^2$. Likewise, finding the optimum precoders $\boldsymbol{p}_k$ which minimize the transmit power P_T for all possible values of v is equivalent to minimizing the overall transmit power. Therefore, the problem can be equivalently solved via the optimization for all the values of v, that is

$$\min_{\{\boldsymbol{p}_k\}_{k=1}^{K}} P_T = \sum_{k=1}^{K} \|\boldsymbol{p}_k\|_2^2 \quad \text{subject to} \quad \mathrm{E}\left[R_k|\, v\right] \geq \rho_k,\ \forall k. \tag{5.5}$$

Due to the conditional expectation this problem is hard to solve without further assumptions. The optimization problem (5.5) is non-convex, and our strategy in the following is to find an approximation that allows us to reformulate the problem with new constraints based on the average MSE.

5.1.1. Problem Formulation

Let us introduce the MIMO MSE for the BC, i.e $\mathrm{MSE}_k^{\mathrm{BC}} = \mathrm{E}[|s_k - \hat{s}_k|^2]$, as an extension emerging from the *Single-Input Single-Output* (SISO) MSE computed in (3.20)

$$\mathrm{MSE}_k^{\mathrm{BC}} = 1 - 2\Re\left\{\boldsymbol{f}_k^{\mathrm{H}}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{p}_k\right\} + \boldsymbol{f}_k^{\mathrm{H}}\boldsymbol{H}_k^{\mathrm{H}}\sum_{i=1}^{K}\boldsymbol{p}_i\boldsymbol{p}_i^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{f}_k + \boldsymbol{f}_k^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{\eta}_k}\boldsymbol{f}_k. \tag{5.6}$$

Recall that the users perfectly know their own channel. Hence, any meaningful receive filters are functions of the channel state (see [37]). The receive filters that minimize the MSE measure read as

$$\begin{aligned}\boldsymbol{f}_k^{\mathrm{MMSE}} &= \mathrm{argmin}_{\boldsymbol{f}_k}\, \mathrm{E}\left[|s_k - \hat{s}_k|^2 \big|\, \boldsymbol{H}_k\right] \\ &= \mathrm{argmin}_{\boldsymbol{f}_k}\, \mathrm{E}\left[\left|s_k - \boldsymbol{f}_k^{\mathrm{H}}\boldsymbol{H}_k^{\mathrm{H}}\sum_{i=1}^{K}\boldsymbol{p}_i s_i + \boldsymbol{f}_k^{\mathrm{H}}\boldsymbol{\eta}_k\right|^2 |\boldsymbol{H}_k\right].\end{aligned} \tag{5.7}$$

In order to compute the derivative of the $\mathrm{MSE}_k^{\mathrm{BC}}$ in (5.6) with respect to $\boldsymbol{f}_k^*$, we first rewrite the MSE as $\mathrm{tr}(\mathrm{MSE}_k^{\mathrm{BC}})$ and employ the results in Section C.6 of Appendix C to calculate the following derivative

$$\frac{\partial\, \mathrm{tr}\left(\mathrm{MSE}_k^{\mathrm{BC}}\right)}{\partial \boldsymbol{f}_k^*} = -\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{p}_k + \left(\boldsymbol{H}_k^{\mathrm{H}}\sum_{i=1}^{K}\boldsymbol{p}_i\boldsymbol{p}_i^{\mathrm{H}}\boldsymbol{H}_k + \boldsymbol{C}_{\boldsymbol{\eta}_k}\right)\boldsymbol{f}_k. \tag{5.8}$$

Thus, we get the optimal MSE filters equating this expression to zero, i.e.

$$\boldsymbol{f}_k^{\mathrm{MMSE}} = \left(\boldsymbol{H}_k^{\mathrm{H}}\sum_{i=1}^{K}\boldsymbol{p}_i\boldsymbol{p}_i^{\mathrm{H}}\boldsymbol{H}_k + \boldsymbol{C}_{\boldsymbol{\eta}_k}\right)^{-1}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{p}_k, \tag{5.9}$$

and the minimum MSE for user k is easily calculated plugging (5.9) into (5.6), i.e.

$$\begin{aligned}\mathrm{MMSE}_k^{\mathrm{BC}} &= 1 - \boldsymbol{f}_k^{\mathrm{MMSE,H}} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{p}_k \\ &= 1 - \boldsymbol{p}_k^{\mathrm{H}} \boldsymbol{H}_k \left(\boldsymbol{H}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i \boldsymbol{p}_i^{\mathrm{H}} \boldsymbol{H}_k + \boldsymbol{C}_{\boldsymbol{\eta}_k} \right)^{-1} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{p}_k. \end{aligned} \tag{5.10}$$

Consider now the vectors $\boldsymbol{c}_k = \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{p}_k \in \mathbb{C}^R$ and $\boldsymbol{b}_k = \boldsymbol{c}_k^{\mathrm{H}}$, and the matrix $\boldsymbol{X}_k \in \mathbb{C}^{R \times R} = \boldsymbol{H}_k^{\mathrm{H}} \sum_{i \neq k} \boldsymbol{p}_i \boldsymbol{p}_i^{\mathrm{H}} \boldsymbol{H}_k + \boldsymbol{C}_{\boldsymbol{\eta}_k}$. Then, $\mathrm{MMSE}_k^{\mathrm{BC}} = 1 - \boldsymbol{b}_k \left(\boldsymbol{X}_k + \boldsymbol{c}_k \boldsymbol{b}_k \right)^{-1} \boldsymbol{c}_k$ and its inverse after applying the matrix inversion lemma (see Section C.4 of Appendix C) reads as

$$\begin{aligned}\mathrm{MMSE}_k^{\mathrm{BC},-1} &= \left(1 + \boldsymbol{b}_k \boldsymbol{X}_k^{-1} \boldsymbol{c}_k \right)^{-1} \\ &= \left(1 + \boldsymbol{p}_k^{\mathrm{H}} \boldsymbol{H}_k \left(\boldsymbol{H}_k^{\mathrm{H}} \sum_{i \neq k} \boldsymbol{p}_i \boldsymbol{p}_i^{\mathrm{H}} \boldsymbol{H}_k + \boldsymbol{C}_{\boldsymbol{\eta}_k} \right)^{-1} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{p}_k \right)^{-1}. \end{aligned} \tag{5.11}$$

Applying the determinant to the last expression, we get

$$\begin{aligned}\det\left(\mathrm{MMSE}_k^{\mathrm{BC},-1} \right) &= \det \left(1 + \boldsymbol{p}_k^{\mathrm{H}} \boldsymbol{H}_k \left(\boldsymbol{H}_k^{\mathrm{H}} \sum_{i \neq k} \boldsymbol{p}_i \boldsymbol{p}_i^{\mathrm{H}} \boldsymbol{H}_k + \boldsymbol{C}_{\boldsymbol{\eta}_k} \right)^{-1} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{p}_k \right)^{-1} \\ &= \det \left(\mathbf{I}_R + \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{p}_k \boldsymbol{p}_k^{\mathrm{H}} \boldsymbol{H}_k \left(\boldsymbol{H}_k^{\mathrm{H}} \sum_{i \neq k} \boldsymbol{p}_i \boldsymbol{p}_i^{\mathrm{H}} \boldsymbol{H}_k + \boldsymbol{C}_{\boldsymbol{\eta}_k} \right)^{-1} \right)^{-1}, \end{aligned} \tag{5.12}$$

where the last expression is obtained by employing the Sylvester's theorem (see Section C.2 of Appendix C). Hence, we arrive at the following relationship between the rate and the MMSE as

$$R_k = -\log_2 \left(\mathrm{MMSE}_k^{\mathrm{BC}} \right). \tag{5.13}$$

Recall that we have made the assumption of perfect CSIR and imperfect CSIT. Due to that, it is not possible for us to employ the relationship of (5.13), since we are interested in certain average per-user rate constraints. Analogously, bearing in mind the partial CSIT, v, the MSE is the MIMO extension of that from Section 3.3.2 of Chapter 3, that is

$$\overline{\mathrm{MSE}}_k^{\mathrm{BC}} = \mathrm{E}\left[1 - 2\Re\left\{ \boldsymbol{f}_k^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{p}_k \right\} + \boldsymbol{f}_k^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i \boldsymbol{p}_i^{\mathrm{H}} \boldsymbol{H}_k \boldsymbol{f}_k + \boldsymbol{f}_k^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{\eta}_k} \boldsymbol{f}_k | \, v \right], \tag{5.14}$$

where the expectation is taken over all the possible channel realizations for a given realization of the partial CSI v. Remember that the pdf $f_{\boldsymbol{H}_k|v}(\boldsymbol{H}_k|v)$ is known at the transmitter.

Since the users know their corresponding channels $\boldsymbol{H}_k$, the optimal MSE receive filters from (5.9) hold. Then, the average MMSE at the BC is given by $\overline{\text{MMSE}}_k^{\text{BC}} = \text{E}[\text{MMSE}_k^{\text{BC}}|\, v]$, i.e.

$$\begin{aligned}\overline{\text{MMSE}}_k^{\text{BC}} &= \text{E}\left[1 - \boldsymbol{f}_k^{\text{MMSE,H}} \boldsymbol{H}_k^{\text{H}} \boldsymbol{p}_k |\, v\right] \\ &= \text{E}\left[1 - \boldsymbol{p}_k^{\text{H}} \boldsymbol{H}_k \left(\boldsymbol{H}_k^{\text{H}} \sum_{i=1}^{K} \boldsymbol{p}_i \boldsymbol{p}_i^{\text{H}} \boldsymbol{H}_k + \boldsymbol{C}_{\boldsymbol{\eta}_k}\right)^{-1} \boldsymbol{H}_k^{\text{H}} \boldsymbol{p}_k |\, v\right]. \end{aligned} \tag{5.15}$$

Our goal is to ensure minimum average rates ρ_k for all the users. Due to Jensen's inequality (see Appendix A) and the concavity of $\log_2(\cdot)$, we have $\log_2(\text{E}[x]) \geq \text{E}[\log_2(x)]$. Since the instantaneous data rate can be expressed as $R_k = -\log_2(\text{MMSE}_k^{\text{BC}})$, we have that

$$\text{E}[R_k|\, v] = \text{E}[-\log_2(\text{MMSE}_k^{\text{BC}})|\, v] \geq -\log_2(\text{E}[\text{MMSE}_k^{\text{BC}}|\, v]). \tag{5.16}$$

Based on the above discussion, it is possible to circumvent the difficult optimization of the average rates and focus on the average MSE instead. This way the new QoS constraints are expressed as maximum MSEs, as follows

$$\overline{\text{MMSE}}_k^{\text{BC}} \leq 2^{-\rho_k}, \tag{5.17}$$

and the average rate satisfies the inequality

$$\text{E}[R_k|\, v] \geq -\log_2\left(\text{E}\left[1 - \boldsymbol{p}_k^{\text{H}} \boldsymbol{H}_k \left(\boldsymbol{H}_k^{\text{H}} \sum_{i=1}^{K} \boldsymbol{p}_i \boldsymbol{p}_i^{\text{H}} \boldsymbol{H}_k + \boldsymbol{C}_{\boldsymbol{\eta}_k}\right)^{-1} \boldsymbol{H}_k^{\text{H}} \boldsymbol{p}_k |\, v\right]\right). \tag{5.18}$$

In other words, when ensuring an average MMSE, a minimum average rate is guaranteed. From our simulation results we observe that the gap between the two performance measures is small. An example of how tight the gap is can be found in Appendix A. Note that employing the average MMSE instead of the original average rate constraints establishes an upper bound for the minimum total transmit power.

Thus, the optimization problem (5.6) can be reformulated as follows

$$\min_{\{\boldsymbol{p}_k, \boldsymbol{f}_k\}_{k=1}^K} P_T = \sum_{k=1}^{K} \|\boldsymbol{p}_k\|_2^2 \quad \text{subject to} \quad \overline{\text{MMSE}}_k^{\text{BC}} \leq 2^{-\rho_k},\ \forall k. \tag{5.19}$$

Observe that the new problem formulation includes the optimization of the BC receive filters, in addition to the BC precoders. Therefore, we have to jointly optimize both filters to minimize the MSE. Since this problem is easier to solve than the former one, we will focus on minimizing the total transmit power for given average MMSE constraints for which the original average rate constraints hold. The optimization of the filters is studied in the ensuing section.

5.1.2. Exploiting MSE Duality

In the previous section the BC MMSE filters are found and given by (5.9). Then, we obtained the optimal filters for the problem reformulation presented in (5.19). However, we also have to design the optimal BC precoders $\boldsymbol{p}_k$.

Note that since the derivation of the receive filters $\boldsymbol{f}_k$ is straightforward, the precoders $\boldsymbol{p}_k$ are coupled and it is not possible to individually optimize them following the same procedure as for the receive filters. For this reason, we leverage the average MSE duality presented in [8] to avoid such inconvenience.

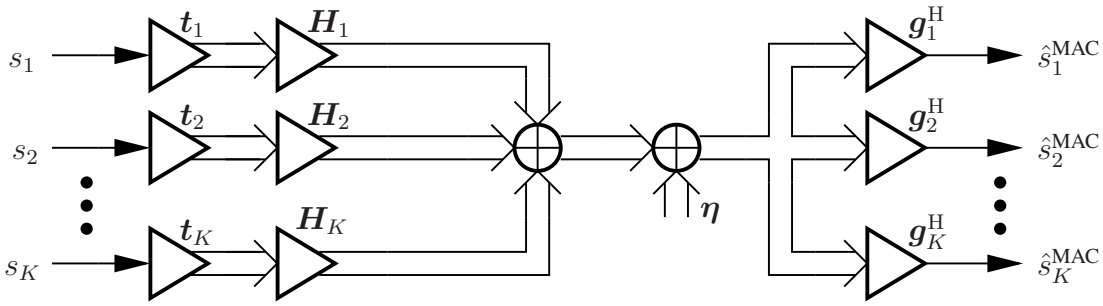


Figure 5.2: MIMO Multiple Access Channel.

Let us first introduce the dual MIMO MAC system model depicted in Fig. 5.2. The kth precoder is $\boldsymbol{t}_k \in \mathbb{C}^R$. The transmit signal propagates over the channel $\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2} \in \mathbb{C}^{N \times R}$, where $\boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2}$ is the Cholesky decomposition of $\boldsymbol{C}_{\boldsymbol{\eta}_k} = \boldsymbol{C}_{\boldsymbol{\eta}_k}^{\mathrm{H}/2} \boldsymbol{C}_{\boldsymbol{\eta}_k}^{1/2}$ (see Section C.5 of Appendix C). The received signal is perturbed by the AWGN$\boldsymbol{\eta} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_N)$ and filtered with the receiver $\boldsymbol{g}_k \in \mathbb{C}^N$ to get the estimated symbol of user k, i.e.

$$\hat{s}_k^{\mathrm{MAC}} = \boldsymbol{g}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{H}_i \boldsymbol{C}_{\boldsymbol{\eta}_i}^{-\mathrm{H}/2} \boldsymbol{t}_i s_i + \boldsymbol{\eta}. \tag{5.20}$$

Note that the MAC receivers $\boldsymbol{g}_k$ depend on the partial CSI v, whereas the MAC precoders $\boldsymbol{t}_k$ are functions of the current channel state. The MSE for the MISO system model was presented in Section 3.2 of Chapter 3 and is easily extended to the MIMO scenario as follows

$$\mathrm{MSE}_k^{\mathrm{MAC}} = 1 - 2\Re\left\{\boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2} \boldsymbol{t}_k\right\} + \boldsymbol{g}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{H}_i \boldsymbol{C}_{\boldsymbol{\eta}_i}^{-\mathrm{H}/2} \boldsymbol{t}_i \boldsymbol{t}_i^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{\eta}_i}^{-1/2} \boldsymbol{H}_i^{\mathrm{H}} \boldsymbol{g}_k + \|\boldsymbol{g}_k\|_2^2 . \tag{5.21}$$

Note that the previous expression holds for both perfect CSIT and CSIR. In our case, the average MSE for imperfect CSIR in the dual MAC, $\overline{\mathrm{MSE}}_k^{\mathrm{MAC}}$, has to be computed as the

expectation over the possible channel realizations for given v, i.e.

$$\overline{\mathrm{MSE}}_k^{\mathrm{MAC}} = \mathrm{E}\left[1 - 2\Re\left\{\boldsymbol{g}_k^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2}\boldsymbol{t}_k\right\} + \sum_{i=1}^{K}\left|\boldsymbol{g}_k^{\mathrm{H}}\boldsymbol{H}_i\boldsymbol{C}_{\boldsymbol{\eta}_i}^{-\mathrm{H}/2}\boldsymbol{t}_i\right|^2 + \|\boldsymbol{g}_k\|_2^2 \,|\, v\right]. \quad (5.22)$$

We now present the conversion formulas to switch from the BC to the MAC, and vice-versa, preserving both the average MSE and the total transmit power. Following the same procedure described in Section 3.3, we define the relationship between the BC and the MAC for given MAC filters as

$$\boldsymbol{p}_k = \alpha_k\boldsymbol{g}_k \quad \text{and} \quad \boldsymbol{f}_k = \alpha_k^{-1}\boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2}\boldsymbol{t}_k, \quad (5.23)$$

with $\{\alpha_k\}_{k=1}^K \in \mathbb{R}^+$. Accordingly, the average MSE in the BC given by (5.14), $\overline{\mathrm{MSE}}_k^{\mathrm{BC}}$, is rewritten as

$$\overline{\mathrm{MSE}}_k^{\mathrm{BC}} = \mathrm{E}\left[1 - 2\Re\left\{\boldsymbol{t}_k^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{\eta}_k}^{-1/2}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{g}_k\right\} + \alpha_k^{-2}\|\boldsymbol{t}_k\|_2^2 + \sum_{i=1}^{K}\frac{\alpha_i^2}{\alpha_k^2}\left|\boldsymbol{g}_i^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2}\boldsymbol{t}_k\right|^2 \,|\, v\right]. \quad (5.24)$$

Equating this expression to (5.22), we get the K equalities that have to be fulfilled. Let us define the vector containing the scalar factors α_k^2, i.e. $\boldsymbol{a} = [\alpha_1^2, \ldots, \alpha_K^2]^{\mathrm{T}}$, and the vector $\varsigma_i = \mathrm{E}[\|\boldsymbol{t}_i\|^2 \,|\, v] \in \mathbb{R}_0^+$, which contains the transmit power for all the users. Such definition enable us to rewrite the equalities in matrix form as

$$\boldsymbol{\Gamma}\boldsymbol{a} = \boldsymbol{\varsigma}, \quad (5.25)$$

where the entries of $\boldsymbol{\Gamma} \in \mathbb{R}^{K\times K}$ are given by

$$[\boldsymbol{\Gamma}]_{k,j} = \begin{cases} \sum_{i\neq k}\mathrm{E}[|\boldsymbol{g}_k^{\mathrm{H}}\boldsymbol{H}_i\boldsymbol{C}_{\boldsymbol{\eta}_i}^{-\mathrm{H}/2}\boldsymbol{t}_i|^2 \,|\, v] + \|\boldsymbol{g}_k\|_2^2 & j = k, \\ -\mathrm{E}[|\boldsymbol{g}_j^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2}\boldsymbol{t}_k|^2 \,|\, v] & j \neq k. \end{cases}$$

Note that $\boldsymbol{\Gamma}$ is non-singular since it is diagonally dominant (see Section 3.3 of Chapter 3 for further details). Additionally, $\boldsymbol{\Gamma}$ has positive diagonal and non-positive off-diagonal entries. Thus, $\boldsymbol{\Gamma}^{-1}$ has non-negative entries and the scalar factors in $\boldsymbol{a}$ read

$$\boldsymbol{a} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\varsigma}. \quad (5.26)$$

Note that since $\alpha_k^2 \geq 0$, $\forall k$, we can get $\{\alpha_k\}_{k=1}^K \in \mathbb{R}^+$, as desired.

Analogously, the conversion from MAC to BC is now addressed. This is necessary for the *Alternate Optimization* (AO) procedure employed to optimize the filters that we will explain in the following. Therefore, for given BC filters, the MAC filters achieving the same average MSEs and employing identical total transmit power can be found [37]

(see also the MISO case in Section 3.3). The duality can be obtained in the same way as for the MAC to BC conversion using the following relationships

$$\boldsymbol{g}_k = \beta_k^{-1}\boldsymbol{p}_k \quad \text{and} \quad \boldsymbol{t}_k = \beta_k \boldsymbol{C}_{\boldsymbol{\eta}_k}^{\mathrm{H}/2}\boldsymbol{f}_k. \tag{5.27}$$

Now, substituting (5.27) into the average MSE expression for the dual MAC of (5.22), we get

$$\overline{\mathrm{MSE}}_k^{\mathrm{MAC}} = \mathrm{E}\left[1 - 2\Re\{\boldsymbol{p}_k^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{f}_k\} + \sum_{i=1}^{K}\frac{\beta_i^2}{\beta_k^2}\left|\boldsymbol{p}_k^{\mathrm{H}}\boldsymbol{H}_i\boldsymbol{f}_i\right|^2 + \beta_k^{-2}\left\|\boldsymbol{p}_k\right\|_2^2 \mid v\right]. \tag{5.28}$$

Then, equating (5.28) to the $\overline{\mathrm{MSE}}_k^{\mathrm{BC}}$ expression given by (5.14) and multiplying both sides by β_k^2, we get

$$\beta_k^2\,\mathrm{E}\left[\sum_{i=1}^{K}|\boldsymbol{f}_k^{\mathrm{H}}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{p}_i|^2|\, v\right] + \beta_k^2\,\mathrm{E}\left[\boldsymbol{f}_k^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{\eta}_k}\boldsymbol{f}_k|\, v\right] = \mathrm{E}\left[\sum_{i=1}^{K}\beta_i^2\left|\boldsymbol{p}_k^{\mathrm{H}}\boldsymbol{H}_i\boldsymbol{f}_i\right|^2 \mid v\right] + \left\|\boldsymbol{p}_k\right\|_2^2. \tag{5.29}$$

We now equate the average MSEs in the BC and the MAC for the K users, and rewrite them in matrix form as

$$\boldsymbol{\Omega}\boldsymbol{b} = \boldsymbol{\tau}, \tag{5.30}$$

where $\boldsymbol{b} = [\beta_1^2, \ldots, \beta_K^2]^{\mathrm{T}}$ contains the scalar factors for the conversion, and $\boldsymbol{\tau} = \left[\left\|\boldsymbol{p}_1\right\|_2^2, \ldots, \left\|\boldsymbol{p}_K\right\|_2^2\right]^{\mathrm{T}}$ includes the transmit power for all the users. Thus, the entries of $\boldsymbol{\Omega}$ are given by

$$[\boldsymbol{\Omega}]_{k,j} = \begin{cases} \sum_{i\neq k}\mathrm{E}\left[\left|\boldsymbol{f}_k^{\mathrm{H}}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{p}_i\right|^2 \mid v\right] + \mathrm{E}\left[\boldsymbol{f}_k^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{\eta}_k}\boldsymbol{f}_k|\, v\right] & j = k, \\ -\,\mathrm{E}\left[\left|\boldsymbol{f}_j^{\mathrm{H}}\boldsymbol{H}_j^{\mathrm{H}}\boldsymbol{p}_k\right|^2 \mid v\right] & j \neq k. \end{cases} \tag{5.31}$$

Note that, just like $\boldsymbol{\Gamma}$ for the MAC to BC conversion, $\boldsymbol{\Omega}$ is non-singular since it is diagonally dominant. Additionally, it has positive diagonal and non-positive off-diagonal entries. Thus, $\boldsymbol{\Omega}^{-1}$ has non-negative entries and the resulting $\{\beta_k^2\}_{k=1}^K$ from $\boldsymbol{b} = \boldsymbol{\Omega}^{-1}\boldsymbol{\tau}$ are non-negative. In other words, we can always find $\beta_k \in \mathbb{R}^+$ such that $\overline{\mathrm{MSE}}_k^{\mathrm{BC}} = \overline{\mathrm{MSE}}_k^{\mathrm{MAC}}$, $\forall k$.

From the average MSE BC/MAC duality result, we conclude that the BC average MMSE in the optimization problem of (5.19) can be achieved in the dual MAC by exploiting the relationship of (5.27) for MIMO systems. Moreover, it has been shown that the total transmit power employed in both domains remains unchanged. Therefore, we formulate the optimization problem in the dual MAC as follows

$$\min_{\{\boldsymbol{g}_k,\boldsymbol{t}_k\}_{k=1}^K} P_T^{\mathrm{MAC}} = \sum_{k=1}^{K}\mathrm{E}\left[\left\|\boldsymbol{t}_k\right\|_2^2 \mid v\right] \quad \text{subject to} \quad \overline{\mathrm{MMSE}}_k^{\mathrm{MAC}} \leq 2^{-\rho_k},\ \forall k, \tag{5.32}$$

where the minimum total power required to fulfill the QoS restrictions P_T^{MAC} is the same as P_T in (5.19).

Recall that the BC precoders $\boldsymbol{p}_k$ are coupled. To avoid this difficulty, we can equivalently compute the optimal MSE receive filters in the dual MAC such that $\boldsymbol{g}_k$ are individually calculated as the MMSE filters, $\boldsymbol{g}_k^{\text{MMSE}}$, i.e.

$$\frac{\partial \operatorname{tr}\left(\overline{\text{MSE}}_k^{\text{MAC}}\right)}{\partial \boldsymbol{g}_k^*} = -\operatorname{E}\left[\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\text{H}/2} \boldsymbol{t}_k |\, v\right] + \left(\sum_{i=1}^{K} \operatorname{E}\left[\boldsymbol{H}_i \boldsymbol{C}_{\boldsymbol{\eta}_i}^{-\text{H}/2} \boldsymbol{t}_i \boldsymbol{t}_i^{\text{H}} \boldsymbol{C}_{\boldsymbol{\eta}_i}^{-1/2} \boldsymbol{H}_i^{\text{H}} |\, v\right] + \mathbf{I}_N\right) \boldsymbol{g}_k. \tag{5.33}$$

Note that to compute the derivative we take into account the linear property of the expectation $\operatorname{E}[\cdot]$ and the equality $\operatorname{tr}(\overline{\text{MSE}}_k^{\text{MAC}}) = \overline{\text{MSE}}_k^{\text{MAC}}$. Then, it is possible to employ the results shown in Section C.6 of Appendix C. After equating the last expression to zero, we obtain the optimal MSE MAC receive filters

$$\boldsymbol{g}_k^{\text{MMSE}} = \left(\sum_{i=1}^{K} \operatorname{E}\left[\boldsymbol{H}_i \boldsymbol{C}_{\boldsymbol{\eta}_i}^{-\text{H}/2} \boldsymbol{t}_i \boldsymbol{t}_i^{\text{H}} \boldsymbol{C}_{\boldsymbol{\eta}_i}^{-1/2} \boldsymbol{H}_i^{\text{H}} |\, v\right] + \mathbf{I}_N\right)^{-1} \operatorname{E}\left[\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\text{H}/2} \boldsymbol{t}_k |\, v\right]. \tag{5.34}$$

Both problem formulations, the corresponding to the BC in (5.19) and the one to the MAC in (5.32), allow for a simple computation of the optimal receivers although the precoders fulfilling the QoS constraints are difficult to find. Therefore, we propose an AO where the BC receivers are found via (5.9) for given precoders $\boldsymbol{p}_k$, and the BC precoders (including the power allocation) are computed in the dual MAC for given $\boldsymbol{f}_k$, as we will demonstrate in the ensuing section.

5.1.3. Power Allocation via Interference Functions

In the previous section we have proposed a method to optimally design the transmit and receive MSE filters by means of AO. Both filters are individually optimized for every user, since the BC/MAC MSE duality allows us to update the transmit and receive filters for a certain user, without any changing of the MSE of the other users. Observe that the optimization problem (5.32) aims at minimizing the total transmit power subject to certain average MMSE constraints for all the users. To achieve those targets, finding the optimal filters as before is not enough and an adaptation of the transmit power is necessary. Such powers are referred to as power allocation. Finding the minimum total power is a difficult task, since increasing the power for one the users is translated into larger interference for the other ones. Thus, we resort to the framework proposed in [38] to find the minimum total power required to achieve the QoS restrictions.

The power allocation requires the computation of the expectations in (5.22) and (5.34). Note that since the users perfectly know their channel $\boldsymbol{H}_k$, the optimal MAC transmit filters $\boldsymbol{t}_k$ depend on $\boldsymbol{H}_k$, whereas the receive filters depend on the imperfect CSI v. We propose to approximate the expectations performing Monte Carlo numerical integration employing M realizations of the random variable given by the imperfect CSI v, i.e. the pdfs $f_{\boldsymbol{H}_k|v}(\boldsymbol{H}_k|v)$. Assuming the channel ergodicity, this approximation collapses to the real expectation when $M \to \infty$.

Recall the error model from (5.3). For given CSI v we know the channel expectations and the error covariance matrices for all the users, i.e. $\{\bar{\boldsymbol{H}}_k, \boldsymbol{C}_k\}_{k=1}^K$. Therefore, the channel realizations are generated as

$$\hat{\boldsymbol{H}}_k^{(m)} = \bar{\boldsymbol{H}}_k + \tilde{\boldsymbol{H}}_k^{(m)}, \tag{5.35}$$

where $\tilde{\boldsymbol{H}}_k^{(m)} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{C}_{\tilde{\boldsymbol{H}}_k})$.

Let us define the matrix $\boldsymbol{\Theta}_k = [\boldsymbol{H}_k^{(1)}\boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2}, \ldots, \boldsymbol{H}_k^{(M)}\boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2}] \in \mathbb{C}^{N \times RM}$ containing the M channel realizations corresponding to the user k. We split up the MAC precoders $\boldsymbol{t}_k^{(m)}$ for every channel realization into the average transmit power ξ_k and the normalized transmit filters $\boldsymbol{\tau}_k^{(m)}$, such that $\sqrt{\xi_k}\boldsymbol{\tau}_k^{(m)} = \boldsymbol{t}_k^{(m)}$, with $\sum_{m=1}^M ||\boldsymbol{\tau}_k^{(m)}||_2^2 = M$ and $\xi_k = \frac{1}{M}\sum_{m=1}^M ||\boldsymbol{t}_k^{(m)}||_2^2$. For the matrix form computation, we introduce an additional matrix $\boldsymbol{T}_k \in \mathbb{C}^{RM \times R}$, that is

$$\boldsymbol{T}_k = 1/\sqrt{\xi_k}\,\mathrm{blockdiag}\left(\boldsymbol{t}_k^{(1)}, \ldots, \boldsymbol{t}_k^{(M)}\right). \tag{5.36}$$

Employing the new notation, and taking into account that $\boldsymbol{T}_k\boldsymbol{1} = [\boldsymbol{\tau}_k^{(1),\mathrm{T}}, \ldots, \boldsymbol{\tau}_k^{(M),\mathrm{T}}]^{\mathrm{T}}$ we approximate the average MAC MSE from (5.22) as

$$\overline{\mathrm{MSE}}_k^{\mathrm{MAC}} = 1 - \frac{2}{M}\Re\left\{\boldsymbol{g}_k^{\mathrm{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\boldsymbol{1}\sqrt{\xi}_k\right\} + \|\boldsymbol{g}_k\|_2^2 + \frac{1}{M}\boldsymbol{g}_k^{\mathrm{H}}\sum_{i=1}^K \xi_i\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}}\boldsymbol{g}_k. \tag{5.37}$$

Consequently, we obtain an approximation of the MAC receive filters using the previous matrices as follows

$$\boldsymbol{g}_k^{\mathrm{MMSE}} = \left(\frac{1}{M}\sum_{i=1}^K \xi_i\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}} + \mathbf{I}_N\right)^{-1}\frac{1}{M}\sqrt{\xi_k}\boldsymbol{\Theta}_k\boldsymbol{T}_k\boldsymbol{1}. \tag{5.38}$$

Thus, the average MMSE in the dual MAC is also approximated by substituting (5.38) into (5.37), which leads to

$$\overline{\mathrm{MMSE}}_k^{\mathrm{MAC}} = 1 - \frac{1}{M^2}\xi_k\boldsymbol{1}^{\mathrm{T}}\boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{\Theta}_k^{\mathrm{H}}\left(\frac{1}{M}\sum_{i=1}^K \xi_i\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}} + \mathbf{I}_N\right)^{-1}\boldsymbol{\Theta}_k\boldsymbol{T}_k\boldsymbol{1}. \tag{5.39}$$

Observe that the formulation of the MIMO MAC average MMSE is very close to that for the MISO MAC average MMSE in (4.57) (Section 4.2.3 of Chapter 4). However, the dimension of the matrices $\boldsymbol{T}_k$ and $\boldsymbol{\Theta}_k$ is different from that of (4.57), and the vector $\mathbf{1}$ did not appeared in (4.57). Hence, the developments to be directly applied from MISO to MIMO scenarios will not be repeated again but referenced instead in the following.

As we previously mentioned, to design the power allocation we employ the so-called *standard* interference functions presented in [38]. A description of this framework can be found in Appendix B.

First, consider that the MAC receive filters $\boldsymbol{g}_k$ resulting from the updating as MMSE filters are kept fixed. To allow for an adaptation of such filters, when the power allocation is updated, additional scalar receive filters, denoted as r_k, are introduced. Replacing $\boldsymbol{g}_k$ by $r_k\tilde{\boldsymbol{g}}_k$ in (5.37) leads to

$$\overline{\mathrm{MSE}}_k^{\mathrm{MAC}} = 1-\frac{2}{M}\Re\left\{r_k^*\tilde{\boldsymbol{g}}_k^{\mathrm{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}\sqrt{\xi_k}\right\}+\frac{1}{M}\,|r_k|^2\sum_{i=1}^{K}\xi_i\tilde{\boldsymbol{g}}_k^{\mathrm{H}}\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}}\tilde{\boldsymbol{g}}_k+|r_k|^2\,\|\tilde{\boldsymbol{g}}_k\|_2^2\,. \tag{5.40}$$

The corresponding optimal scalar receivers, which minimize (5.40), are given by (see the derivation in Section 4.2.3)

$$r_k^{\mathrm{MMSE}} = \frac{\frac{1}{M}\tilde{\boldsymbol{g}}_k^{\mathrm{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}\sqrt{\xi_k}}{\frac{1}{M}\sum_{i=1}^{K}\xi_i\tilde{\boldsymbol{g}}_k^{\mathrm{H}}\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}}\tilde{\boldsymbol{g}}_k+\|\tilde{\boldsymbol{g}}_k\|_2^2}. \tag{5.41}$$

Substituting r_k^{MMSE} in (5.40) gives the $\overline{\mathrm{MMSE}}_{k,\mathrm{scalar}}^{\mathrm{MAC}}$. With the definition of

$$z_k(\boldsymbol{\xi}) = \frac{1}{M}\sum_{i=1}^{K}\xi_i\tilde{\boldsymbol{g}}_k^{\mathrm{H}}\boldsymbol{\Theta}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Theta}_i^{\mathrm{H}}\tilde{\boldsymbol{g}}_k - \frac{\xi_k}{M^2}\left|\tilde{\boldsymbol{g}}_k^{\mathrm{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}\right|^2+\|\tilde{\boldsymbol{g}}_k\|_2^2,$$

the minimum MSE reads as

$$\overline{\mathrm{MMSE}}_{k,\mathrm{scalar}}^{\mathrm{MAC}} = \frac{1}{\xi_k}\,\frac{1}{\frac{1}{\xi_k}+\frac{1}{M^2}\,\frac{1}{z_k(\boldsymbol{\xi})}\left|\tilde{\boldsymbol{g}}_k^{\mathrm{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}\right|^2}. \tag{5.42}$$

Note that this expression is equivalent to that of (5.39) when the MMSE MAC receive filters (5.38) are used, as previously shown for the MISO scenario in the previous chapter (see Section 4.2.3).

In summary, we presented the optimal receive filter $\boldsymbol{g}_k^{\mathrm{MMSE}}$ and then introduced the normalization variable r_k for power updating in the AO. With this new notation, a new MMSE expression is obtained for given normalized MAC precoders $\boldsymbol{T}_k$, i.e., for given normalized BC receivers. We now introduce the interference function $J_k(\boldsymbol{\xi})$ depending on $\tilde{\boldsymbol{g}}_k$ and r_k, for given $\boldsymbol{T}_k$, $\forall k$, as follows

$$J_k(\boldsymbol{\xi}) = \left(\frac{1}{\xi_k}+\frac{1}{M^2}\left|\tilde{\boldsymbol{g}}_k^{\mathrm{H}}\boldsymbol{\Theta}_k\boldsymbol{T}_k\mathbf{1}\right|^2 z_k^{-1}\left(\boldsymbol{\xi}\right)\right)^{-1}. \tag{5.43}$$

Note that by using this function the optimal normalization variables r_k^{MMSE} are implicitly used.

We propose to incorporate the functions $J_k(\boldsymbol{\xi})$ to find the optimal $\tilde{\boldsymbol{g}}_k$ and r_k for given MAC precoders $\boldsymbol{T}_k$, taking advantage of the properties of *standard* interference functions. Consequently, we need to check if the required conditions for the function $J_k(\boldsymbol{\xi})$ to be *standard* are satisfied. The proof is analogous to that from Section 4.2.3 and we can conclude that $J_k(\boldsymbol{\xi})$ satisfies positivity, scalability, and monotonicity.

For the MISO scenario we had defined an additional interference function $I_k(\boldsymbol{\xi})$, which only depended on $\boldsymbol{g}_k$ and $\boldsymbol{T}_k$ since the scalar normalization factor r_k was not included. The two interference functions, $I_k(\boldsymbol{\xi})$ and $J_k(\boldsymbol{\xi})$, provided the same result if the receive filters were properly updated in the fixed point iteration of $J_k(\boldsymbol{\xi})$. However, the function $I_k(\boldsymbol{\xi})$ required additional computational complexity, and for that reason will not be considered for the MIMO scenario.

Consider now the problem formulation in the dual MAC obtained in (5.32). The QoS average rate constraints are lower bounded by the new average MMSE ones. Recall that (5.42) is equivalent to (5.39) when the MMSE MAC receive filters are used, as in (5.38). Moreover, the interference function $J(\boldsymbol{\xi})$ is given by $J(\boldsymbol{\xi}) = \xi_k \overline{\text{MMSE}}_{k,\text{scalar}}^{\text{MAC}}$. Combining these facts, we get the following reformulation of the optimization problem for given MAC precoders

$$\min_{\boldsymbol{\xi},\{\tilde{\boldsymbol{g}}_k\}_{k=1}^K} \sum_{i=1}^{K} \xi_i \quad \text{subject to} \quad J_k\left(\boldsymbol{\xi}\right)/\xi_k \leq 2^{-\rho_k},\ \forall k. \tag{5.44}$$

Due to the properties of $J_k(\boldsymbol{\xi})$, $\boldsymbol{J}(\boldsymbol{\xi}) = [J_1(\boldsymbol{\xi}), \ldots, J_K(\boldsymbol{\xi})]$ is a *standard* interference function. Therefore, it is possible to take advantage of the useful characteristics presented in [38] and its extension in [70]. This way the QoS average rate requirements are approximated by the inequality $\boldsymbol{\xi} \geq \boldsymbol{Q}^{-1}\boldsymbol{J}(\boldsymbol{\xi})$, with $\boldsymbol{Q} = \text{diag}(2^{-\rho_1}, \ldots, 2^{-\rho_K})$. Then, the fixed point iteration $\boldsymbol{\xi}^{(\ell)} = \boldsymbol{Q}^{-1}\boldsymbol{J}(\boldsymbol{\xi}^{(\ell-1)})$ converges to the global optimum of (5.44). The proof for the joint convergence in $\boldsymbol{\xi}$ and $\tilde{\boldsymbol{g}}_k$ was shown in Section 4.2.3.

Taking into account the characteristics of the *standard* interference functions, the power allocation converges to an unique solution, $\boldsymbol{\xi}^{\text{opt}}$. On the contrary, the MAC receive filters convergence is not unique. For example, another optimal receive filter is given by the product of $\boldsymbol{g}_k^{\text{opt}}$ by $\mathrm{e}^{j\phi}$.

So far the optimal MAC receivers including the power allocation i.e., the BC precoders, are found. To complete the AO, the BC receivers $\boldsymbol{f}_k$ are computed in the BC via (5.9). Therefore, the AO process is complete and leads us to the algorithm proposed in the following section.

5.1.4. Algorithmic Solution

In this section, the algorithm referred to as Algorithm 5.1: PM.MIMO is presented. Such an algorithm implements the methods proposed to perform the AO that solves the optimization problems (5.32) and (5.19). The two formulations were shown to be equivalent by the BC/MAC MSE duality, and the solution obtained via (5.44). The AO provides an upper bound for the total transmit power needed to fulfill the original restrictions from the power minimization based on the average rate constraints (5.5).

Algorithm 5.1: PM.MIMO. Power Minimization by AO in the MIMO BC

1: $\ell \leftarrow 0$, initialize $\boldsymbol{p}_i^{(0)}$, $\forall i$
2: **repeat**
3: $\quad \ell \leftarrow \ell + 1$
4: $\quad$ **for** $m = 1$ to M **do**
5: $\quad\quad \boldsymbol{f}_k^{(\ell-1,m)} \leftarrow$ update BC receiver using (5.9), $\forall k$
6: $\quad$ **end for**
7: $\quad \boldsymbol{t}_k^{(\ell-1,m)}, \tilde{\boldsymbol{g}}_k^{(\ell-1)}, \forall k, \forall m \leftarrow$ BC-to-MAC conversion (see Section 5.1.2)
8: $\quad \xi_k^{(\ell-1)} \leftarrow \frac{1}{M} \sum_{m=1}^{M} ||\boldsymbol{t}_k^{(\ell-1,m)}||_2^2$, $\forall k$
9: $\quad \boldsymbol{\tau}_k^{(\ell-1,m)} \rightarrow 1/\sqrt{\xi_k^{(\ell-1)} \boldsymbol{t}_k^{(\ell-1,m)}}$, $\forall k$
10: $\quad \boldsymbol{T}_k^{(\ell)} \leftarrow \text{blockdiag}(\boldsymbol{\tau}_k^{(\ell-1,1)}, \ldots, \boldsymbol{\tau}_k^{(\ell-1,M)})$, $\forall k$
11: $\quad \boldsymbol{J}(\boldsymbol{\xi}^{(\ell-1)}) \leftarrow$ update interference function (5.43)
12: $\quad \xi_k^{(\ell)} \leftarrow 2^{\rho_k} J_k(\boldsymbol{\xi}^{(\ell-1)})$, $\forall k$
13: $\quad \tilde{\boldsymbol{g}}_k^{(\ell)} \leftarrow$ update MAC receiver (5.38), $\forall k$
14: $\quad$ **for** $m = 1$ to M **do**
15: $\quad\quad \boldsymbol{t}_k^{(\ell,m)} \leftarrow \sqrt{\xi_k^{(\ell)}} [\boldsymbol{T}_k^{(\ell)}]_{(m-1)R+1:mR,m}$ update MAC precoder, $\forall k$
16: $\quad$ **end for**
17: $\quad \boldsymbol{p}_k^{(\ell)}, \boldsymbol{f}_k^{(\ell,m)}, \forall k, \forall m \leftarrow$ MAC-to-BC conversion (see Section 5.1.2)
18: **until** $||\boldsymbol{\xi}^{(\ell)} - \boldsymbol{\xi}^{(\ell-1)}||_1 \leq \delta$

Algorithm 5.1: PM.MIMO is the extension to the MIMO BC of the Algorithm 4.3: PM.MISO.ICSI.2 presented in the previous chapter for the MISO BC. Fortunately, we have shown that the same procedures can be applied for the MIMO and the MISO problem formulations. The main difference lies in the dimension of the BC receivers that slightly changes the pseudocode. Note that, since equivalent properties can be demonstrated for the more general MIMO case, we do not include a MIMO version of Algorithm 4.2: PM.MISO.ICSI.1 because of the convergence issues discussed in Section 4.2.4 of Chapter 4.

The precoders are randomly initialized in the line 1. Note that the initialization does

not affect to the final solution since the minimum total transmit power has been proven to be unique. Next, the BC receive filters are updated to the MMSE filters for every channel realization (recall that there exists perfect CSIR). After that, we switch to the MAC by using the relationships in (5.23) for given BC transmit and receive filters (line 7). Then, the total average power is computed in the line 8, allowing to obtain the normalized MAC precoders in the line 9. The line 10 computes the matrix collecting the normalized precoders for every channel realization.

In the line 11 the new value for the interference function is calculated and used afterwards in the line 12 to find the new power allocation. The MAC receive filters have to be updated accordingly in the line 13 to preserve the equivalence between the conventional MMSE and the version including the normalization factor used to define the interference function $J(\boldsymbol{\xi})$. This new power allocation is included in the MAC precoders in the line 15, and then, the updated versions of the corresponding BC filters are computed in the line 17 via the relationships (5.27). We compare the difference between the total power obtained in the current iteration and that achieved in the previous one with a certain threshold, which depends on the desired accuracy, to decide whether we have reached convergence or not (see the line 18).

Just like Algorithm 4.3: PM.MISO.ICSI.2 from Section 4.2.4 of Chapter 4, Algorithm 5.1: PM.MIMO converges to the optimum of (5.19). Note that the filters update reduce the MMSE of every iteration. Moreover, the power allocation adaptation reduces the total transmit power required to achieve the MMSE constraints. Since both the MMSE and the total transmit power are lower bounded by zero, it is straightforward to see that the proposed algorithm converges. However, as previously mentioned, the solution is an upper bound of the required total transmit power necessary to fulfill the average rate constraints, since the average MMSE restrictions are more stringent.

5.1.5. Simulation Results

In this section we present some numerical results to illustrate the performance of the proposed algorithm. The simulation setup consists on a BS with $N = 4$ transmit antennas and $K = 4$ users equipped with $R = 2$ antennas each. The noise is considered to be equally distributed for all the users, $\boldsymbol{\eta}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{\eta}_k})$, with $\boldsymbol{C}_{\boldsymbol{\eta}_k} = \mathbf{I}_R, \forall k$. The MIMO channel realizations are generated taking into account the model given by (5.3). Accordingly, $\boldsymbol{H}_k^{(m)} = \bar{\boldsymbol{H}}_k + \tilde{\boldsymbol{H}}_k^{(m)}$, with $m = \{1, \ldots, M\}$, $M = 1000$, and $\bar{\boldsymbol{H}}_k$ being the imperfect CSI error modeled as $\tilde{\boldsymbol{H}}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_N)$.

Considering this scenario, we impose the average rate restrictions $\rho_1 = 0.5146, \rho_2 = 0.737, \rho_3 = 1$, and $\rho_4 = 0.2345$ bits per channel use, respectively. The corresponding lower bounds for the average MMSE are $\varepsilon_k = 2^{-\rho_k}, \forall k$, i.e., $\varepsilon_1 = 0.7, \varepsilon_2 = 0.6, \varepsilon_3 = 0.5$, and $\varepsilon_4 = 0.85$. The gap between the total powers achieved at a given iteration and at

the previous one is compared to the threshold, denoted by δ, which is set to 10^{-4}. Initial precoders are randomly generated as previously discussed.

Figure 5.3 depicts the evolution of the average rates during the execution of Algorithm 5.1: PM.MIMO. Note that the QoS are not strictly fulfilled but a little gap arising from the average MMSE approximation of the targets has appeared. On the other hand, Fig. 5.4 shows the average MMSEs corresponding to the previous average rates. Observe how the restrictions are now fulfilled with the equality, since the aim is to minimize the total transmit power. Then, the average MMSE targets $\varepsilon_1 = 0.7, \varepsilon_2 = 0.6, \varepsilon_3 = 0.5$, and $\varepsilon_4 = 0.85$ match the exact values. Finally, the evolution of the total power is given in Fig. 5.5. The total power starts over 9 dB for random initial precoders and reduces gradually to about 0 dB after 25 iterations.

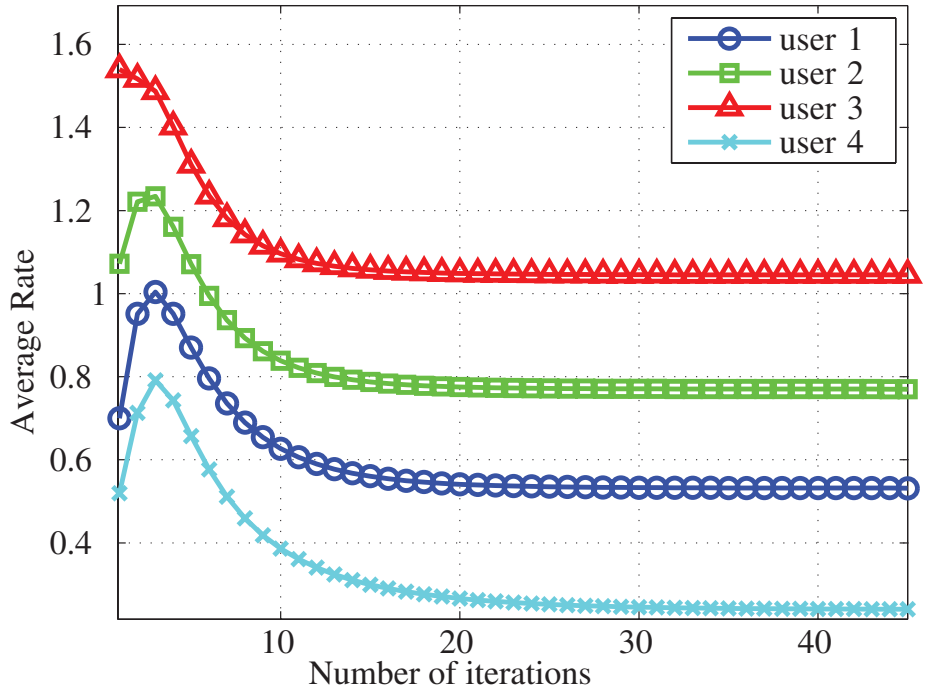


Figure 5.3: Power Minimization in the MIMO BC with Imperfect CSI: Rate vs. Number of Iterations.

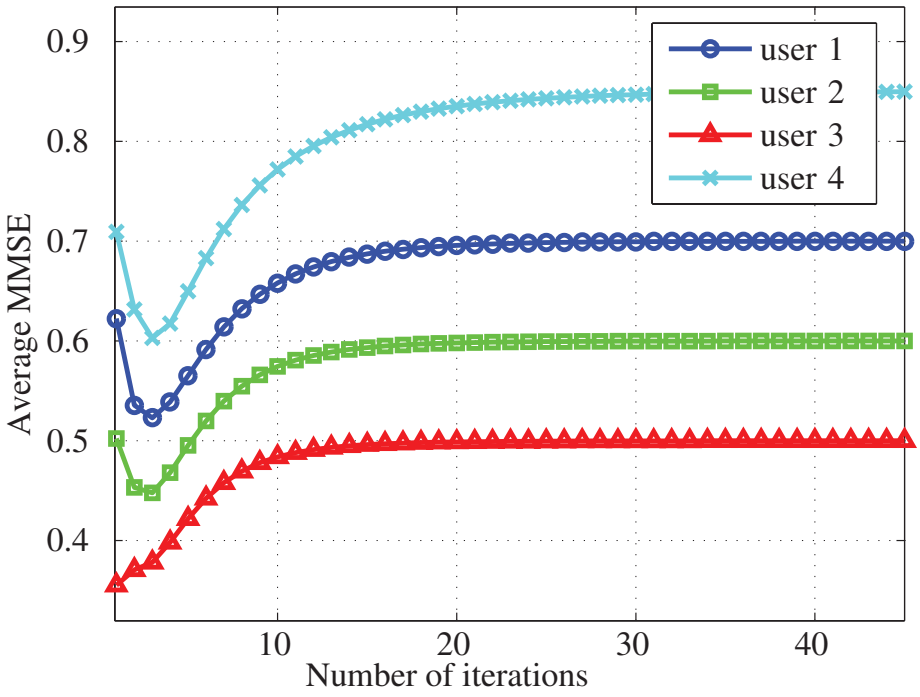


Figure 5.4: Power Minimization in the MIMO BC with Imperfect CSI: MMSE vs. Number of Iterations.

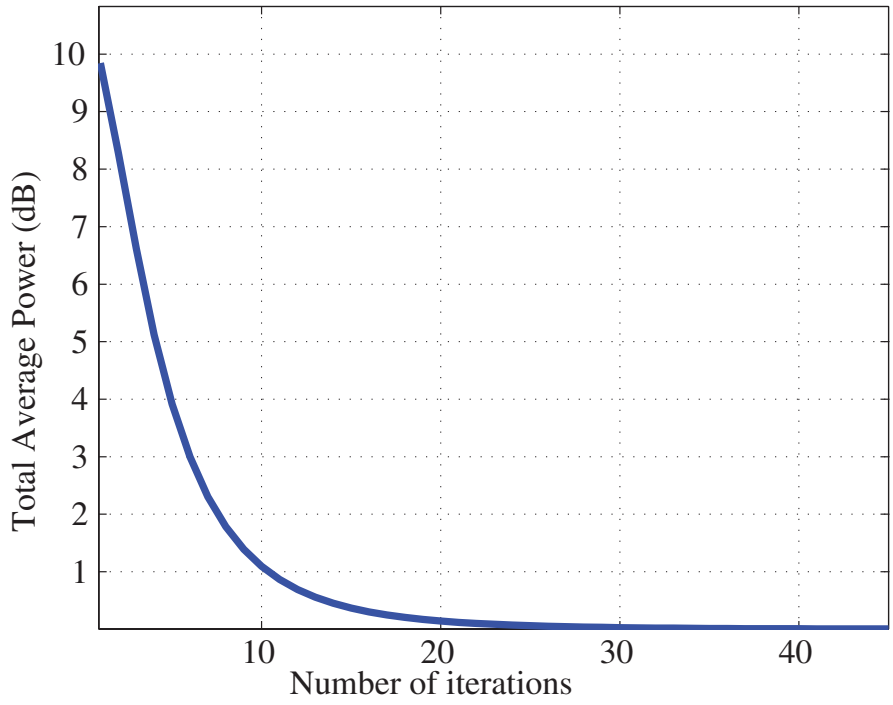


Figure 5.5: Power Minimization in the MIMO BC with Imperfect CSI: Power vs. Number of Iterations.

5.2. Feasibility Region

In this section we analyze the feasibility region of the power minimization problem given by (5.32). The feasibility of the average MMSE targets $\{2^{-\rho_k}\}_{k=1}^{K}$ guarantees the feasibility of the QoS given as average rate targets since the average MMSE restrictions are more stringent than the original average rate ones. The proposed test is fundamental to determine the convergence of the power minimization algorithms presented in Chapters 4 and 5 when imperfect CSIT is assumed, and is an extension to a more general scenario of the test presented in [40] where a MISO BC with perfect CSI at both ends of the communication link was considered.

5.2.1. Feasibility Region in the SIMO MAC

In this section we revise the feasibility region studied in [40]. Consider the *Single-Input Multiple-Output* (SIMO) MAC presented in Fig. 2.3 of Chapter 2. In such a setup the users equipped with one transmit antenna send K independent data signals to the N-antenna BS. The data signals are precoded with the linear filter $t_k \in \mathbb{C}$ and filtered by $\boldsymbol{g}_k \in \mathbb{C}^N$ at the BC. The SIMO channel and the noise are $\boldsymbol{h}_k \in \mathbb{C}^N$, and $\boldsymbol{\eta} \sim \mathbb{C}_{\mathcal{N}}(\mathbf{0}, \sigma^2\mathbf{I}_N)$, respectively, with σ^2 being the noise variance.

As shown in Section 4.1.2 of Chapter 4, employing the optimal MMSE receiver $\boldsymbol{g}_k^{\text{MMSE}}$ leads to the corresponding MMSE as follows

$$\text{MMSE}_k = 1 - |t_k|^2 \boldsymbol{h}_k^{\text{H}} \left(\sum_{i=1}^{K} \boldsymbol{h}_i |t_i|^2 \boldsymbol{h}_i^{\text{H}} + \sigma^2 \mathbf{I}_N \right)^{-1} \boldsymbol{h}_k. \tag{5.45}$$

We now define the matrices $\boldsymbol{T} \in \mathbb{C}^{K\times K}$, and $\boldsymbol{H} \in \mathbb{C}^{N\times K}$ as

$$\boldsymbol{T} = \text{diag}\left(t_1, \ldots, t_K\right), \tag{5.46}$$

$$\boldsymbol{H} = \left[\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K\right]. \tag{5.47}$$

Given that notation, the MMSE for the user k is expressed as

$$\text{MMSE}_k = 1 - |t_k|^2 \boldsymbol{h}_k^{\text{H}} \left(\boldsymbol{H}\boldsymbol{T}\boldsymbol{T}^{\text{H}}\boldsymbol{H}^{\text{H}} + \sigma^2\mathbf{I}_N \right)^{-1} \boldsymbol{h}_k \tag{5.48}$$

$$= \left[\mathbf{I}_K - \boldsymbol{T}^{\text{H}}\boldsymbol{H}^{\text{H}} \left(\boldsymbol{H}\boldsymbol{T}\boldsymbol{T}^{\text{H}}\boldsymbol{H}^{\text{H}} + \sigma^2\mathbf{I}_N \right)^{-1} \boldsymbol{H}\boldsymbol{T} \right]_{k,k}. \tag{5.49}$$

And the sum of the MMSEs for all the users can be easily obtained using the trace operator as follows

$$\sum_{k=1}^{K} \text{MMSE}_k = \text{tr}\left(\mathbf{I}_K - \boldsymbol{T}^{\text{H}}\boldsymbol{H}^{\text{H}} \left(\boldsymbol{H}\boldsymbol{T}\boldsymbol{T}^{\text{H}}\boldsymbol{H}^{\text{H}} + \sigma^2\mathbf{I}_N \right)^{-1} \boldsymbol{H}\boldsymbol{T} \right) \tag{5.50}$$

$$= K - \text{tr}\left(\boldsymbol{T}^{\text{H}}\boldsymbol{H}^{\text{H}} \left(\boldsymbol{H}\boldsymbol{T}\boldsymbol{T}^{\text{H}}\boldsymbol{H}^{\text{H}} + \sigma^2\mathbf{I}_N \right)^{-1} \boldsymbol{H}\boldsymbol{T} \right). \tag{5.51}$$

Now, by applying the trace operator properties (see Section C.3 of Appendix C), we get

$$\sum_{k=1}^{K} \mathrm{MMSE}_k = K - \mathrm{tr}\left(\boldsymbol{H}\boldsymbol{T}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\left(\boldsymbol{H}\boldsymbol{T}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}} + \sigma^2\mathbf{I}_N\right)^{-1}\right) \tag{5.52}$$

$$= K - N + \mathrm{tr}\left(\mathbf{I}_N - \boldsymbol{H}\boldsymbol{T}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\left(\boldsymbol{H}\boldsymbol{T}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}} + \sigma^2\mathbf{I}_N\right)^{-1}\right) \tag{5.53}$$

$$= K - N + \mathrm{tr}\left(\mathbf{I}_N + \sigma^{-2}\boldsymbol{H}\boldsymbol{T}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\right)^{-1}. \tag{5.54}$$

Since the matrix inversion in (5.54) is positive definite (see Section C.5 of Appendix C), its trace is lower bounded by $N - \mathrm{rank}(\boldsymbol{H})$, and then

$$\sum_{k=1}^{K} \mathrm{MMSE}_k \geq K - \mathrm{rank}\left(\boldsymbol{H}\right), \tag{5.55}$$

where the equality is asymptotically achieved when the total transmit power tends to infinity. It is reasonable to assume that the vector channels are not linearly dependent and then $\mathrm{rank}(\boldsymbol{H}) = \min\{N, K\}$.

Note from (5.55) that, if the number of antennas at the BS is large enough, i.e. $N \geq K$, the MMSEs can be made arbitrarily small simultaneously for all the users when the total transmit power tends to infinity. Observe also that, due to the duality presented in Section 3.3.1 of Chapter 3, the results shown in this section can be applied to the sum MMSE in the BC.

5.2.2. Feasibility Region for the MIMO MAC under Imperfect CSIR

We now extend the previous study to the more general case where MIMO channels and imperfect MAC CSIR are considered. In such scenario, the interferences cannot be completely removed even when the number of transmit antennas is larger than the number of users, i.e. $N \geq K$. Consequently, increasing the transmit power does not necessarily lead to a reduction of the MMSEs for all the users because, although it increases its received power, it also increases the power of the intra-user interference. In certain scenarios, the QoS constraints may require that some users achieve low MMSE values that may be unfeasible even though the transmit power is increased unlimitedly. That is in accordance with the following fact: when imperfect MAC CSIR is considered, the MMSE for all the users cannot be reduced arbitrarily in a simultaneous way, contrarily to the case where there exist perfect CSI on both the transmitters and the receiver in the MAC.

In the following we present a feasibility test to determine whether it is possible or not to accomplish the QoS constraints $\overline{\mathrm{MMSE}}_k^{\mathrm{MAC}} = 2^{-\rho_k}$.

Let us start substituting $\boldsymbol{g}_k^{\text{MMSE}}$, given by (5.34), into the $\overline{\text{MSE}}_k^{\text{MAC}}$ expression given by (5.22) to determine the average MMSE in the MAC

$$\overline{\text{MMSE}}_k^{\text{MAC}} = 1 - \text{E}\left[\boldsymbol{t}_k^{\text{H}} \boldsymbol{C}_{\boldsymbol{\eta}_k}^{-1/2} \boldsymbol{H}_k^{\text{H}} |\, v\right] \left(\sigma^2 \boldsymbol{I}_N + \sum_{i=1}^{K} \text{E}[\boldsymbol{H}_i \boldsymbol{C}_{\boldsymbol{\eta}_i}^{-\text{H}/2} \boldsymbol{t}_i \boldsymbol{t}_i^{\text{H}} \boldsymbol{C}_{\boldsymbol{\eta}_i}^{-1/2} \boldsymbol{H}_i^{\text{H}} |\, v]\right)^{-1} \text{E}\left[\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\text{H}/2} \boldsymbol{t}_k |\, v\right], \tag{5.56}$$

where we introduced the noise covariance in the dual MAC, i.e. $\boldsymbol{C}_{\boldsymbol{\eta}} = \sigma^2 \mathbf{I}_N$. We now introduce the matrix

$$\boldsymbol{\Upsilon} = [\boldsymbol{H}_1 \boldsymbol{C}_{\boldsymbol{\eta}_1}^{-\text{H}/2}, \ldots, \boldsymbol{H}_K \boldsymbol{C}_{\boldsymbol{\eta}_K}^{-\text{H}/2}] \, \text{blockdiag}(\boldsymbol{t}_1, \ldots, \boldsymbol{t}_K), \tag{5.57}$$

and rewrite (5.56) as follows

$$\overline{\text{MMSE}}_k^{\text{MAC}} = 1 - \left[\text{E}[\boldsymbol{\Upsilon}^{\text{H}}|v] \left(\text{E}[\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^{\text{H}}|v] + \sigma^2 \mathbf{I}_N\right)^{-1} \text{E}[\boldsymbol{\Upsilon}|v]\right]_{k,k}. \tag{5.58}$$

Hence, the sum average MMSE is rewritten using the trace operator as

$$\sum_{i=1}^{K} \overline{\text{MMSE}}_i^{\text{MAC}} = K - \text{tr}\left(\text{E}[\boldsymbol{\Upsilon}^{\text{H}}|v] \left(\text{E}[\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^{\text{H}}|v] + \sigma^2 \mathbf{I}_N\right)^{-1} \text{E}[\boldsymbol{\Upsilon}|v]\right). \tag{5.59}$$

When $K \geq N$ and the channel knowledge is perfect at both sides of the link, the expectations in (5.59) are removed and the sum average MMSE can be made arbitrarily small [40]. However, due to the imperfect CSI at the MAC receiver, we cannot reduce the average MMSE as much as desired for all the users at the same time.

Expression (5.59) allows to determine the region where the feasible average MMSEs lie. Indeed, setting the MAC thermal noise variance to zero (i.e., $\sigma^2 = 0$) we obtain a lower bound for the sum average MMSE for any finite total average power allocation, i.e.

$$\sum_{i=1}^{K} \overline{\text{MMSE}}_i^{\text{MAC}} > K - \text{tr}\{\boldsymbol{X}\}, \tag{5.60}$$

where $\boldsymbol{X} = \text{E}[\boldsymbol{\Upsilon}^{\text{H}}|v](\text{E}[\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^{\text{H}}|v])^{-1} \text{E}[\boldsymbol{\Upsilon}|v]$. The bound is asymptotically achieved when the powers for all the users reach infinity. Therefore, we formulate a necessary condition for the feasibility of QoS targets: any power allocation with finite sum power achieves an MMSE tuple $\{\overline{\text{MMSE}}_i^{\text{MAC}}\}_{i=1}^{K}$ inside the polytope

$$\mathcal{P} = \left\{ \{\overline{\text{MMSE}}_i^{\text{MAC}}\}_{i=1}^{K} \,|\, \sum_{i=1}^{K} \overline{\text{MMSE}}_i^{\text{MAC}} \geq K - \text{tr}\,\{\boldsymbol{X}\} \right\}. \tag{5.61}$$

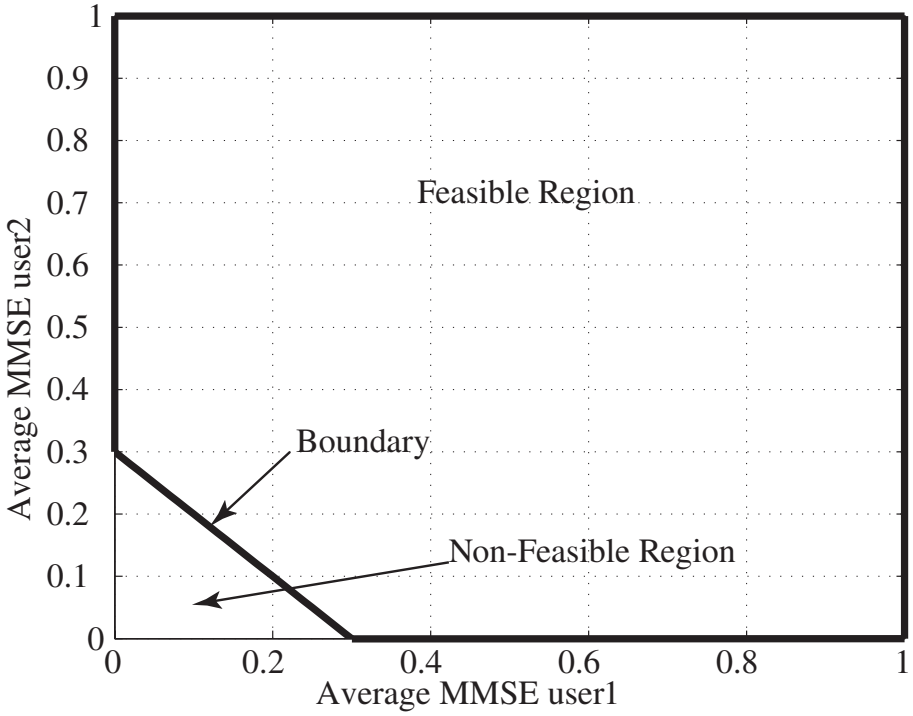


Figure 5.6: Example of Feasibility Region in a 2-User Scenario.

An example of the feasibility region considering the sum-MMSE lower bounded by 0.3 is shown in Fig. 5.6. The boundary is reached when the total transmit power asymptotically reaches infinite and separates the feasible and unfeasible regions.

So far we have found a necessary condition for the feasibility of the QoS targets, i.e., any power allocation with finite sum power achieves an MMSE tuple $\{\overline{\text{MMSE}}_i^{\text{MAC}}\}_{i=1}^K$ inside the polytope $\mathcal{P}$. In order to prove that $\mathcal{P}$ is the feasible set of solutions to the power minimization problem, we must prove the converse, i.e., that there exists a power allocation for any tuple inside the polytope $\mathcal{P}$. The mapping from $\{\overline{\text{MMSE}}_i^{\text{MAC}}\}_{i=1}^K$ to the power allocation results from equating $\overline{\text{MMSE}}_k^{\text{MAC}}$ with the target ε_k. The resulting fixed point is unique due to the properties of the interference functions [38]. Then, if the fixed point exists, the aforementioned mapping is bijective. Let $\boldsymbol{f}(\boldsymbol{x};\boldsymbol{c})$ be a multivariate function that depends on a vector of independent variables $\boldsymbol{x}$ and a vector of parameters $\boldsymbol{c}$. Such function has a fixed point $\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x};\boldsymbol{c})$ if it satisfies the following set of sufficient

conditions [72]

$$\boldsymbol{f}(\mathbf{0};\boldsymbol{c}) \geq \mathbf{0}, \tag{5.62}$$

$$\exists\, \boldsymbol{a} > \mathbf{0} \quad \text{such that} \quad \boldsymbol{f}(\boldsymbol{a};\boldsymbol{c}) > \boldsymbol{a}, \tag{5.63}$$

$$\exists\, \boldsymbol{b} > \boldsymbol{a} \quad \text{such that} \quad \boldsymbol{f}(\boldsymbol{b};\boldsymbol{c}) < \boldsymbol{b}. \tag{5.64}$$

We now define $\varepsilon_k = 2^{-\rho_k}$ as the average MMSE targets, that is, the lower bounds of the original average rate based on QoS constraints (see 5.17), and $\boldsymbol{\varepsilon} = [\varepsilon_1, \ldots, \varepsilon_K]^{\mathrm{T}}$ as the vector that collects all such targets. We also introduce the following definitions

$$\boldsymbol{\varphi}_k = \frac{1}{\sqrt{\xi_k}} \mathrm{E}\left[\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2} \boldsymbol{t}_k | \, v\right], \tag{5.65}$$

$$\begin{aligned}\boldsymbol{\Phi}_k = \frac{1}{\xi_k} \mathrm{E}\Big[&\left(\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2} \boldsymbol{t}_k - \mathrm{E}\left[\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2} \boldsymbol{t}_k | \, v\right]\right) \\ &\left(\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2} t_k - \mathrm{E}\left[\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2} \boldsymbol{t}_k | \, v\right]\right)^{\mathrm{H}} | \, v\Big], \text{ and}\end{aligned} \tag{5.66}$$

$$\boldsymbol{A}_k = \sum_{i=1}^{K} \xi_i \boldsymbol{\Phi}_i + \sum_{j \neq k} \xi_j \boldsymbol{\varphi}_j \boldsymbol{\varphi}_j^{\mathrm{H}} + \sigma^2 \mathbf{I}_N, \tag{5.67}$$

which allows us to rewrite the MMSE (5.56) as

$$\overline{\mathrm{MMSE}}_k^{\mathrm{MAC}} = 1 - \boldsymbol{\varphi}_k^{\mathrm{H}} \left(\boldsymbol{A}_k + \xi_k \boldsymbol{\varphi}_k \boldsymbol{\varphi}_k^{\mathrm{H}}\right)^{-1} \boldsymbol{\varphi}_k. \tag{5.68}$$

Applying the matrix inversion lemma to the last expression (see Appendix Section C.4 of C), enables us to eventually get

$$\overline{\mathrm{MMSE}}_k^{\mathrm{MAC}} = \left(1 + \xi_k \boldsymbol{\varphi}_k^{\mathrm{H}} \boldsymbol{A}_k^{-1} \boldsymbol{\varphi}_k\right)^{-1}. \tag{5.69}$$

We now define the function $f_k(\boldsymbol{\xi};\boldsymbol{\varepsilon})$ from (5.69) equating $\overline{\mathrm{MMSE}}_k^{\mathrm{MAC}} = \varepsilon_k$ at some fixed point $\boldsymbol{\xi}$

$$f_k(\boldsymbol{\xi};\boldsymbol{\varepsilon}) := \left(\varepsilon_k^{-1} - 1\right) \left(\boldsymbol{\varphi}_k^{\mathrm{H}} \boldsymbol{A}_k^{-1} \boldsymbol{\varphi}_k\right)^{-1} \forall k. \tag{5.70}$$

We next show that the fixed points $\xi_k = f_k(\boldsymbol{\xi};\boldsymbol{\varepsilon})$ correspond to the optimal power allocation vectors $\boldsymbol{\xi}^{\mathrm{opt}}$ for which $\overline{\mathrm{MMSE}}_k^{\mathrm{MAC}} = \varepsilon_k$, $\forall k$. To do so, we show in the following that the function $\boldsymbol{f}(\boldsymbol{\xi};\boldsymbol{\varepsilon}) = [f_1(\boldsymbol{\xi};\boldsymbol{\varepsilon}), \ldots, f_K(\boldsymbol{\xi};\boldsymbol{\varepsilon})]^{\mathrm{T}}$ satisfies the fixed point conditions (5.62), (5.63), and (5.64).

The first requirement given in (5.62) is easy to show since if the transmit power is $\boldsymbol{\xi} = \mathbf{0}$, the inter-user interference drops out and the matrix $\boldsymbol{A}_k$ reduces to $\sigma^2 \mathbf{I}_N$. Then,

$$f_k(\mathbf{0};\boldsymbol{\varepsilon}) = \frac{1 - \varepsilon_k}{\varepsilon_k} \frac{\sigma^2}{\|\boldsymbol{\varphi}_k\|_2^2}. \tag{5.71}$$

Note that $f_k(\mathbf{0};\boldsymbol{\varepsilon}) \geq 0$ as long as $0 < \varepsilon_k \leq 1$. Moreover, (5.71) also provides a lower bound for $f_k(\boldsymbol{\xi};\varepsilon)$, i.e.

$$f_k(\boldsymbol{\xi};\boldsymbol{\varepsilon}) \geq \frac{1-\varepsilon_k}{\varepsilon_k} \frac{\sigma^2}{\|\boldsymbol{\varphi}_k\|_2^2}, \tag{5.72}$$

for any $\boldsymbol{\xi} \geq \mathbf{0}$.

The second condition given in (5.63) is also easy to prove. Indeed, let $\bar{a}$ be the minimum element of $\boldsymbol{f}(\mathbf{0};\boldsymbol{\varepsilon})$. Hence, $\boldsymbol{f}(\boldsymbol{\xi};\boldsymbol{\varepsilon}) \geq \bar{a}\mathbf{1}$, for any $\boldsymbol{\xi} \geq \mathbf{0}$. Note from (5.72) that $\bar{a} > 0$ as long as $\varepsilon_k < 1$. Observe now that the power allocation $\boldsymbol{\xi} = a\mathbf{1}$, with $a < \bar{a}$, gives

$$\boldsymbol{f}(a\mathbf{1};\boldsymbol{\varepsilon}) \geq \bar{a}\mathbf{1} > a\mathbf{1}, \tag{5.73}$$

thus satisfying (5.63).

The proof for the condition (5.64) is more involved so we will divide the problem into two cases depending on the number of users and transmit antennas.

A) $\mathbf{N \geq K}$. This is the case where the number of transmit antennas is greater than or equal to the number of users. We start searching for an upper bound for $f_k(\boldsymbol{\xi};\boldsymbol{\varepsilon})$, or equivalently, a lower bound for the inverse term in (5.70). To do so, we introduce the following matrices

$$\boldsymbol{B}_{\bar{k}} = [\boldsymbol{\varphi}_{i_1}, \ldots, \boldsymbol{\varphi}_{i_{K-1}}]_{i_j \neq k, \forall j}, \tag{5.74}$$

$$\boldsymbol{\Xi}_{\bar{k}} = \mathrm{diag}(\xi_i)_{i \neq k}, \tag{5.75}$$

which allow us to rewrite the second summand in $\boldsymbol{A}_k$ from (5.67) as

$$\sum_{i \neq k} \xi_i \boldsymbol{\varphi}_i \boldsymbol{\varphi}_i^{\mathrm{H}} = \boldsymbol{B}_{\bar{k}} \boldsymbol{\Xi}_{\bar{k}} \boldsymbol{B}_{\bar{k}}^{\mathrm{H}}. \tag{5.76}$$

Additionally, we define the matrix $\boldsymbol{\Phi} \in \mathbb{C}^{N \times N}$ as follows

$$\boldsymbol{\Phi} = \sum_{i=1}^{K} \xi_i \boldsymbol{\Phi}_i + \sigma^2 \mathbf{I}_N, \tag{5.77}$$

which allows us to rewrite the matrix $\boldsymbol{A}_k$ as

$$\boldsymbol{A}_k = \boldsymbol{\Phi} + \boldsymbol{B}_{\bar{k}} \boldsymbol{\Xi}_{\bar{k}} \boldsymbol{B}_{\bar{k}}^{\mathrm{H}}. \tag{5.78}$$

Applying now the matrix inversion lemma of Appendix C it is possible to obtain the inverse of $\boldsymbol{A}_k$ in the way

$$\boldsymbol{A}_k^{-1} = \boldsymbol{\Phi}^{-1} \left[\mathbf{I}_N - \boldsymbol{B}_{\bar{k}} \left(\boldsymbol{\Xi}_{\bar{k}}^{-1} + \boldsymbol{B}_{\bar{k}}^{\mathrm{H}} \boldsymbol{\Phi}^{-1} \boldsymbol{B}_{\bar{k}} \right)^{-1} \boldsymbol{B}_{\bar{k}}^{\mathrm{H}} \boldsymbol{\Phi}^{-1} \right]. \tag{5.79}$$

Thus, defining $\boldsymbol{\psi}_k = \boldsymbol{\Phi}^{-1/2}\boldsymbol{\varphi}_k \in \mathbb{C}^N$ and $\boldsymbol{D}_{\bar{k}} = \boldsymbol{\Phi}^{-1/2}\boldsymbol{B}_{\bar{k}} \in \mathbb{C}^{N\times K-1}$, and letting the total power grow without restriction leads us, eventually, to the lower bound

$$\boldsymbol{\varphi}_k^{\mathrm{H}}\boldsymbol{A}_k^{-1}\boldsymbol{\varphi}_k \geq \boldsymbol{\psi}_k^{\mathrm{H}}\left(\mathbf{I}_N - \boldsymbol{D}_{\bar{k}}\left(\boldsymbol{D}_{\bar{k}}^{\mathrm{H}}\boldsymbol{D}_{\bar{k}}\right)^{-1}\boldsymbol{D}_{\bar{k}}^{\mathrm{H}}\right)\boldsymbol{\psi}_k, \tag{5.80}$$

and the corresponding upper bound

$$f_k(\boldsymbol{\xi};\boldsymbol{\varepsilon}) \leq \frac{1-\varepsilon_k}{\varepsilon_k}\left(\boldsymbol{\psi}_k^{\mathrm{H}}\left(\mathbf{I}_N - \boldsymbol{D}_{\bar{k}}\left(\boldsymbol{D}_{\bar{k}}^{\mathrm{H}}\boldsymbol{D}_{\bar{k}}\right)^{-1}\boldsymbol{D}_{\bar{k}}^{\mathrm{H}}\right)\boldsymbol{\psi}_k\right)^{-1}. \tag{5.81}$$

Notice that the matrix $\boldsymbol{D}_{\bar{k}}^{\mathrm{H}}\boldsymbol{D}_{\bar{k}}$ in (5.80) and (5.81) is non-singular when $N \geq K$. Observe that the equality in the last expression holds for $\xi_k \to \infty, \forall k$. Since $\boldsymbol{f}(\boldsymbol{\xi};\boldsymbol{\varepsilon}) \geq \bar{a}\mathbf{1} > a\mathbf{1}$, for any $\boldsymbol{\xi} \geq \mathbf{0}$, sets a lower bound as stated in (5.73) to fulfill the second requirement given by (5.63), we only have to find $\boldsymbol{b}$ such that

$$b_k > \left(\frac{1}{\varepsilon_k} - 1\right)\left(\boldsymbol{\psi}_k^{\mathrm{H}}(\mathbf{I} - \boldsymbol{D}_{\bar{k}}\left(\boldsymbol{D}_{\bar{k}}^{\mathrm{H}}\boldsymbol{D}_{\bar{k}}\right)^{-1}\boldsymbol{D}_{\bar{k}}^{\mathrm{H}})\boldsymbol{\psi}_k\right)^{-1}, \tag{5.82}$$

to complete the proof for the third requirement of (5.64) when $N \geq K$, i.e. $b_k > f_k(a\mathbf{1},\boldsymbol{\varepsilon}) > a$, as required.

B) $\mathbf{N} < \mathbf{K}$. We now focus on the case when the number of transmit antennas is smaller than the number of users. The power allocation is set to $\boldsymbol{b} = \alpha\boldsymbol{b}_0$, where $\boldsymbol{b}_0$ belongs to the simplex

$$\mathcal{S} = \{\boldsymbol{x} | \sum_k x_k = 1 \text{ and } x_k \geq 0\ \forall k\}. \tag{5.83}$$

For $\alpha \to \infty$ (or $\sigma^2 \to 0$) and $\boldsymbol{b}_0 > \mathbf{0}$, we can rewrite (5.70) as

$$f_k^{\infty}(\boldsymbol{b}_0;\boldsymbol{\varepsilon}) := \frac{\frac{1}{\varepsilon_k - 1}}{\boldsymbol{\varphi}_k^{\mathrm{H}}\Big(\sum_i b_{0,i}\boldsymbol{\Phi}_i + \sum_{j\neq k} b_{0,j}\boldsymbol{\varphi}_j\boldsymbol{\varphi}_j^{\mathrm{H}}\Big)^{-1}\boldsymbol{\varphi}_k}.$$

The average MMSE targets collected in $\boldsymbol{\varepsilon}$ have to satisfy the equality (5.60) for $\alpha \to \infty$, i.e., a tuple $\boldsymbol{\varepsilon}$ that lies in the region separating feasible from unfeasible targets, i.e. $\mathcal{B} = \{\boldsymbol{\varepsilon} | \mathbf{1}^{\mathrm{T}}\boldsymbol{\varepsilon} = K - \mathrm{tr}(\boldsymbol{X})\}$. Note that $\boldsymbol{b}_0 = \boldsymbol{f}^{\infty}(\boldsymbol{b}_0;\boldsymbol{\varepsilon})$ is a fixed point of $\boldsymbol{f}^{\infty}$ but we need to verify the bijective mapping in order to complete the proof, that is, for any average MMSE target tuple $\boldsymbol{\varepsilon} \in \mathcal{B}$ there exists a unique power allocation $\boldsymbol{b} = \alpha\boldsymbol{b}_0$, with $\alpha \to \infty$.

First, we define the *Signal to Interference-plus-Noise Ratio* (SINR) as $\mathrm{SINR} = 1/\overline{\mathrm{MMSE}}_k^{\mathrm{MAC}} - 1$. In the limit case when the power allocation tends to infinity, $\alpha \to \infty$, we obtain the following expression for the *Signal to Interference Ratio* (SIR) as

$$\mathrm{SIR}_k = b_{0,k}\boldsymbol{\varphi}_k^{\mathrm{H}}\Big(\sum_{i=1}^{K} b_{0,i}\boldsymbol{\Phi}_i + \sum_{j\neq k} b_{0,j}\boldsymbol{\varphi}_j\boldsymbol{\varphi}_j^{\mathrm{H}}\Big)^{-1}\boldsymbol{\varphi}_k,$$

from which we rewrite $\text{SIR}_k = b_{0,k}(\mathcal{Q}_k(\boldsymbol{b}_0))^{-1}$. Thus, we can use the properties of the function $\mathcal{Q}_k(\boldsymbol{b}_0)$, (see [39]), to guarantee the existence and uniqueness of the optimal power allocation for the following balancing problem

$$\max_{r,\boldsymbol{b}_0} r \quad \text{subject to} \quad \frac{b_{0,k}}{\mathcal{Q}_k\left(\boldsymbol{b}_0\right)} = r\,\text{SIR}_{0,k} \quad \forall k \in \{1,\ldots,K\}. \tag{5.84}$$

Since we established a relationship between the SIR and the $\overline{\text{MMSE}}^{\text{MAC}}$ when we let the power grow without restriction, i.e. $\alpha \to \infty$, we use the bound for $\mathbf{1}\boldsymbol{\varepsilon}$ to find the optimal balancing level r of (5.84) using

$$\sum_{i=1}^{K} \frac{1}{1 + r\,\text{SIR}_{0,i}} = K - \text{tr}\{\boldsymbol{X}\}. \tag{5.85}$$

The previous equation only has a single solution since the functions $(1 + r\,\text{SIR}_{0,k})^{-1}$ are monotonically decreasing with $r > 0$, $\forall k$. For example, if we obtain the SIR targets from the MMSE targets lying in the region of interest $\mathcal{B}$, (5.85) is fulfilled with $r = 1$.

So far we have shown that a unique power allocation $\boldsymbol{b} = \alpha\boldsymbol{b}_0$, with $\boldsymbol{b}_0 \in \mathcal{S}$ and $\alpha \to \infty$, always exists for any MMSE tuple in the region separating feasible from unfeasible targets $\boldsymbol{\varepsilon}' \in \mathcal{B}$, such that $\boldsymbol{f}(\boldsymbol{b};\boldsymbol{\varepsilon}') = \boldsymbol{b}$. Note that $\boldsymbol{f}(\boldsymbol{b};\boldsymbol{\varepsilon})$ is decreasing in $\boldsymbol{\varepsilon}$ and we can prove that the third requirement of (5.64) is also fulfilled for $N < K$ due to the fact that for any target $\boldsymbol{\varepsilon} > \boldsymbol{\varepsilon}'$ we have

$$\boldsymbol{f}(\boldsymbol{b};\boldsymbol{\varepsilon}) < \boldsymbol{b}. \tag{5.86}$$

In summary, the power minimization problem of (4.44) has a solution, i.e., the MMSE QoS targets $\boldsymbol{\varepsilon} = [2^{-\rho_1},\ldots,2^{-\rho_K}]^{\text{T}}$ are feasible, if and only if $\boldsymbol{\varepsilon} \in \mathcal{P}$, with $\mathcal{P}$ defined in (5.61).

Recall that our result is a generalization of Theorem III.1 in [40], where the feasibility was studied for the case of perfect CSIT. In particular, the conditioned expectations in $\boldsymbol{X}$ from (5.60) are removed for error-free CSI, i.e., $\boldsymbol{X} = \boldsymbol{\Upsilon}^{\text{H}}(\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^{\text{H}})^{-1}\boldsymbol{\Upsilon} = \mathbf{I}_N$. Then, the bound on the sum average MMSE reads as $\sum_{i=1}^{K} \overline{\text{MMSE}}_i^{\text{MAC}} > K - \text{tr}\{\boldsymbol{X}\} = K - N$. From practical results, our observation is that $\text{tr}\{\boldsymbol{X}\} \approx R\ \forall k$, considering $\boldsymbol{h}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{h}_k})$, with $\boldsymbol{C}_{\boldsymbol{h}_k} = \text{E}[\boldsymbol{h}_k\boldsymbol{h}_k^{\text{H}}|v] = \sigma_{h_k}^2\mathbf{I}_{RN}$, where R is the number of per-user antennas and $\boldsymbol{h}_k = \text{vec}(\boldsymbol{H}_k)$. Taking that into account, we can infer that the feasible region for single-antenna receivers equals that of SISO systems for this particular imperfect CSI. Note also that, contrary to the perfect CSI case, $\text{tr}\{\boldsymbol{X}\}$ needs not be an integer number.

5.2.3. Simulation Results

In this section, we show the results of some numerical simulations carried out to validate the results obtained previously. We consider a MIMO BC with $N = 4$ antennas at the BS that sends data signals to $K = 4$ users equipped with $R = 2$ antennas. The imperfect CSI is translated into $M = 1000$ channel realizations $\boldsymbol{H}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, R\mathbf{I}_N)$, $\forall k$. Therefore, $\boldsymbol{h}_k = \mathrm{vec}(\boldsymbol{H}_k) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{h}_k})$. Recall our observation from the previous section, i.e., $\sum_{i=1}^{K} \overline{\mathrm{MMSE}}_i^{\mathrm{MAC}} > K - R$. Then, $\sum_{i=1}^{K} \overline{\mathrm{MMSE}}_i^{\mathrm{MAC}} > 2$ for this setup. Bearing that in mind, we run the Algorithm 5.1: PM.MIMO from Section 5.1.4 with the unfeasible average rate targets $\rho_1 = 1.3219$, $\rho_2 = 0.7370$, $\rho_3 = 1.737$, and $\rho_4 = 1$ bits per channel use, and the corresponding average MMSE lower bounds $\varepsilon_1 = 0.4$, $\varepsilon_2 = 0.6$, $\varepsilon_3 = 0.3$, and $\varepsilon_4 = 0.5$, such that $\sum_{k=1}^{4} \varepsilon_k = 1.8 < 2$. The number of iterations is limited to 50 since the algorithm shall not converge for unfeasible targets.

Figure 5.7 depicts the gradual reduction of the average MMSEs for all the users. However, the desired values cannot be simultaneously reached for all of them. On the other hand, Fig. 5.8 shows the behavior of the total transmit power during the algorithm execution. Observe that the power increases over 36 dB when trying to get the average MMSE targets. However, as we previously discussed, the QoS constraints are not feasible.

5.3. Sum-MSE Lower Bound

This section focuses on the bound already presented in the previous section. In particular, we will focus on expression (5.60) to compute the lower bound of the summation of the average MMSEs for all the users. Note, however, that this expression depends on the MAC precoders (or the dual BC receive filters). Such an equation also provides critical information about the feasibility of the approximated average rate targets. Therefore, we will develop a method to find the filters that minimize (5.60) in the following in order to obtain a better understanding of the bound (5.60). Let us consider the MSE expression in the MAC when we let the transmit power to increase unlimitedly. In other words, we neglect the noise and obtain

$$\mathrm{MSE}_k^{\mathrm{MAC}} = 1 - 2\Re\left\{\boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{H}_k \boldsymbol{t}_k\right\} + \sum_{i=1}^{K} \boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{H}_i \boldsymbol{t}_i \boldsymbol{t}_i^{\mathrm{H}} \boldsymbol{H}_i^{\mathrm{H}} \boldsymbol{g}_k. \tag{5.87}$$

The sum-MSE in such a case is computed summing up the MSE for all the users to get

$$\sum_{k=1}^{K} \mathrm{MSE}_k^{\mathrm{MAC}} = K - 2\Re\left\{\sum_{k=1}^{K} \boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{H}_k \boldsymbol{t}_k\right\} + \sum_{k=1}^{K}\sum_{i=1}^{K} \boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{H}_i \boldsymbol{t}_i \boldsymbol{t}_i^{\mathrm{H}} \boldsymbol{H}_i^{\mathrm{H}} \boldsymbol{g}_k. \tag{5.88}$$

We now find the optimal MAC precoders, that is, the filters $\boldsymbol{t}_k$ minimizing the sum of the MSEs for all the users. Observe that, when we try to find the MMSE filters, the MAC

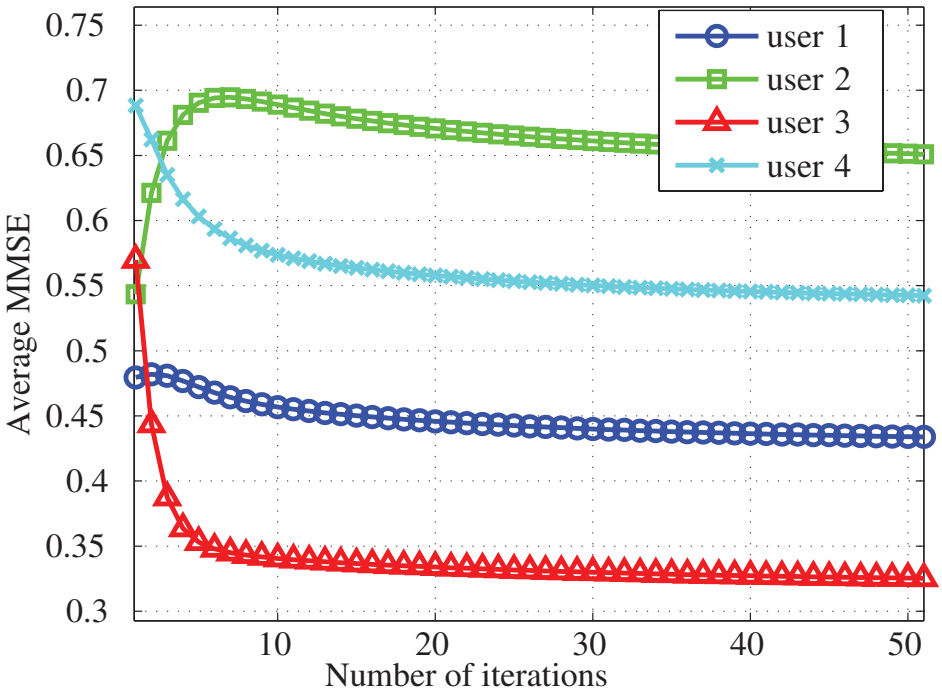


Figure 5.7: Example of Power Minimization Algorithm Execution for Unfeasible Targets: MMSE vs. Number of Iterations.

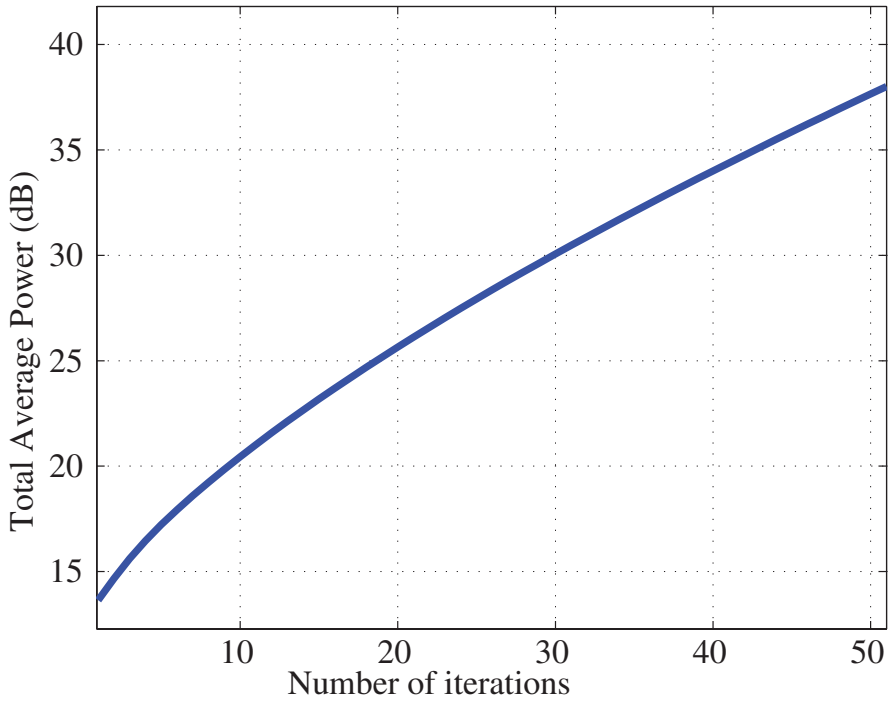


Figure 5.8: Example of Power Minimization Algorithm Execution for Unfeasible Targets: Total Power vs. Number of Iterations.

transmit filters are computed in the BC due to the dependence of the MSE of the user k with the transmit filter $\boldsymbol{t}_j$, for $j \neq k$. Nevertheless, for the sum-MSE such a problem does not hold. Hence, in order to find the MAC transmit filters that minimize (5.88), we calculate the derivative of $\sum_{k=1}^{K} \mathrm{tr}\left(\mathrm{MSE}_k^{\mathrm{MAC}}\right) = \sum_{k=1}^{K} \mathrm{MSE}_k^{\mathrm{MAC}}$ with respect to $\boldsymbol{t}_k$ using the results presented in Section C.6 of Appendix C, i.e.

$$\frac{\partial \sum_{k=1}^{K} \mathrm{tr}\left(\mathrm{MSE}_k^{\mathrm{MAC}}\right)}{\partial \boldsymbol{t}_k^*} = -\boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{g}_k + \boldsymbol{H}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{g}_i \boldsymbol{g}_i^{\mathrm{H}} \boldsymbol{H}_k \boldsymbol{t}_k. \tag{5.89}$$

Hence, the MAC transmit filters read as

$$\boldsymbol{t}_k = \left(\boldsymbol{H}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{g}_i \boldsymbol{g}_i^{\mathrm{H}} \boldsymbol{H}_k \right)^{-1} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{g}_k, \tag{5.90}$$

and substituting into (5.88) we get

$$\sum_{k=1}^{K} \mathrm{MSE}_k^{\mathrm{MAC}} = K - \sum_{k=1}^{K} \boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{H}_k \boldsymbol{t}_k = K - \sum_{k=1}^{K} \boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{H}_k \left(\boldsymbol{H}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{g}_i \boldsymbol{g}_i^{\mathrm{H}} \boldsymbol{H}_k \right)^{-1} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{g}_k. \tag{5.91}$$

Remember now that along this work we have considered perfect CSIT and imperfect CSIR in the MAC. Thus, (5.90) holds and the average of (5.91) conditioned to the imperfect CSI information available at the BS, v, has to be considered. Additionally, we define the matrix that collects the receive filters for all the users as $\boldsymbol{G} = [\boldsymbol{g}_1, \ldots, \boldsymbol{g}_K]$, such that $\sum_{i=1}^{K} \boldsymbol{g}_i \boldsymbol{g}_i^{\mathrm{H}} = \boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}$, and $\boldsymbol{g}_k = \boldsymbol{G}\boldsymbol{e}_k$, with $\boldsymbol{e}_k$ being the canonical vector. The average sum-MSE is then

$$\mathrm{E}\left[\sum_{k=1}^{K} \mathrm{MSE}_k^{\mathrm{MAC}} |\, v \right] = K - \sum_{k=1}^{K} \boldsymbol{e}_k^{\mathrm{T}} \mathrm{E}\left[\boldsymbol{G}^{\mathrm{H}} \boldsymbol{H}_k \left(\boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{G}\boldsymbol{G}^{\mathrm{H}} \boldsymbol{H}_k \right)^{-1} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{G} |\, v \right] \boldsymbol{e}_k. \tag{5.92}$$

To find the receive filters that minimize the sum-MSE, $\boldsymbol{G}^{\mathrm{opt}}$, the trace is again introduced into the average sum-MSE, that is

$$\begin{aligned} \mathrm{E}\left[\sum_{k=1}^{K} \mathrm{MSE}_k^{\mathrm{MAC}} |\, v \right] &= K - \sum_{k=1}^{K} \mathrm{tr}\left(\boldsymbol{e}_k^{\mathrm{T}} \mathrm{E}\left[\boldsymbol{G}^{\mathrm{H}} \boldsymbol{H}_k \left(\boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{G}\boldsymbol{G}^{\mathrm{H}} \boldsymbol{H}_k \right)^{-1} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{G} |\, v \right] \boldsymbol{e}_k \right) \\ &= K - \sum_{k=1}^{K} \mathrm{tr}\left(\mathrm{E}\left[\left(\boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{G}\boldsymbol{G}^{\mathrm{H}} \boldsymbol{H}_k \right)^{-1} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{G}\boldsymbol{e}_k \boldsymbol{e}_k^{\mathrm{T}} \boldsymbol{G}^{\mathrm{H}} \boldsymbol{H}_k |\, v \right] \right). \end{aligned} \tag{5.93}$$

Accordingly, in order to simplify the computation of the derivative of (5.92) with respect to $\boldsymbol{G}$ we first determine the following derivative

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{G}^*}\sum_{k=1}^{K}\mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}_k\left(\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}_k\right)^{-1}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}|\, v\right]\boldsymbol{e}_k\boldsymbol{e}_k^{\mathrm{T}}\right)\\ =\sum_{k=1}^{K}\mathrm{E}\left[\boldsymbol{H}_k\left(\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}_k\right)^{-1}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}|\, v\right]\boldsymbol{e}_k\boldsymbol{e}_k^{\mathrm{T}},\end{aligned} \tag{5.94}$$

where we have considered $(\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}_k)^{-1}$ as a constant. On the other hand, we have

$$\begin{aligned}&\frac{\partial}{\partial[\boldsymbol{G}^*]_{i,j}}\sum_{k=1}^{K}\mathrm{tr}\left(\mathrm{E}\left[\left(\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}_k\right)^{-1}\boldsymbol{A}_k|\, v\right]\right)\\ &\quad=-\sum_{k=1}^{K}\mathrm{tr}\left(\mathrm{E}\left[\left(\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}_k\right)^{-1}\boldsymbol{A}_k\left(\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}_k\right)^{-1}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}\boldsymbol{e}_j\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{H}_k|\, v\right]\right)\\ &\quad=-\sum_{k=1}^{K}\boldsymbol{e}_i^{\mathrm{T}}\,\mathrm{E}\left[\boldsymbol{H}_k\left(\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}_k\right)^{-1}\boldsymbol{A}_k\left(\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}_k\right)^{-1}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}|\, v\right]\boldsymbol{e}_j,\end{aligned} \tag{5.95}$$

with $\boldsymbol{A}_k=\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}\boldsymbol{e}_k\boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}_k$. Finally, adding up (5.94) with the matrix resulting from the components of (5.95), we obtain the following gradient for the user m

$$\begin{aligned}\frac{\partial\,\mathrm{E}\left[\sum_{k=1}^{K}\mathrm{MSE}_k^{\mathrm{MAC}}|\, v\right]}{\partial\boldsymbol{G}^*}\boldsymbol{e}_m&=-\,\mathrm{E}\left[\boldsymbol{H}_m\left(\boldsymbol{H}_m^{\mathrm{H}}\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}_m\right)^{-1}\boldsymbol{H}_m^{\mathrm{H}}\boldsymbol{G}|\, v\right]\boldsymbol{e}_m\\ &+\sum_{k=1}^{K}\mathrm{E}\left[\boldsymbol{H}_k\left(\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}_k\right)^{-1}\boldsymbol{A}_k\left(\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}_k\right)^{-1}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}|\, v\right]\boldsymbol{e}_m.\end{aligned} \tag{5.96}$$

Since finding a close-form solution to $\boldsymbol{G}^{\mathrm{opt}}$ is a difficult task, we propose to employ a steepest descent algorithm. Moreover, due to the fact that minimizing the sum-MSE (5.92) with respect to $\boldsymbol{G}$ is an unconstrained convex problem (see Section F.1 of Appendix F), the algorithm converges to the optimum filters $\boldsymbol{G}^{\mathrm{opt}}$.

5.3.1. Algorithmic Solution

The algorithm referred to as Algorithm 5.2: Opt.Rx finds the MAC receive filters $\boldsymbol{g}_k$ that minimize the average sum-MSE lower bound presented in (5.93). All the precoders are randomly initialized in the line 1. Inside the outer loop, the current version of the matrices employed to compute the gradient of (5.96) are calculated (line 4). Next, the

Algorithm 5.2: Opt.Rx Optimum Receive Filter for Sum-MSE Lower Bound

1: $\ell \leftarrow 0$, randomly initialize $\boldsymbol{G}^{(0)}$
2: **repeat**
3: $\quad \ell \leftarrow \ell + 1$
4: $\quad \boldsymbol{\Phi}^{(\ell)}, \boldsymbol{\Theta}^{(\ell)} \leftarrow$ compute the matrices from (5.96)
5: $\quad s \leftarrow 1$, set the step size to starting value
6: $\quad$ **repeat**
7: $\quad\quad \boldsymbol{g}_k^{(\ell)} \leftarrow \boldsymbol{g}_k^{(\ell-1)} - s(-\boldsymbol{\Phi}^{(\ell)}\boldsymbol{e}_k + \boldsymbol{\Theta}^{(\ell)}\boldsymbol{e}_k)$, $\forall k$ (see (5.96))
8: $\quad\quad s \leftarrow \frac{s}{2}$
9: $\quad$ **until** $\sum_{i=1}^{K} \mathrm{E}[\mathrm{MSE}_k(\boldsymbol{G}^{\ell})|\, v] - \sum_{i=1}^{K} \mathrm{E}[\mathrm{MSE}_k(\boldsymbol{G}^{\ell-1})|\, v] < 0$
10: **until** $|\sum_{i=1}^{K} \mathrm{E}[\mathrm{MSE}_k(\boldsymbol{G}^{\ell})|\, v] - \sum_{i=1}^{K} \mathrm{E}[\mathrm{MSE}_k(\boldsymbol{G}^{\ell-1})|\, v]| < \delta$

step size is initialized to 1 (line 5). The line 7 inside the inner loop computes the new MAC receive filters for all the users using the gradient and the step size s. The step size is reduced in the subsequent iterations of the inner loop when the condition in the line 9 is not fulfilled. In other words, if the average sum-MSE lower bound is larger than the one from the previous iteration, the step size is reduced (line 8). The line 10 decides if the convergence has been reached or not, depending on the performance measure reduction obtained in two consecutive iterations.

5.3.2. Simulation Results

We now consider the MIMO MAC from Fig. 5.2 to evaluate the impact of the imperfect CSI quality on the average sum-MSE lower bound presented in Section 5.3. To that end, we use a simple model for channel estimation, that is, a training sequence is sent over the channel and is also affected by the thermal noise. At the BS, the channel is estimated depending on the received signal.

Let $\boldsymbol{h}_k \in \mathbb{C}^{NR}$ be the channel vector for the user k such that $\boldsymbol{h}_k = \mathrm{vec}(\boldsymbol{H}_k)$, with N and R being the number of antennas at the BS and the users, respectively. In the following, we drop the subindex k for the shake of notation brevity. We assume that the channel vector and the thermal noise $\boldsymbol{\eta} \in \mathbb{C}^{NR}$ are zero-mean circularly symmetric complex Gaussian and independent, i.e. $\boldsymbol{h} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{h}})$ and $\boldsymbol{\eta} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{\eta}})$, respectively. We now introduce the training sequence $\boldsymbol{S} \in \mathbb{C}^{NR \times NR}$ such that $\boldsymbol{S} = \sqrt{p}\boldsymbol{\Sigma}$, with $||\boldsymbol{\Sigma}||_{\mathrm{F}}^2 = NR$, and $\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S} = p\mathbf{I}_{NR}$. Thus, the received signal at the BS is

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{h} + \boldsymbol{\eta}. \tag{5.97}$$

We stack the channel and the noise to obtain a new circularly symmetric complex Gaussian random vector $[\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{h}^{\mathrm{T}}]^{\mathrm{T}}$. Taking into account that the property holds for any

linear combination [73], this leads us to

$$\begin{pmatrix} \mathbf{I}_{NR} & \boldsymbol{S} \\ \mathbf{0} & \mathbf{I}_{NR} \end{pmatrix} \begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{h} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{h} \end{bmatrix} \sim \mathcal{N}_{\mathbb{C}} \left(\mathbf{0}, \begin{pmatrix} \boldsymbol{C}_{\boldsymbol{y}} & \boldsymbol{C}_{\boldsymbol{y},\boldsymbol{h}} \\ \boldsymbol{C}_{\boldsymbol{h},\boldsymbol{y}} & \boldsymbol{C}_{\boldsymbol{h}} \end{pmatrix} \right), \tag{5.98}$$

where the covariance matrices read as

$$\boldsymbol{C}_{\boldsymbol{y}} = \mathrm{E}\left[\boldsymbol{y}\boldsymbol{y}^{\mathrm{H}}\right] = \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}}, \tag{5.99}$$

$$\boldsymbol{C}_{\boldsymbol{y},\boldsymbol{h}} = \mathrm{E}\left[\boldsymbol{y}\boldsymbol{h}^{\mathrm{H}}\right] = \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}} = \boldsymbol{C}_{\boldsymbol{h},\boldsymbol{y}}^{\mathrm{H}}, \tag{5.100}$$

and $\boldsymbol{y} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}})$.

To obtain the pdf of the channel $\boldsymbol{h}$ given the observation $\boldsymbol{y}$, $f_{\boldsymbol{h}|\boldsymbol{y}}(\boldsymbol{h}|\boldsymbol{y})$, we employ the Bayes' rule $f_{\boldsymbol{h}|\boldsymbol{y}}(\boldsymbol{h}|\boldsymbol{y}) = f_{\boldsymbol{h},\boldsymbol{y}}(\boldsymbol{h},\boldsymbol{y})/f_{\boldsymbol{y}}(\boldsymbol{y})$ to get (see Section E.1 of Appendix E)

$$f_{\boldsymbol{h}|\boldsymbol{y}}(\boldsymbol{h}|\boldsymbol{y}) = \frac{1}{\pi^{NR} \det\left(\boldsymbol{C}_{\boldsymbol{h}} - \boldsymbol{C}_{\boldsymbol{h},\boldsymbol{y}}\boldsymbol{C}_{\boldsymbol{y}}^{-1}\boldsymbol{C}_{\boldsymbol{y},\boldsymbol{h}}\right)} \, \mathrm{e}^{-Q(\boldsymbol{h},\boldsymbol{y})}, \tag{5.101}$$

where

$$Q\left(\boldsymbol{h},\boldsymbol{y}\right) = \left(\boldsymbol{h} - \boldsymbol{C}_{\boldsymbol{h},\boldsymbol{y}}\boldsymbol{C}_{\boldsymbol{y}}^{-1}\boldsymbol{y}\right)^{\mathrm{H}} \left(\boldsymbol{C}_{\boldsymbol{h}} - \boldsymbol{C}_{\boldsymbol{h},\boldsymbol{y}}\boldsymbol{C}_{\boldsymbol{y}}^{-1}\boldsymbol{C}_{\boldsymbol{y},\boldsymbol{h}}\right)^{-1} \left(\boldsymbol{h} - \boldsymbol{C}_{\boldsymbol{h},\boldsymbol{y}}\boldsymbol{C}_{\boldsymbol{y}}^{-1}\boldsymbol{y}\right). \tag{5.102}$$

The first and second order moments are then

$$\mathrm{E}\left[\boldsymbol{h}|\boldsymbol{y}\right] = \boldsymbol{C}_{\boldsymbol{h},\boldsymbol{y}}\boldsymbol{C}_{\boldsymbol{y}}^{-1}\boldsymbol{y} = \boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} \left(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}}\right)^{-1} \boldsymbol{y}, \tag{5.103}$$

$$\boldsymbol{C}_{\boldsymbol{h}|\boldsymbol{y}} = \boldsymbol{C}_{\boldsymbol{h}} - \boldsymbol{C}_{\boldsymbol{h},\boldsymbol{y}}\boldsymbol{C}_{\boldsymbol{y}}^{-1}\boldsymbol{C}_{\boldsymbol{y},\boldsymbol{h}} = \boldsymbol{C}_{\boldsymbol{h}} - \boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} \left(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}}\right)^{-1} \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}. \tag{5.104}$$

We consider a MIMO scenario with $N = 4$ antennas at the BC, and $K = 4$ users equipped with $R = 2$ antennas each. The CSI quality in this setup comes from the *Signal-to-Noise Ratio* (SNR) of the received signal. In our case, we define the SNR as the ratio between the transmit and the noise powers, i.e.

$$\text{SNR} = \frac{\|\boldsymbol{S}\|_{\mathrm{F}}^2}{\mathrm{tr}\left(\boldsymbol{C}_{\boldsymbol{\eta}}\right)} = \frac{p}{\sigma^2}, \tag{5.105}$$

where we consider $\boldsymbol{C}_{\boldsymbol{\eta}} = \sigma^2\mathbf{I}_{NR}$, with $\sigma^2 = 1$ in the numerical experiments. Therefore, the transmit power p is modified so we can get SNRs from 0 to 40 dB. With $\boldsymbol{h}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_{NR})$ for all the users, we generate a realization of $\boldsymbol{y}_k$, $\forall k$. Then, we generate $M = 2000$ realizations of the random variable $\boldsymbol{h}|\boldsymbol{y}$, which is described by its mean of (5.103) and covariance matrix of (5.104), as we have previously shown. The generated random variables are then introduced as the channel realizations for Algorithm 5.2: Opt.Rx.

On the one hand, Fig. 5.9 shows the average sum-MSE lower bound reached for different SNRs, that is, the best performance that can be theoretically reached when we

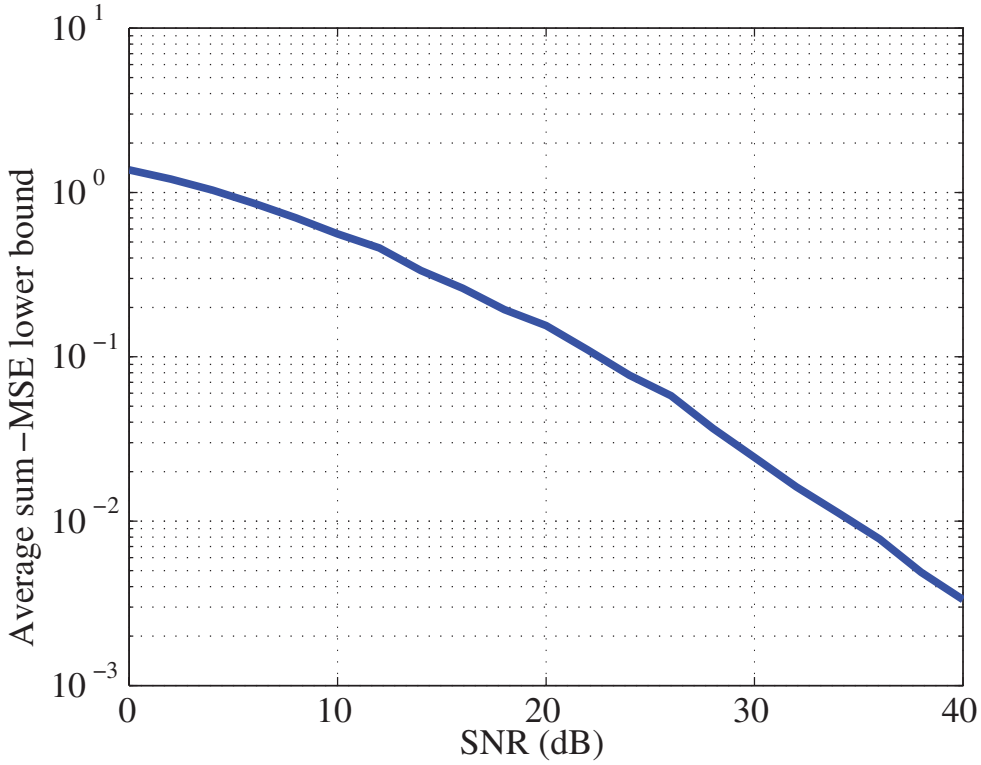


Figure 5.9: Average sum-MSE Lower Bound vs. SNR.

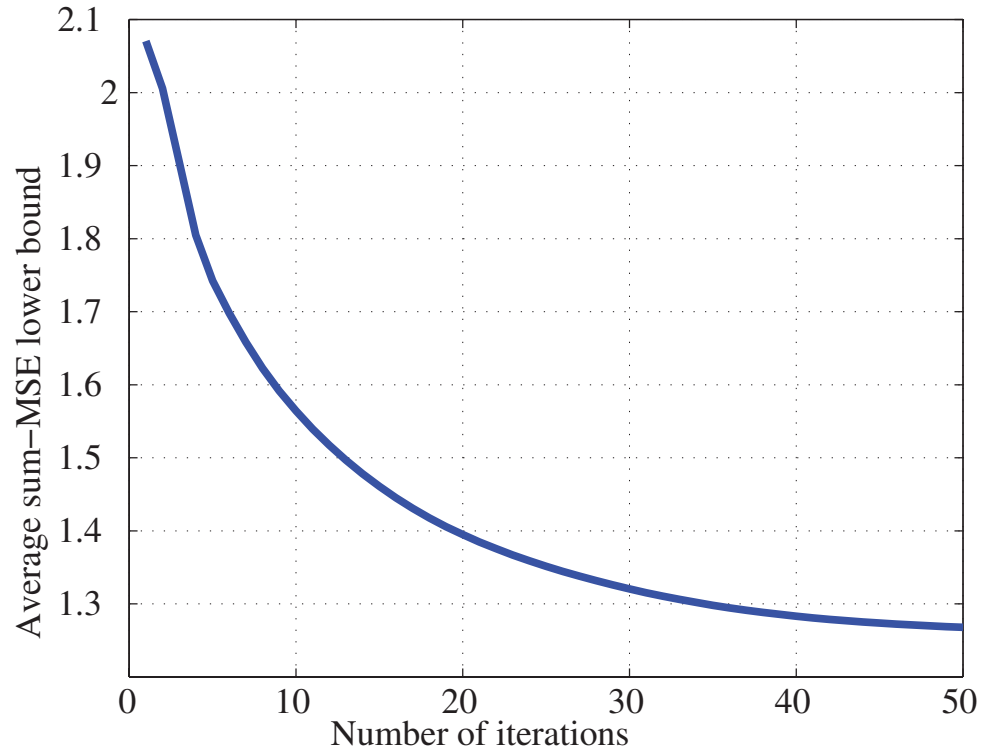


Figure 5.10: Convergence of Algorithm 5.3: Average sum-MSE Lower Bound vs. Number of Iterations.

increase the power without limit. Note that the impact of the CSI quality is critical and the lower bound drastically reduces for large SNR values. On the other hand, Fig. 5.10 depicts the evolution of the average sum-MSE lower bound with the filters computed with Algorithm 5.2: Opt.Rx. Here, the convergence of the steepest descent algorithm can be seen for an execution example.

5.4. Rate Balancing Problem

So far we have considered the design of the precoders and the receivers in a MIMO BC to minimize the transmit power fulfilling certain QoS constraints. However, when the QoS constraints are rather stringent, the minimum transmit power may be unacceptably high for the system or, as explained in the previous section, the problem may be unfeasible. To avoid such harmful situations, we face the problem of designing the MIMO BC using a different perspective. On the one hand, we relax the constraint that all the users experience a set of average rates given by $\{\rho_k\}_{k=1}^K$ and, on the other, we limit the transmit power to a given value P_T. More specifically, we propose that the per-user average rate targets be scaled by the same scale factor $\varsigma \in \mathbb{R}^+$ and that such factor be optimized, together with the precoders and the receivers, for given transmit power P_T. When the optimum ς^{opt} is obtained, the user k enjoys an average rate $\rho_k^{\text{opt}} = \varsigma^{\text{opt}}\rho_k$ and, correspondingly, a certain QoS. The scaling of all the user rates by the same factor in a multiuser communication system is referred to as rate balancing in the literature.

Other examples of balancing problems can be found for the SINR [12], or MMSE [34] for perfect CSIR and CSIT. Moreover, the imperfect CSIT in the BC is studied in [31] for average MMSE.

Recall the optimization problem proposed in (5.5). We now describe the converse problem where the transmit power equal toP_T is one of the constraints to be fulfilled. The aim is to get the best QoS constraints as possible taking into account a given relationship between the per-user average rates, that is

$$\max_{\{\varsigma,\boldsymbol{p}_k\}_{k=1}^K} \varsigma \quad \text{subject to} \quad \mathrm{E}[R_k|v] \geq \varsigma\rho_k,\ \forall k, \quad \text{and} \quad \sum_{i=1}^{K} \|\boldsymbol{p}_i\|_2^2 \leq P_T. \tag{5.106}$$

Notice from the previous problem formulation that the design of the precoders $\boldsymbol{p}_k$ has to be jointly optimized with the balancing factor ς. One remarkable difference with respect to the power minimization problem of (5.5) is that the optimization problem in (5.106) is always feasible. In other words, we can relax the problem restrictions as much as needed, via reducing ς until the QoS constraints are achieved using a total transmit power smaller or equal to P_T.

We now follow the same procedure as in Section 5.1.1, where the constraints were lower bounded (see (5.17)) using the average MMSE to transform the problem into a more tractable one. Then, we obtain the new restrictions

$$\overline{\mathrm{MMSE}}_k^{\mathrm{BC}} \leq 2^{-\varsigma\rho_k}. \tag{5.107}$$

Taking that into account, the MIMO BC balancing problem proposed to be solved in this section is expressed as follows

$$\max_{\{\varsigma,\boldsymbol{p}_k,\boldsymbol{f}_k\}_{k=1}^K} \varsigma \quad \text{subject to} \quad \overline{\mathrm{MMSE}}_k^{\mathrm{BC}} \leq 2^{-\varsigma\rho_k},\ \forall k, \quad \text{and} \quad \sum_{i=1}^K \|\boldsymbol{p}_i\|_2^2 \leq P_T. \tag{5.108}$$

Note now that this maximization depends on the receive filters $\boldsymbol{f}_k$ for the average MMSE lower bound. Thus, following an argumentation similar to that presented in Section 5.1.2, the optimization problem (5.108) can be equivalently rewritten in the dual MAC to allow for an individual adaptation of the BC precoders. Analogously to the development carried out for the power minimization problem in Section 5.1.3, we propose to use the *standard* interference functions to obtain a new reformulation of the problem for given MAC transmit filters $\boldsymbol{t}_k,\ \forall k$, that reads as

$$\max_{\{\varsigma,\xi_k,\boldsymbol{g}_k\}_{k=1}^K} \varsigma \quad \text{subject to} \quad \frac{J_k\left(\boldsymbol{\xi}\right)}{\xi_k} \leq 2^{-\varsigma\rho_k}, \text{ and } \sum_{i=1}^K \xi_i \leq P_T, \tag{5.109}$$

where $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_K]^T$ is the power allocation vector, $\boldsymbol{g}_k$ are the dual MAC receivers, and $J_k\left(\boldsymbol{\xi}\right)$ are the interference functions as given by (5.43).

Note that the Algorithm 5.1: PM.MIMO presented in Section 5.1.4 can be used to determine $\boldsymbol{\xi}^{\mathrm{opt}}$ and $\boldsymbol{g}_k^{\mathrm{opt}}$ for given value ς. This algorithm by itself is not valid in this section because $\boldsymbol{\xi}^{\mathrm{opt}}$ may not satisfy the power constraint, i.e., $\sum_{i=1}^K \xi_i^{\mathrm{opt}}$ may be larger than P_T. However, it can be combined with a bisection method to solve (5.109).

The bisection method is commonly used to find roots of a continuous function $f(x)$ in a certain interval $[a, b]$, such that $\mathrm{sgn}(f(a)) = -\mathrm{sgn}(f(b))$. The process is very simple and consists on selecting subintervals where the root lies until certain accuracy is reached. Although there exist quicker methods, the bisection search is robust and suitable for our off-line optimization problem. Figure 5.11 shows three iterations of the bisection search for an hypothetical relationship between the balance level and the total power. Note that the search can be performed since the function is monotonically increasing with the total transmit power $\boldsymbol{\xi}$, as we will discuss in the following. We will also show how the search interval is reduced at every iteration.

Indeed, let us start setting two feasible rate balancing values ς^{L} and ς^{H}, such that the optimum, ς^{opt}, lies in between, i.e. $\varsigma^{\mathrm{L}} \leq \varsigma^{\mathrm{opt}} \leq \varsigma^{\mathrm{H}}$. Let $\boldsymbol{\xi}^{\mathrm{L}}$ and $\boldsymbol{\xi}^{\mathrm{H}}$ be the optimum

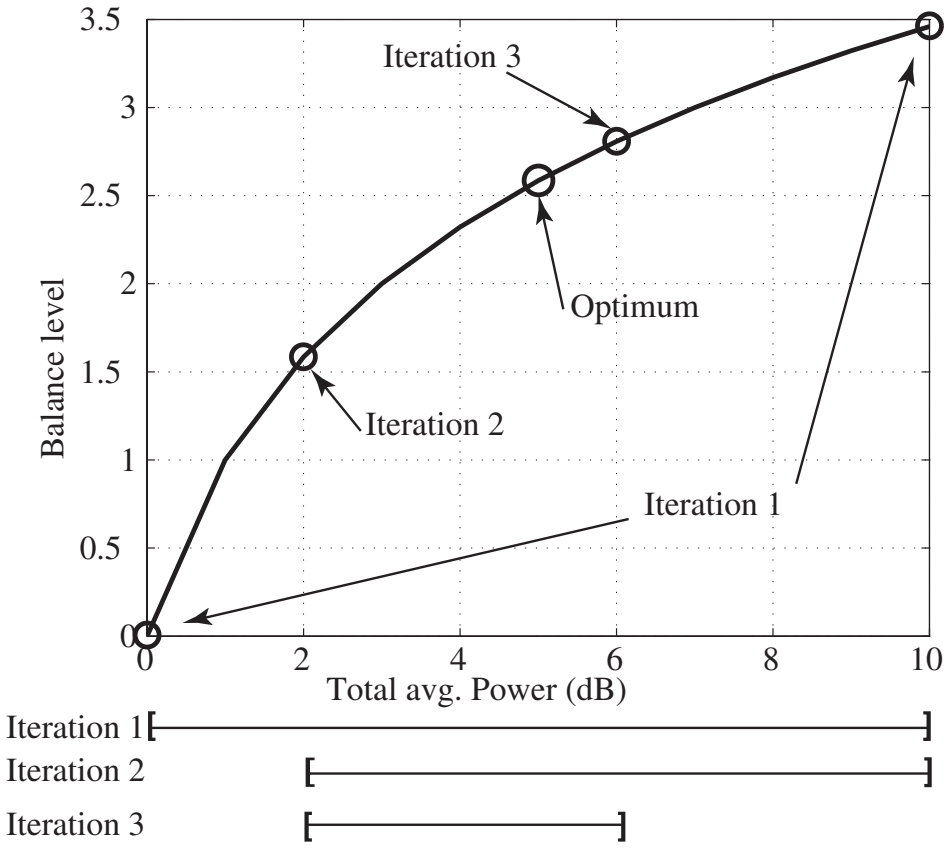


Figure 5.11: Example of Bisection Search.

power allocation vectors corresponding to ς^{L} and ς^{H}, respectively. Such optimal power allocation vectors satisfy on the one hand

$$\frac{I_k(\boldsymbol{\xi}^{\mathrm{L}})}{\xi_k^{\mathrm{L}}} = 2^{-\varsigma^{\mathrm{L}}\rho_k} \quad \text{and} \quad \frac{I_k(\boldsymbol{\xi}^{\mathrm{H}})}{\xi_k^{\mathrm{H}}} = 2^{-\varsigma^{\mathrm{H}}\rho_k}, \tag{5.110}$$

and on the other one

$$\sum_{i=1}^{K} \xi_i^{\mathrm{L}} \leq \sum_{i=1}^{K} \xi_i^{\mathrm{opt}} \leq \sum_{i=1}^{K} \xi_i^{\mathrm{H}}, \tag{5.111}$$

as we will show in the following.

We now introduce the average MMSE balancing factors

$$\epsilon_k = \frac{2^{-\varsigma\rho_k}}{2^{-\rho_k}} = 2^{-\rho_k(\varsigma-1)}. \tag{5.112}$$

Note that increasing the balance level ς produces smaller scaling factors ϵ_k, $\forall k$. Analogously, reducing the balance level translates into larger MMSE scaling factors. Let ϵ_k^{L} and ϵ_k^{H} be the MMSE scaling factors corresponding to ς^{L} and ς^{H}, respectively. Accordingly, $\epsilon_k^{\mathrm{L}} \geq \epsilon_k^{\mathrm{opt}} \geq \epsilon_k^{\mathrm{H}}$.

With the goal of proving that a bisection search can be performed, we consider $\epsilon_k^{\mathrm{L}} = a\epsilon_k^{\mathrm{opt}}$, with $a > 1$. The constraints in (5.109) are fulfilled with equality when $\epsilon_k = \epsilon_k^{\mathrm{opt}}$ and $\boldsymbol{\xi} = \boldsymbol{\xi}^{\mathrm{opt}}$. Hence

$$a\epsilon_k^{\mathrm{opt}} 2^{-\rho_k} = a\frac{J_k(\boldsymbol{\xi}^{\mathrm{opt}})}{\xi_k^{\mathrm{opt}}}, \tag{5.113}$$

which means that increasing the MMSE targets leads to smaller transmit power (i.e. $\xi_k = a^{-1}\xi_k^{\mathrm{opt}}$, $\forall k$) when we keep the interference as constant. Moreover, notice that keeping this fact sets an upper bound for the interference with the reduced transmit powers, $J_k(a^{-1}\boldsymbol{\xi}^{\mathrm{opt}}) < J_k(\boldsymbol{\xi}^{\mathrm{opt}})$, due to the scalability property of the *standard* interference functions. Therefore, the power needed to fulfill the constraint with equality is lower than $a^{-1}\boldsymbol{\xi}^{\mathrm{opt}}$, and $\mathbf{1}^{\mathrm{T}}\boldsymbol{\xi}^{\mathrm{L}} < a^{-1}\mathbf{1}^{\mathrm{T}}\boldsymbol{\xi}^{\mathrm{opt}} < P_T$ holds.

We now prove the relationship in the reverse direction, that is, a power reduction translates into larger scaling factors ϵ_k, $\forall k$. Let us consider the power reduction $\boldsymbol{A}\boldsymbol{\xi}^{\mathrm{opt}}$, with $\boldsymbol{A} = \mathrm{diag}(a_1, \cdots, a_K) < \mathbf{I}_K$, that leads to certain average MMSE scaling factor $\tilde{\epsilon}_k$ for some user k, i.e.

$$\tilde{\epsilon}_k 2^{-\rho_k} = \frac{1}{a_k \xi_k^{\mathrm{opt}}} I_k(\boldsymbol{A}\boldsymbol{\xi}^{\mathrm{opt}}). \tag{5.114}$$

Since no assumption about the user k has been made, we focus on the user with the largest power reduction, that is k', such that $a_{k'} \leq a_k$, $\forall k$. Consequently,

$$\tilde{\epsilon}_{k'} 2^{-\rho_{k'}} = \frac{I_{k'}\left(\boldsymbol{A}\boldsymbol{\xi}^{\mathrm{opt}}\right)}{a_{k'}\xi_{k'}^{\mathrm{opt}}} \geq \frac{I_{k'}\left(a_{k'}\boldsymbol{\xi}^{\mathrm{opt}}\right)}{a_{k'}\xi_{k'}^{\mathrm{opt}}} > \frac{I_{k'}\left(\boldsymbol{\xi}^{\mathrm{opt}}\right)}{\xi_{k'}^{\mathrm{opt}}} = \epsilon_{k'}^{\mathrm{opt}} 2^{-\rho_{k'}}, \tag{5.115}$$

where again we take advantage of the scalability property. Therefore, $\tilde{\epsilon}_{k'} > \epsilon_{k'}^{\text{opt}}$ for $\boldsymbol{\xi}^{\text{opt}} > \boldsymbol{A}\boldsymbol{\xi}^{\text{opt}}$. Remember from (5.112) that a larger $\tilde{\epsilon}_{k'}$ comes from a smaller ς. Then, the inequality also holds for the rest of the users and we get

$$\tilde{\epsilon}_k > \epsilon_k^{\text{opt}}, \forall k \quad \text{for} \quad \boldsymbol{\xi}^{\text{opt}} > \boldsymbol{A}\boldsymbol{\xi}^{\text{opt}}. \tag{5.116}$$

We have previously shown that relaxing the balancing level ϵ_k^{opt} implies a power reduction with respect to $\boldsymbol{\xi}^{\text{opt}}$. Hence, we conclude that a power reduction entails a lower balancing level ς, and vice-versa, when the precoders, the receive filters, and the power allocation vectors are optimum for every balancing level.

Finally, we conclude that it is possible to reduce the gap between ς^{L} and ς^{H} via bisection search explained before. That way, the optimum balancing level ς^{opt} for the total average transmit power $\sum_{k=1}^{K} \xi_k^{\text{opt}} = P_T$ can be found.

5.4.1. Rate Balancing Algorithm

Algorithm 5.3: Rate.Balancing. Rate Balancing in the MIMO BC

1: $\ell \leftarrow 0$, initialize $\varsigma^{\text{L},(0)}, \varsigma^{\text{H},(0)}$
2: find $\boldsymbol{\xi}^{\text{H},(0)} \leq \boldsymbol{\xi}^{\text{L},(0)}$ via power minimization Algorithm 5.1: PM.MIMO
3: **repeat**
4: $\quad \ell \leftarrow \ell + 1$
5: $\quad \varsigma^{(\ell)} \leftarrow \sqrt{\varsigma^{\text{L},(\ell-1)}\varsigma^{\text{H},(\ell-1)}}$ new balancing candidate
6: $\quad$ find $\boldsymbol{\xi}^{(\ell)}$ for $\varsigma^{(\ell)}$ via power minimization Algorithm 5.1: PM.MIMO
7: $\quad$ **if** $\sum_{i=1}^{K} \xi_i^{(\ell)} < P_T$ **then**
8: $\quad\quad \varsigma^{\text{H},(\ell)} \leftarrow \varsigma^{(\ell)}$, $\varsigma^{\text{L},(\ell)} \leftarrow \varsigma^{\text{L},(\ell-1)}$ update balance candidates
9: $\quad$ **else**
10: $\quad\quad \varsigma^{\text{L},(\ell)} \leftarrow \varsigma^{(\ell)}$, $\varsigma^{\text{H},(\ell)} \leftarrow \varsigma^{\text{H},(\ell-1)}$ update balance candidates
11: $\quad$ **end if**
12: **until** $|\sum_{i=1}^{K} \xi_i^{(\ell)} - P_T| < \delta$

Algorithm 5.3: Rate.Balancing presents the steps to solve the optimization problem (5.109). The algorithm is initialized with two balancing levels $\varsigma^{\text{L},(0)}$ and $\varsigma^{\text{H},(0)}$ (line 1). Next, their corresponding vector power allocation vectors, $\boldsymbol{\xi}^{\text{H},(0)}$ and $\boldsymbol{\xi}^{\text{L},(0)}$, are computed via the power minimization of Algorithm 5.1: PM.MIMO (line 2). Observe that the power constraint P_T lies between these two powers. Next, the algorithm enters a loop that first computes a new balancing level as the geometric mean of the balancing levels obtained in the previous iteration (line 5). Then, the power allocation vector for this new balancing level is computed using the Algorithm 5.1: PM.MIMO (line 6). Next, we check

whether the power obtained is lower than the power constraint or not (line 7) and update the balancing levels accordingly (lines 8 and 10). Finally, we test if the current power has the desired accuracy to finish the iteration loop (line 12).

The proof for the convergence of the Algorithm 5.3: Rate.Balancing depends on the feasibility of the power minimization problem with the average MMSE targets $2^{-\varsigma^{\mathrm{H},(0)}\rho_k}$, $\forall k$. Indeed, recall that the feasibility region is described in Section 5.2 as a bounded polytope and that the initial balancing levels $\varsigma^{\mathrm{L},(0)}$ and $\varsigma^{\mathrm{H},(0)}$ are chosen such as $\varsigma^{\mathrm{L},(0)} \leq \varsigma^{\mathrm{opt}} \leq \varsigma^{\mathrm{H},(0)}$. Hence, if $2^{-\varsigma^{\mathrm{H},(0)}\rho_k}$, $\forall k$, lies inside the polytope so does $2^{-a\varsigma^{\mathrm{H},(0)}\rho_k}$, $\forall k$, for any $0 \leq a < 1$. Taking into account that the average MMSE given by $\frac{1}{\xi_k^{(\ell)}} J_k(\boldsymbol{\xi}^{(\ell)})$ is monotonically decreasing with $\boldsymbol{\xi}^{(\ell)}$, the bisection procedure reduces the gaps $\varsigma^{\mathrm{H},(\ell)} - \varsigma^{\mathrm{L},(\ell)}$ and $|\mathbf{1}^{\mathrm{T}}\boldsymbol{\xi}^{(\ell)} - P_T|$ at every iteration until a desired accuracy δ is achieved.

5.4.2. Simulation Results

This section focuses on the performance of Algorithm 5.3: Rate.Balancing. This algorithm solves the optimization problem (5.108) by means of Algorithm 5.1: PM.MIMO and a bisection process for which it is necessary to decide two starting points, $\varsigma^{\mathrm{L},(0)}$ and $\varsigma^{\mathrm{H},(0)}$, such that the optimum balancing level lies in between, i.e., $\varsigma^{\mathrm{L},(0)} \leq \varsigma^{\mathrm{opt}} \leq \varsigma^{H,(0)}$.

To evaluate the proposed method, we consider the same setup as in Section 4.2.5 of Chapter 4, that is, the BS equipped with $N = 4$ transmit antennas sends information to $K = 4$ single-antenna users. A number of $M = 1000$ channel realizations generated by the model $\hat{\boldsymbol{h}}_k^{(m)} = \bar{\boldsymbol{h}}_k + \tilde{\boldsymbol{h}}_k^{(m)}$, for $k = \{1, \ldots, K\}$ and $m = \{1, \ldots, M\}$, with $\tilde{\boldsymbol{h}}_k^{(m)} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\tilde{\boldsymbol{h}}_k})$, are used. The thermal noise is considered to be equal for all the users, i.e. $\sigma_{\eta_k}^2 = 1$, $\forall k$. The rate targets are $\rho_1 = 0.5146, \rho_2 = 0.737, \rho_3 = 1$, and $\rho_4 = 0.2345$ bits per channel use, respectively, as considered in Section 4.2.5. We scale them with different balancing candidates to obtain the rate targets for such a candidate. The threshold to check if convergence has been reached or not is set to $\delta = 10^{-2}$.

Taking into account the numerical results obtained in Section 4.2.5, we set the available total average transmit power to $P_T = 3$ dB leading to an expected balancing level of approximately one, i.e. $\varsigma^{\mathrm{opt}} = 1$. Therefore, we pick $\varsigma^{\mathrm{L},(0)} = 0.6$ and $\varsigma^{\mathrm{H},(0)} = 1.3$, from which $\varsigma^{\mathrm{opt}} \in [0.6, 1.3]$. Figure 5.12 plots the average power versus the balancing level for the different iterations of the bisection algorithm. The two initial values correspond to the points located at left and right vertical axis in the figure. Note that the search interval reduces as the algorithm progresses until it converges after five iterations to the point $\varsigma^{\mathrm{opt}} = 0.99659$ and $P_T = 3.0072$ dB. This is in accordance with the experimental results obtained in Section 4.2.5 for the same power constraint leading to 1, the optimum balancing level, as it must be.

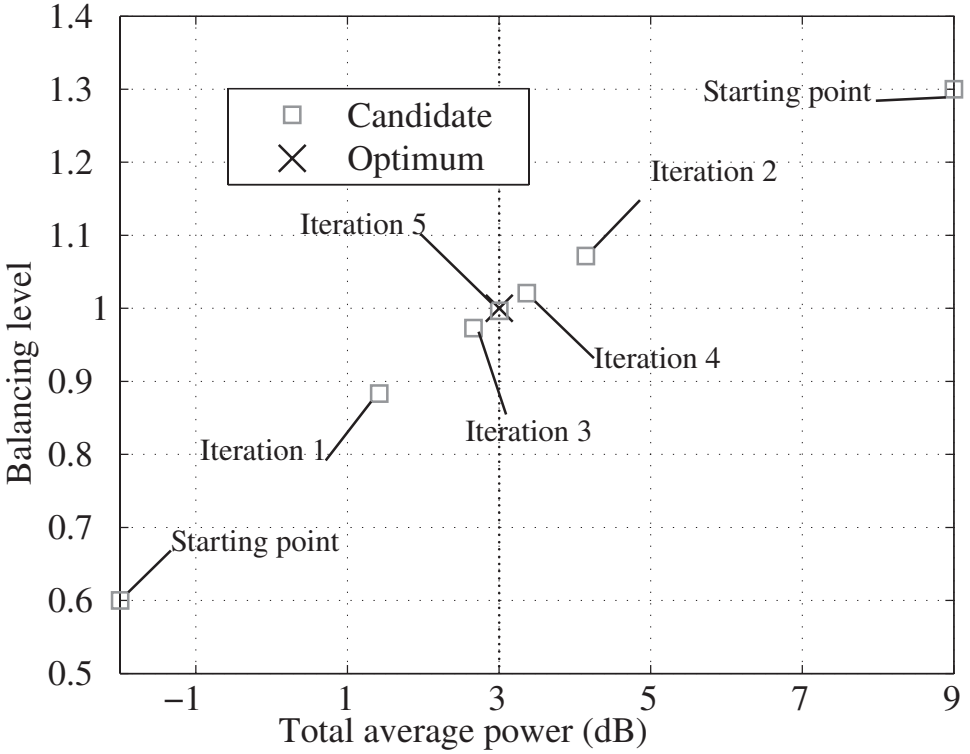


Figure 5.12: MMSE Balancing: Balance Level Candidates vs. Total Average Power in dB.

We also performed a computer experiment to compare our approach to that presented in [31], where a duality was proposed to solve several optimization problems considering a scenario where the users estimate their instantaneous channels and share the information with the BS through an error-free link. More specifically (see Section V of [31]) the following robust weighted MSE Min-Max problem was addressed

$$\min_{\{\boldsymbol{p}_k,\boldsymbol{f}_k\}_{k=1}^K} \max_i \frac{\overline{\text{MMSE}}_i^{\text{BC}}}{w_i} \quad \text{subject to} \quad \sum_{j=1}^K \|\boldsymbol{p}_j\|_2^2 \leq P_T, \tag{5.117}$$

where w_i is the weight for the ith user. The robust precoders and the filters are designed via an AO process where the direction of the filters is computed exploiting the BC/MAC duality, and the power allocation is calculated solving an eigensystem [34]. The optimum solution to (5.117) fulfills $\sum_{i=1}^K \|\boldsymbol{p}_i\|_2^2 = P_T$ and $\overline{\text{MMSE}}_k^{\text{BC}}/w_k = w^{\text{opt}}$, $\forall k$, and it is reached after a few iterations with an error precision for the min max ratio w^{opt} of 10^{-4}, as can be seen in Fig. 5.13.

This min max problem can be seen as a balancing problem with $w_k = 2^{-\rho_k}$. Thus, we

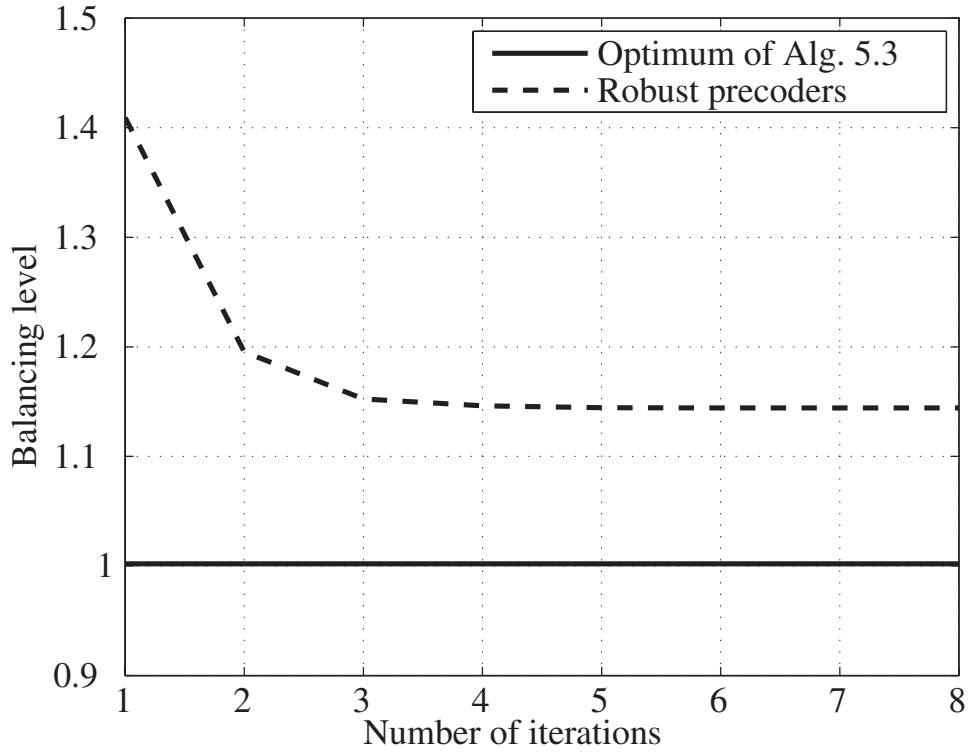


Figure 5.13: Robust Transceiver: Balance Level vs. Number of Iterations.

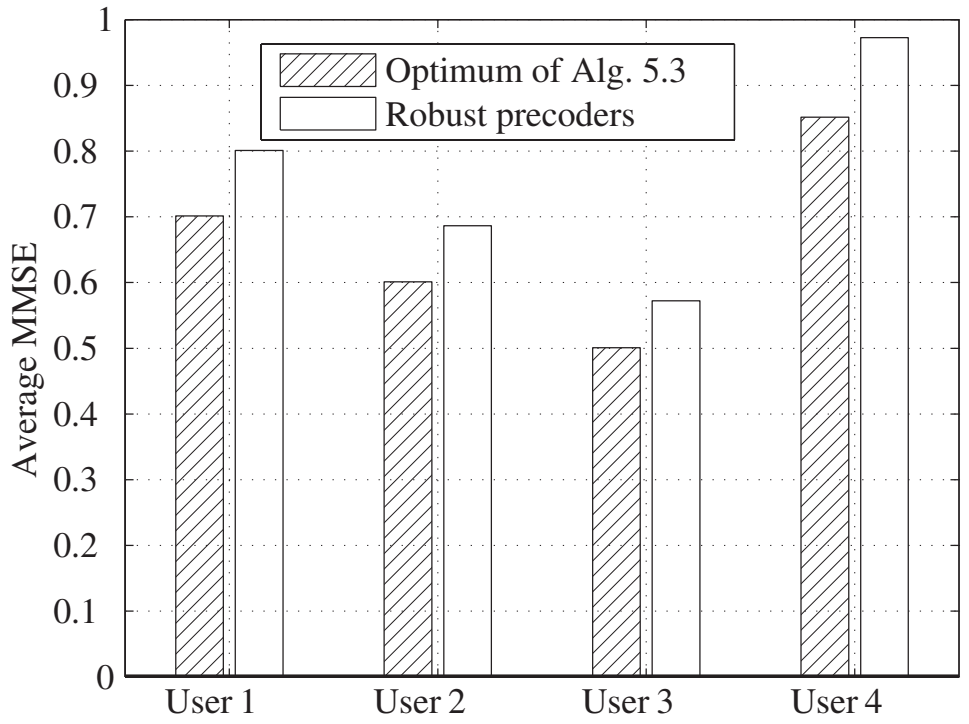


Figure 5.14: Robust Transceiver: Average MMSEs with Algorithm 5.3 vs. Average MMSEs for Robust Transceivers.

represent in Fig. 5.14 the comparison between the solutions employing robust transceivers and the one proposed in this work. As it can be seen from the figure, the proposed Algorithm 5.3: Rate.Balancing performs better, that is, $\varsigma^{\text{opt}} = 0.99659$ is closer than $w^{\text{opt}} = 1.1442$ to 1. However, the algorithm presented in [31] could manage the scenario with imperfect CSI at both the transmitter and the receivers. In addition, it is more efficient since the Algorithm 5.3: Rate.Balancing has to solve the power minimization problem for every balancing candidate.

5.5. Conclusions

In this chapter the minimization of the transmit power subject to QoS constraints under imperfect CSIT in the MIMO BC is addressed. The solution proposed is the extension to that developed in the previous chapter for the MISO BC. Simulation results show the performance of the adapted algorithm.

The optimization problem solved in this chapter is meaningless if the QoS constraints, given as average rate restrictions, are unfeasible. Since we employ a MMSE-based lower bound instead of the original average rate targets, a study of the average MMSE region for the MIMO MAC has been realized. Previous works had considered the MMSE feasibility region for perfect CSIT and MISO BC. Our study, however, is more general and allows to tackle the imperfect CSIT scenario. Moreover, for the particular case of Rayleigh channel realizations, we observed certain behavior that was shown in computer experiments also presented here.

Since the feasibility bound depends on the precoders employed for the partial CSIT scenario, we investigated a method to achieve the sum-MSE lower bound. Specifically, such method consists on employing a steepest descent algorithm to find the optimal MMSE transmit and receive filters, taking into account that the power can be increased unlimitedly. Using such algorithmic solution, we studied the impact of the CSI quality in the aforementioned sum-MSE lower bound. Numerical results were presented to exhibit the drawback caused by low precision CSIT.

A different problem formulation was also considered in this chapter: the average rate balancing. Contrary to the power minimization, the balancing problem was shown to be always feasible since the constraints can be relaxed while keeping some equilibrium between the user rates. Our proposal uses a bisection search, for which the rate targets increase and decrease monotonically with the total transmit power. This way, the optimal average rate targets can be found after few iterations. The proposed algorithm was compared to similar ones. Although our algorithmic solution is computationally costly, it outperforms the previous methods.

Chapter 6

Transmit Power Minimization in Broadcast Channels: Multiple Streams, OFDM, and Feedback Design

In this chapter we consider some additional issues to the power minimization in the *Broadcast Channel* (BC). Note that the *Base Station* (BS) has usually more degrees of freedom than the receivers. Therefore, it is appropriate trying to mitigate the interference between the users by applying precoding at transmission. In this chapter, the design of linear precoders is considered in several practical scenarios.

The first scenario considers the minimization of the transmit power in a BC where the BS is able to allocate several streams for each user. This is an extension to the model considered in Chapter 5, where the *Multiple-Input Multiple-Output* (MIMO) BC was considered with only one stream per user. Such a scenario is interesting if the objective of the MIMO feature is to increase reliability, since the probability of being affected by fading in all the independent paths at the same time is low. Compare this situation with that of multiplexing several streams for every user, hence taking advantage of the MIMO spatial multiplexing to increase the speed of the communication link. Considering multiple streams means an important change in the system model, since the dimension of both the transmit and receive filters have to be adapted accordingly. Therefore, this extension has a big impact on the problem formulation taking into account that we can choose different per-stream target rates without changing the per-user target rate. Thus, we end up with a nested optimization problem where we do not only have to find the optimal precoders but also the optimal per-stream rate constraints.

The second scenario considers *Orthogonal Frequency Division Multiplexing* (OFDM) modulation, a technique widely employed in current communication standards such as *Long-Term Evolution* (LTE) or digital television (DVB-T, DVB-H). OFDM is helpful in high speed communication systems where the frequency bandwidth is large, due to

its ability of transforming wideband frequency-selective fading channels into multiple flat fading narrowband channels. The combination of OFDM with MIMO allows high communication rates, e.g. as happened for the *Worldwide Interoperability for Microwave Access* (WiMAX) standard, and it implies a practical extension for the BC previously studied. It also adds more complexity to the system model due to the role of the frequency dimension. Therefore, we provide a problem formulation suitable for the MIMO-OFDM BC and propose an algorithm that minimizes the total transmit power subject to per-user average rate constraints.

The design of the linear transmit filters requires that the BS acquires the *Channel State Information* (CSI) of the different receivers. In case of *Frequency-Division Duplex* (FDD) systems that knowledge is obtained via a feedback link, which is usually band-limited. In this scenario, we consider every user estimates their channel response and then, the information is combined prior to be sent back to the BS as an entry of a codebook. Once the CSI is received by the BS, it is employed to decide the precoders that will be used, according to the information provided by all the users.

In this chapter we propose to perform an unique and joint optimization that includes the design of the transmit filters and the codebook. This joint optimization has been considered in several previous works. For example, for OFDM underwater acoustic channels [74], the Lloyd's algorithm [75] is employed to design the feedback. Additional examples can be found for the single-user *Multiple-Input Single-Output* (MISO) [76] and multiuser MISO [77–81] scenarios.

Different applications of the Lloyd's algorithm have also been studied. For example, in [82] the single antenna system model capacity is studied under certain assumptions, whereas the distortion outage probability in a sensor network with limited feedback is addressed in [83]. The MIMO BC scenario is considered in [84], where the zero-forcing precoders are designed together with the channel quantizer.

6.1. Power Minimization in the Multiple Stream MIMO BC

In this section, we focus on the application of the methods proposed in previous chapters when the BS allocates several streams for every user. The additional spatial dimension is possible due to the use of multiple antennas at both the BS and the users, i.e. MIMO BC. We consider that there is imperfect *Channel State Information at the Transmitter* (CSIT) whereas the users perfectly know their corresponding channels [85].

The multiple stream MIMO BC depicted in Fig. 6.1 is an extension of the system model shown in Fig. 5.1 of Chapter 5. Again, K users, with R antennas each, receive the information sent from a centralized transmitter with N antennas. The data

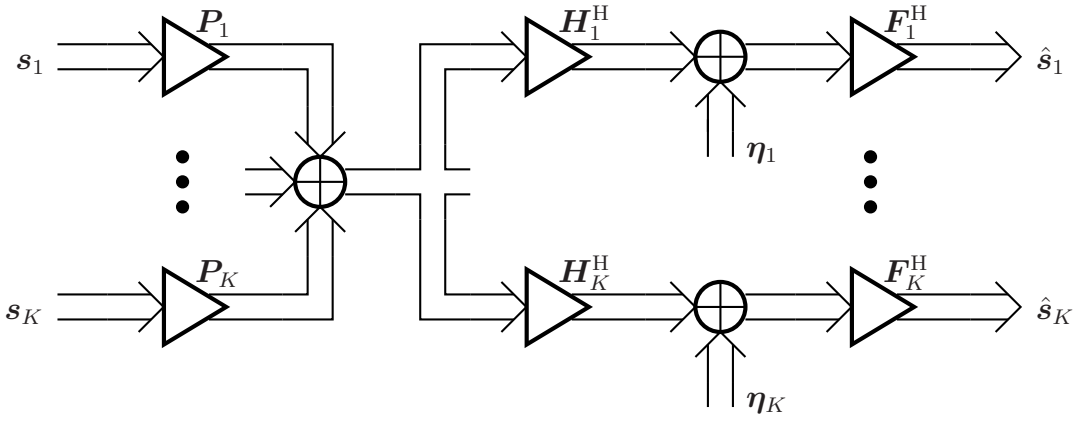


Figure 6.1: Multiple Stream MIMO BC System Model.

symbols are now vectors $\boldsymbol{s}_k \in \mathbb{C}^{d_k}$ comprising the d_k data streams sent to for the kth user, $k \in \{1, \ldots, K\}$. Such data vectors are zero-mean Gaussian with the covariance $\mathrm{E}[\boldsymbol{s}_k \boldsymbol{s}_k^{\mathrm{H}}] = \mathbf{I}_{d_k}$, $\boldsymbol{s}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_{d_k})$. Moreover, the signals corresponding to different users are independent, i.e. $\mathrm{E}[\boldsymbol{s}_k \boldsymbol{s}_l^{\mathrm{H}}] = \mathbf{0}$ when $l \neq k$. Prior to be transmitted, the data vector is precoded by the matrix $\boldsymbol{P}_k \in \mathbb{C}^{N \times d_k}$ to produce the signal that propagates over the MIMO channel $\boldsymbol{H}_k \in \mathbb{C}^{N \times R}$. The signal received is then filtered with a linear filter $\boldsymbol{F}_k \in \mathbb{C}^{R \times d_k}$ to produce the following estimates of the transmitted data

$$\hat{\boldsymbol{s}}_k = \boldsymbol{F}_k^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}} \sum\nolimits_{i=1}^{K} \boldsymbol{P}_i \boldsymbol{s}_i + \boldsymbol{F}_k^{\mathrm{H}} \boldsymbol{\eta}_k, \tag{6.1}$$

where $\boldsymbol{\eta}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{\eta}_k})$ is the kth user's *Additive White Gaussian Noise* (AWGN) which is independent of the transmitted symbols. For the multiple stream MIMO BC, the rate in (3.13) is particularized as

$$R_k = \log_2 \det \left(\mathbf{I}_R + \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{P}_k \boldsymbol{P}_k^{\mathrm{H}} \boldsymbol{H}_k \left(\boldsymbol{C}_{\boldsymbol{\eta}_k} + \boldsymbol{H}_k^{\mathrm{H}} \sum_{i \neq k} \boldsymbol{P}_i \boldsymbol{P}_i^{\mathrm{H}} \boldsymbol{H}_k \right)^{-1} \right). \tag{6.2}$$

Note that the total transmit power for this scenario can be obtained as $P_T = \sum_{k=1}^{K} ||\boldsymbol{P}_k||_{\mathrm{F}}^2$.

Similarly to the scenarios presented in Chapters 4 and 5, we assume that perfect CSI is available at the receivers. The transmitter, however, only has a partial knowledge of the CSI modeled by v and the conditional *probability density function* (pdf)s $f_{\boldsymbol{H}_k|v}(\boldsymbol{H}_k|v)$. No additional assumptions regarding $f_{\boldsymbol{H}_k|v}(\boldsymbol{H}_k|v)$ are made. Since the error model is identical to that of Section 5.1 of Chapter 5, the details will not be repeated here. Hence,

$$\boldsymbol{H}_k = \bar{\boldsymbol{H}}_k + \tilde{\boldsymbol{H}}_k, \tag{6.3}$$

where $\bar{\boldsymbol{H}}_k$ is the mean conditioned to v, and $\tilde{\boldsymbol{H}}_k$ is the error due to the imperfect CSIT (see (5.3) for further information).

Partial CSI v has a direct impact on the design of the transmit filters $\{\boldsymbol{P}_k\}_{k=1}^{K}$. Moreover, the expression (6.2) is no longer valid due to the imperfect CSIT, and the ergodic rate will be used instead as the performance measure of interest (see Section 3.1.1 of Chapter 3). According to (3.16), due to Bayes' rule, the average rate is conditioned to the imperfect CSIT leading to

$$\mathrm{E}[R_k|\, v] = \mathrm{E}\left[\log_2 \det\left(\mathbf{I}_R + \boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{P}_k\boldsymbol{P}_k^{\mathrm{H}}\boldsymbol{H}_k\left(\boldsymbol{C}_{\boldsymbol{\eta}_k} + \boldsymbol{H}_k^{\mathrm{H}}\sum_{i\neq k}\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}}\boldsymbol{H}_k\right)^{-1}\right)|\, v\right]. \tag{6.4}$$

Our goal is to minimize the total transmit power P_T while ensuring the kth user's average BC information rate is larger than a given value ρ_k, i.e.

$$\min_{\{\boldsymbol{P}_k\}_{k=1}^{K}} P_T = \sum_{k=1}^{K}\|\boldsymbol{P}_k\|_{\mathrm{F}}^2 \quad \text{subject to} \quad \mathrm{E}\left[R_k|\, v\right] \geq \rho_k\, \forall k, \tag{6.5}$$

where $\boldsymbol{P}_k$ depends on v. Note that $\mathrm{E}[R_k|\, v] \geq \rho_k$ implies that $\mathrm{E}[R_k] \geq \rho_k$, i.e., an ergodic rate of ρ_k is ensured with precoders adapted to different v.

Solving the optimization problem (6.5) is very difficult due to the expectation in the constraints. However, we can exploit the relationship between rate and *Minimum Mean Square Error* (MMSE) to rewrite the constraints of (6.5) in a more manageable way that ensures the original average rate requirements are fulfilled. The approximation successfully employed in Section 5.1.1 of Chapter 5 is not applicable here, as we will see in the following. Also notice that no restriction on the rate of every stream has been made. Therefore, there exists freedom to distribute the per-user rate among the streams. Although we can take advantage of such flexibility, it leads to additional complexity for the optimization problem.

6.1.1. Problem Formulation

First, we extend the definition of the BC *Mean Square Error* (MSE) shown in Section 3.2 of Chapter 3 for the case of vector data signal, i.e.

$$\mathrm{MSE}_k^{\mathrm{BC}} = \mathrm{E}\left[\|\boldsymbol{s}_k - \hat{\boldsymbol{s}}_k\|_2^2\right] = \mathrm{E}\left[\mathrm{tr}\left(\left(\boldsymbol{s}_k - \hat{\boldsymbol{s}}_k\right)\left(\boldsymbol{s}_k - \hat{\boldsymbol{s}}_k\right)^{\mathrm{H}}\right)\right]. \tag{6.6}$$

Elaborating (6.6), we obtain

$$\mathrm{MSE}_k^{\mathrm{BC}} = \mathrm{tr}\left(\mathbf{I}_{d_k} - 2\Re\left\{\boldsymbol{F}_k^{\mathrm{H}}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{P}_k\right\} + \boldsymbol{F}_k^{\mathrm{H}}\boldsymbol{H}_k^{\mathrm{H}}\sum_{i=1}^{K}\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{F}_k + \boldsymbol{F}_k^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{\eta}_k}\boldsymbol{F}_k\right). \tag{6.7}$$

Taking into account the imperfect CSIT, the appropriate MSE measure is the conditioned average MSE (see Section 3.2).

$$\overline{\mathrm{MSE}}_k^{\mathrm{BC}} = \mathrm{E}\left[\mathrm{tr}\left(\mathbf{I}_{d_k} - 2\Re\left\{\boldsymbol{F}_k^{\mathrm{H}}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{P}_k\right\} + \boldsymbol{F}_k^{\mathrm{H}}\boldsymbol{H}_k^{\mathrm{H}}\sum_{i=1}^{K}\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{F}_k + \boldsymbol{F}_k^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{\eta}_k}\boldsymbol{F}_k\right) \mid v\right]. \tag{6.8}$$

This is in accordance with the rate proposed in (6.4).

Note that the optimal receive filters that minimize (6.7) are also the minimizers of (6.8) since there exists perfect *Channel State Information at the Receiver* (CSIR). Such filters are calculated following the same process as in Section 5.1.1 of Chapter 5 for the single stream MIMO case

$$\begin{aligned}\boldsymbol{F}_k^{\mathrm{MMSE}} &= \mathrm{argmin}_{\boldsymbol{F}_k}\,\mathrm{E}\left[\,\|\boldsymbol{s}_k - \hat{\boldsymbol{s}}_k\|_2^2 \big|\, \boldsymbol{H}_k\right] \\ &= \left(\boldsymbol{H}_k^{\mathrm{H}}\sum_{i=1}^{K}\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}}\boldsymbol{H}_k + \boldsymbol{C}_{\boldsymbol{\eta}_k}\right)^{-1}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{P}_k,\end{aligned} \tag{6.9}$$

where the last equality is obtained taking the derivative of the $\mathrm{MSE}_k^{\mathrm{BC}}$ in (6.7) with respect to $\boldsymbol{F}_k^*$, accordingly to the results presented in Appendix C (see Section C.6), and equating such derivative to zero.

We next obtain the average MMSE for user k plugging $\boldsymbol{F}_k^{\mathrm{MMSE}}$ from (6.9) into the average MSE expression (6.8) as follows

$$\begin{aligned}\overline{\mathrm{MMSE}}_k^{\mathrm{BC}} &= \mathrm{E}\left[\mathrm{tr}\left(\mathbf{I}_{d_k} - \boldsymbol{P}_k^{\mathrm{H}}\boldsymbol{H}_k\left(\boldsymbol{H}_k^{\mathrm{H}}\sum_{i=1}^{K}\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}}\boldsymbol{H}_k + \boldsymbol{C}_{\boldsymbol{\eta}_k}\right)^{-1}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{P}_k\right) \mid v\right] \\ &= \mathrm{E}\left[\mathrm{tr}\left(\mathbf{I}_{d_k} + \boldsymbol{P}_k^{\mathrm{H}}\boldsymbol{H}_k\left(\boldsymbol{H}_k^{\mathrm{H}}\sum_{i\neq k}\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}}\boldsymbol{H}_k + \boldsymbol{C}_{\boldsymbol{\eta}_k}\right)^{-1}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{P}_k\right)^{-1} \mid v\right]\end{aligned} \tag{6.10}$$

$$= \mathrm{E}\left[\mathrm{tr}\left(\boldsymbol{\Sigma}_k\right) \mid v\right], \tag{6.11}$$

where (6.10) is obtained exploiting the matrix inversion lemma (see Section C.4 of Appendix C).

Observe now that, applying the Sylvester's theorem (see Section C.2 of Appendix C) to (6.10), the determinant equality leads to

$$\mathrm{E}\left[R_k \mid v\right] = \mathrm{E}\left[\log_2\left(\det\left(\boldsymbol{\Sigma}_k^{-1}\right)\right) \mid v\right]. \tag{6.12}$$

Notice that, contrary to the single stream MIMO BC case, where the relationship between the rate and the MMSE can be reduced to a scalar equality after getting rid of the

determinant via Sylvester's theorem (see (5.16)), we will have to use a different method for the case of multiple stream MIMO approach.

We take advantage in the following of the properties of positive definite matrices (see Section C.5 of Appendix C). Let us consider the matrix

$$\boldsymbol{P}_k^{\mathrm{H}}\boldsymbol{H}_k\left(\boldsymbol{H}_k^{\mathrm{H}}\sum_{i\neq k}\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}}\boldsymbol{H}_k+\boldsymbol{C}_{\boldsymbol{\eta}_k}\right)^{-1}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{P}_k. \tag{6.13}$$

Due to the results presented in Section C.5 of Appendix C for the Gram of a matrix, the product $\boldsymbol{B}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{B}$ with $\boldsymbol{A}$ being positive semidefinite, the inverse, and the linearity, we conclude that the matrix of (6.13) is positive semidefinite. Therefore, $\boldsymbol{\Sigma}_k$ and also $\mathrm{E}\left[\boldsymbol{\Sigma}_k|\, v\right]$ are positive semidefinite matrices.

Let us introduce the eigenvalue decomposition (see Section C.1 of Appendix C)

$$\mathrm{E}[\boldsymbol{\Sigma}_k|\, v]=\boldsymbol{U}_k\boldsymbol{\Lambda}_k\boldsymbol{U}_k^{\mathrm{H}}, \tag{6.14}$$

with the unitary matrix $\boldsymbol{U}_k$ and the diagonal matrix $\boldsymbol{\Lambda}_k=\mathrm{diag}(\lambda_{k,1},\ldots,\lambda_{k,d_k})$ containing the eigenvalues with $\lambda_{k,i}\geq 0,\forall k,i$. The basis $\boldsymbol{U}_k$ allows us to find the spatial decorrelation precoders

$$\boldsymbol{P}_k'=\boldsymbol{P}_k\boldsymbol{U}_k. \tag{6.15}$$

Such precoders remove the off-diagonal elements of $\mathrm{E}[\boldsymbol{\Sigma}_k|\, v]$ for all k, since employing $\boldsymbol{P}_k'$ results in the following diagonal matrix

$$\mathrm{E}\left[\left(\mathbf{I}_{d_k}+\boldsymbol{U}_k^{\mathrm{H}}\boldsymbol{P}_k^{\mathrm{H}}\boldsymbol{H}_k\left(\boldsymbol{H}_k^{\mathrm{H}}\sum_{i\neq k}\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}}\boldsymbol{H}_k+\boldsymbol{C}_{\boldsymbol{\eta}_k}\right)^{-1}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{P}_k\boldsymbol{U}_k\right)^{-1}|\, v\right]. \tag{6.16}$$

Observe that the use of the precoders $\boldsymbol{P}_k'$ does not change the total transmit power

$$\sum_{k=1}^{K}\|\boldsymbol{P}_k'\|_{\mathrm{F}}^2=\sum_{k=1}^{K}\mathrm{tr}\left(\boldsymbol{P}_k\boldsymbol{U}_k\boldsymbol{U}_k^{\mathrm{H}}\boldsymbol{P}_k^{\mathrm{H}}\right)=\sum_{k=1}^{K}\mathrm{tr}\left(\boldsymbol{P}_k\boldsymbol{P}_k^{\mathrm{H}}\right), \tag{6.17}$$

nor the expressions of the average rate of (6.4) and the average MMSE of (6.10). Henceforth, we consider that the spatial decorrelation precoders $\boldsymbol{P}_k'$ can be incorporated without loss of generality.

Recall now the eigenvalue decomposition in (6.14). Thus, the per-user average MMSE in the BC is

$$\overline{\mathrm{MMSE}}_k^{\mathrm{BC}}=\mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{\Sigma}_k|v\right]\right)=\sum_{i=1}^{d_k}\lambda_{k,i}. \tag{6.18}$$

Notice that $\lambda_{k,i}$ can be interpreted as the kth user's ith stream average MMSE, i.e.

$$\overline{\text{MMSE}}_{k,i}^{\text{BC}} = \lambda_{k,i}, \tag{6.19}$$

and, as a consequence, the average MMSE for some user is the sum of the average MMSEs of the streams allocated for such user.

Accordingly, since $f(\boldsymbol{A}) = -\log(\det(\boldsymbol{A}))$, with $\boldsymbol{A}$ positive semidefinite, is convex (see Section F.2 of Appendix F), applying Jensen's inequality to $\mathrm{E}[R_k|\, v]$ yields

$$\mathrm{E}[R_k|\, v] \geq -\log_2 \det\left(\mathrm{E}\left[\boldsymbol{\Sigma}_k|\, v\right]\right) = -\sum_{i=1}^{d_k} \log_2(\lambda_{k,i}). \tag{6.20}$$

Remember that our goal is to minimize the total transmit power while ensuring certain *Quality-of-Service* (QoS) for all the users. Let us now define the per-stream target rate for the user k and the stream i, i.e. $\varrho_{k,i}$. Consequently, to ensure $\mathrm{E}[R_k^{\text{BC}}] \geq \rho_k$ with $\rho_k = \sum_{i=1}^{d_k} \varrho_{k,i}$, we use

$$\overline{\text{MMSE}}_{k,i}^{\text{BC}} = \lambda_{k,i} \leq 2^{-\varrho_{k,i}}. \tag{6.21}$$

In other words, splitting the target rate ρ_k into per-stream target rates $\varrho_{k,i}$ enables us to rewrite (6.5) in terms of average MMSE constraints rather than rate constraints. As long as there is not restriction over the per-stream average rates, we can choose the targets $\varrho_{k,i}$ in a smart way that allows to achieve lower values for the total transmit power. Furthermore, note that when this new per-stream constraints are included, the optimization is more stringent than the original one. Thus, the per-user rate constraints are guaranteed. However, the decision about the distribution among the streams of the per-user rates constitutes an additional level of complexity that has to be taken into account for the minimization problem. For this reason, we propose a nested optimization.

The outer optimization tackles the way of sharing the original average targets between the streams of a certain user, and reads as

$$\min_{\{\boldsymbol{\varrho}_k\}_{k=1}^K} P_T(\boldsymbol{\varrho}) \quad \text{subject to } \mathbf{1}^{\mathrm{T}}\boldsymbol{\varrho}_k = \rho_k, \quad \text{and} \quad \boldsymbol{\varrho}_k \geq \mathbf{0}\ \forall k, \tag{6.22}$$

with $\boldsymbol{\varrho} = [\boldsymbol{\varrho}_1^{\mathrm{T}}, \ldots, \boldsymbol{\varrho}_K^{\mathrm{T}}]^{\mathrm{T}}$, and $\boldsymbol{\varrho}_k = [\varrho_{k,1}, \ldots, \varrho_{k,d_k}]^{\mathrm{T}}$. The inner optimization deals with the minimization of the total transmit power for given per-stream average target rates, and uses the per-stream average MMSE restrictions from (6.21), i.e.

$$P_T(\boldsymbol{\varrho}) = \min_{\{\boldsymbol{P}_k, \boldsymbol{F}_k\}_{k=1}^K} \sum_{k=1}^{K} \|\boldsymbol{P}_k\|_{\mathrm{F}}^2 \quad \text{subject to} \quad \overline{\text{MMSE}}_{k,i}^{\text{BC}} \leq 2^{-\varrho_{k,i}}\ \forall k, i. \tag{6.23}$$

In the ensuing section, we apply the same method of Section 5.1 of Chapter 5 to solve the minimization problem (6.23). Additionally, the solution of (6.23) becomes the starting point for the study of the outer problem (6.22).

6.1.2. Per-Stream MMSE Filters

The advantage of using the spatial decorrelation precoders $\boldsymbol{P}'_k = \boldsymbol{P}_k\boldsymbol{U}_k$ is that the intra-user interference is removed for the average MMSE, as seen in (6.18). Therefore, each of the streams for every user can be treated as virtual users. Consequently, we introduce $s_{k,i} = \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{s}_k$, $\boldsymbol{f}_{k,i} = \boldsymbol{F}_k\boldsymbol{e}_i$, and $\boldsymbol{p}_{k,i} = \boldsymbol{P}_k\boldsymbol{e}_i$ as the data, the receive filter, and the precoder for the kth user's ith stream, respectively.

As previously shown in (6.9), it is straightforward to find the optimal MMSE receive filters, $\boldsymbol{f}_{k,i}^{\mathrm{MMSE}} = \boldsymbol{F}_{k,\mathrm{MMSE}}\boldsymbol{e}_i$ for given channels and precoders. On the other hand, finding the optimal BC precoders according to (6.23) turns out to be a more difficult problem (see Section 5.1.2 of Chapter 5). Nevertheless, we have obtained a solution based on the BC/*Multiple Access Channel* (MAC) MSE duality. According to that duality, the optimal BC precoders $\boldsymbol{p}_{k,i}$ can be found as the optimal receive filters $\boldsymbol{g}_{k,i} \in \mathbb{C}^N$ that minimize the dual MAC MSE

$$\begin{aligned}\overline{\mathrm{MSE}}_{k,i}^{\mathrm{MAC}} = 1 &- 2\Re\left\{\boldsymbol{g}_{k,i}^{\mathrm{H}}\,\mathrm{E}\left[\boldsymbol{H}_k\boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2}\boldsymbol{t}_{k,i}|\, v\right]\right\} \\ &+ \boldsymbol{g}_{k,i}^{\mathrm{H}}\sum_{l=1}^{K}\mathrm{E}\left[\boldsymbol{H}_l\boldsymbol{C}_{\boldsymbol{\eta}_l}^{-\mathrm{H}/2}\boldsymbol{T}_l\boldsymbol{T}_l^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{\eta}_l}^{-1/2}\boldsymbol{H}_l^{\mathrm{H}}|\, v\right]\boldsymbol{g}_{k,i} + \|\boldsymbol{g}_{k,i}\|_2^2\,. \end{aligned} \tag{6.24}$$

The MAC precoders for the kth user's ith stream are denoted as $\boldsymbol{t}_{k,i} = \boldsymbol{T}_k\boldsymbol{e}_i \in \mathbb{C}^R$.

We now solve the power minimization problem (6.23) using the method proposed in Section 5.1.3 of Chapter 5 for the MIMO scenario. To that end, the following matrices collecting M realizations of the Monte Carlo numerical integration are defined

$$\boldsymbol{\Theta}_k = \left[\boldsymbol{H}_k^{(1)}\boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2}, \ldots, \boldsymbol{H}_k^{(M)}\boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2}\right], \tag{6.25}$$

$$\boldsymbol{\Phi}_{k,i} = \mathrm{blockdiag}\left(\boldsymbol{\tau}_{k,i}^{(1)}, \ldots, \boldsymbol{\tau}_{k,i}^{(M)}\right), \tag{6.26}$$

where $\boldsymbol{t}_{k,i}^{(n)} = \sqrt{\xi_{k,i}}\boldsymbol{\tau}_{k,i}^{(n)}$, with $\xi_{k,i} = \frac{1}{M}\sum_{n=1}^{M}\|\boldsymbol{t}_{k,i}^{(n)}\|_2^2$. Note that $\xi_{k,i}$ can be interpreted as the power allocated to the kth user's ith stream. Based on these definitions, the dual MAC MSE can be approximated as

$$\overline{\mathrm{MSE}}_{k,i}^{\mathrm{MAC}} = 1 - \frac{2}{M}\Re\left\{\boldsymbol{g}_{k,i}^{\mathrm{H}}\boldsymbol{\Theta}_k\boldsymbol{\Phi}_{k,i}\mathbf{1}\sqrt{\xi_{k,i}}\right\} + y_{k,i}, \tag{6.27}$$

with

$$y_{k,i} = \frac{1}{M}\sum_{l=1}^{K}\sum_{j=1}^{d_l}\xi_{l,j}\left\|\boldsymbol{g}_{k,i}^{\mathrm{H}}\boldsymbol{\Theta}_l\boldsymbol{\Phi}_{l,j}\right\|_2^2 + \|\boldsymbol{g}_{k,i}\|_2^2\,. \tag{6.28}$$

The optimal MAC receive filters are the MMSE filters $\boldsymbol{g}_{k,i}^{\mathrm{MMSE}}$. Additionally, due to normalization reasons, we introduce the scalar receiver $r_{k,i}$ so that $\boldsymbol{g}_{k,i}^{\mathrm{MMSE}} = r_{k,i}\tilde{\boldsymbol{g}}_{k,i}$.

The optimal scalar receive filter is denoted as $r_{k,i}^{\text{MMSE}}$. Therefore

$$\boldsymbol{g}_{k,i}^{\text{MMSE}} = \left(\sum_{l=1}^{K} \sum_{j=1}^{d_l} \frac{\xi_{l,j}}{M} \boldsymbol{\Theta}_l \boldsymbol{\Phi}_{l,j} \boldsymbol{\Phi}_{l,j}^{\text{H}} \boldsymbol{\Theta}_l^{\text{H}} + \frac{1}{M} \sqrt{\xi_{k,i}} \mathbf{I}_N \right)^{-1} \boldsymbol{\Theta}_k \boldsymbol{\Phi}_{k,i} \mathbf{1}, \tag{6.29}$$

$$r_{k,i}^{\text{MMSE}} = \frac{\frac{1}{M} \tilde{\boldsymbol{g}}_{k,i}^{\text{H}} \boldsymbol{\Theta}_k \boldsymbol{\Phi}_{k,i} \mathbf{1} \sqrt{\xi_{k,i}}}{\frac{1}{M} \tilde{\boldsymbol{g}}_{k,i}^{\text{H}} \sum_{l=1}^{K} \sum_{j=1}^{d_l} \xi_{l,j} \boldsymbol{\Theta}_l \boldsymbol{\Phi}_{l,j} \boldsymbol{\Phi}_{l,j}^{\text{H}} \boldsymbol{\Theta}_l^{\text{H}} \tilde{\boldsymbol{g}}_{k,i} + \|\tilde{\boldsymbol{g}}_{k,i}\|_2^2}. \tag{6.30}$$

Substituting the optimal MMSE scalar filters for given MAC transmit and receive filters, $r_{k,i}^{\text{MMSE}}$, into (6.27) gives the following expression for the kth user ith stream scalar MMSE

$$\overline{\text{MMSE}}_{k,i,\text{scalar}}^{\text{MAC}} = 1 - \frac{\xi_{k,i}}{M^2} \left| \tilde{\boldsymbol{g}}_{k,i}^{\text{H}} \boldsymbol{\Theta}_k \boldsymbol{\Phi}_{k,i} \mathbf{1} \right|^2 y_{k,i}^{-1}, \tag{6.31}$$

where $\boldsymbol{g}_{k,i}$ must be replaced by $\tilde{\boldsymbol{g}}_{k,i}$ in the expression for $y_{k,i}$. For further details involving the development above described, see the discussion in Section 5.1.3. Note that the equations of this section are equal to those presented in Section 5.1.3 if we consider single stream MIMO virtual users by means of the mapping $\{k, i\} \mapsto z = \sum_{j=1}^{k-1} d_j + i$. That is,

$$\begin{aligned} \tilde{\boldsymbol{g}}_z &= \tilde{\boldsymbol{g}}_{k,i}, \\ \boldsymbol{H}_z &= \boldsymbol{H}_k, \\ \boldsymbol{\Phi}_z &= \boldsymbol{\Phi}_{k,i}, \\ \xi_z &= \xi_{k,i}, \end{aligned}$$

are the single stream MIMO MMSE parameters for the virtual user z, replacing those for the multiple stream MIMO MMSE in (6.31).

6.1.3. Optimization of Per-Stream Target Rates

The BC/MAC duality applied in the previous section, together with the new parameters introduced, enables us to rewrite the inner optimization problem of (6.23) as

$$P_T(\boldsymbol{\varrho}) = \min_{\left\{\boldsymbol{\Phi}_{k,i}, \tilde{\boldsymbol{g}}_{k,i}, \xi_{k,i}\right\}_{k,i}^{K,d_k}} \sum_{m=1}^{K} \sum_{n=1}^{d_m} \xi_{m,n} \quad \text{subject to} \quad \overline{\text{MMSE}}_{k,i,\text{scalar}}^{\text{MAC}} \leq 2^{-\varrho_{k,i}} \forall k, \forall i, \tag{6.32}$$

where $\varrho_{k,i}$ is the kth user's ith stream target rate. For given $\varrho_{k,i}$ (collected in $\boldsymbol{\varrho}$), this optimization problem can be solved similarly to the single stream case based on the interference function framework and the *Alternate Optimization* (AO) (see Section 5.1.3).

An important property of (6.32) is that the constraints are fulfilled with equality in the optimum. Otherwise, the average MMSE for some users could be increased leading to a smaller value of the total transmit power. Note that the algorithmic solution proposed in Section 5.1.4 of Chapter 5 also exhibits this property, i.e., $\overline{\text{MMSE}}^{\text{MAC}}_{k,i,\text{scalar}} = 2^{-\varrho_{k,i}}$.

In the multi-stream case considered in this section, the additional outer optimization (6.22) is necessary to optimally distribute the per-user target rate ρ_k over the data streams of user k such that $\rho_k = \sum_{i=1}^{d_k} \varrho_{k,i}$. In the following, we propose to solve the problem optimization (6.22) by means of a gradient-projection algorithm. It consists on minimizing the transmit power following the direction of the gradient. However, the result of the gradient step has to be projected to the set of feasible values fulfilling the original per-user restrictions.

First, we describe the update rule of the per-stream target rates $\varrho_{k,i}$ according to the following iteration

$$\varrho'_{k,i} = \varrho_{k,i} - s\,\frac{\partial P_T(\boldsymbol{\varrho})}{\partial \varrho_{k,i}}, \tag{6.33}$$

with the step size $s > 0$. Notice that the explicit relationship between the total transmit power P_T and the per-stream targets rates $\varrho_{k,i}$ is not known. Therefore, we exploit the laws of partial differentiation to compute the gradient in (6.33), as we will show in the following.

Let us calculate the derivative of $\overline{\text{MMSE}}^{\text{MAC}}_{k,i,\text{scalar}}$ in (6.31) with respect to the power allocation $\xi_{m,n}$, i.e.

$$\frac{\partial \overline{\text{MMSE}}^{\text{MAC}}_{k,i,\text{scalar}}}{\partial \xi_{m,n}}. \tag{6.34}$$

In order to do that, we distinguish two cases:

1. For $m = k$, $n = i$:

$$\frac{\partial \overline{\text{MMSE}}^{\text{MAC}}_{k,i,\text{scalar}}}{\partial \xi_{k,i}} = -\frac{\frac{1}{M^2}\left|\tilde{\boldsymbol{g}}^{\text{H}}_{k,i}\boldsymbol{\Theta}_k\boldsymbol{\Phi}_{k,i}\mathbf{1}\right|^2\left(y_{k,i} - \frac{\xi_{k,i}}{M}\left\|\tilde{\boldsymbol{g}}^{\text{H}}_{k,i}\boldsymbol{\Theta}_k\boldsymbol{\Phi}_{k,i}\right\|_2^2\right)}{y_{k,i}^2}, \tag{6.35}$$

   with $y_{k,i}$ given by (6.28).

2. Otherwise:

$$\frac{\partial \overline{\text{MMSE}}^{\text{MAC}}_{k,i,\text{scalar}}}{\partial \xi_{m,n}} = \frac{\frac{\xi_{k,i}}{M^3}\left|\tilde{\boldsymbol{g}}^{\text{H}}_{k,i}\boldsymbol{\Theta}_k\boldsymbol{\Phi}_{k,i}\mathbf{1}\right|^2\left\|\tilde{\boldsymbol{g}}^{\text{H}}_{k,i}\boldsymbol{\Theta}_m\boldsymbol{\Phi}_{m,n}\right\|_2^2}{y_{k,i}^2}. \tag{6.36}$$

Recall that the transmit power $P_T(\boldsymbol{\varrho}) = \sum_{k=1}^{K}\sum_{i=1}^{d_k}\xi_{k,i}$ depends on the per-stream targets $\varrho_{k,i}$, $\forall k, i$. Additionally, taking into account that the equality $\overline{\text{MMSE}}^{\text{MAC}}_{k,i,\text{scalar}} =$

$2^{-\varrho_{k,i}}$ holds in the solution of (6.32), the gradient

$$\frac{\partial \overline{\text{MMSE}}^{\text{MAC}}_{k,i,\text{scalar}}}{\partial \varrho_{l,j}} = \sum_{m=1}^{K} \sum_{n=1}^{d_m} \frac{\partial \overline{\text{MMSE}}^{\text{MAC}}_{k,i,\text{scalar}}}{\partial \xi_{m,n}} \frac{\partial \xi_{m,n}}{\partial \varrho_{l,j}}, \tag{6.37}$$

is equal to $-\ln(2)2^{-\varrho_{k,i}}$, when $k = l$ and $i = j$, and 0 otherwise. Let us now define the Jacobian matrix $\boldsymbol{J_f}(\boldsymbol{\xi})$ of the function

$$\boldsymbol{f}(\boldsymbol{\xi}) = \left[\overline{\text{MMSE}}^{\text{MAC}}_{1,1,\text{scalar}}, \ldots, \overline{\text{MMSE}}^{\text{MAC}}_{1,d_1,\text{scalar}}, \ldots, \overline{\text{MMSE}}^{\text{MAC}}_{K,d_K,\text{scalar}}\right]^{\text{T}}, \tag{6.38}$$

as follows

$$[\boldsymbol{J_f}(\boldsymbol{\xi})]_{a,b} = \frac{\partial \overline{\text{MMSE}}^{\text{MAC}}_{k,i,\text{scalar}}}{\partial \xi_{l,j}}, \tag{6.39}$$

where d_k is the number of streams for the user k, $a = \sum_{m=1}^{k-1} d_m + i$, and $b = \sum_{m=1}^{l-1} d_m + j$ (see Section C.7 of Appendix C for a discussion about Jacobian matrices). Similarly, the matrix comprising the partial derivatives of the total average power with respect to the per-stream target rates is defined as

$$\boldsymbol{J_\xi}(\boldsymbol{\varrho}) = \frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{\varrho}^{\text{T}}}. \tag{6.40}$$

By employing the previously defined Jacobian matrices, we can equivalently write the partial derivatives of (6.37) as a matrix product. Moreover, since the constraints are fulfilled with equality at the optimum of (6.32), we get

$$\frac{\partial \overline{\text{MMSE}}^{\text{MAC}}_{k,i}}{\partial \varrho_{l,j}} = \left[\boldsymbol{J_f}\left(\boldsymbol{\xi}\right) \boldsymbol{J_\xi}\left(\boldsymbol{\varrho}\right)\right]_{a,b} = -\ln(2)\left[\boldsymbol{W}\right]_{a,b}, \tag{6.41}$$

with $\boldsymbol{W}$ being the matrix collecting the inverse of the average MMSE targets, i.e.

$$\boldsymbol{W} = \text{diag}\left(2^{\varrho_{1,1}}, \ldots, 2^{\varrho_{1,d_1}}, \ldots, 2^{\varrho_{K,d_K}}\right). \tag{6.42}$$

The Jacobian matrix $\boldsymbol{J_\xi}(\boldsymbol{\varrho})$ is now obtained by left multiplying times the inverse of $\boldsymbol{J_f}(\boldsymbol{\xi})$ in (6.41) as follows

$$\boldsymbol{J_\xi}(\boldsymbol{\varrho}) = -\ln(2)\boldsymbol{J_f}(\boldsymbol{\xi})^{-1}\boldsymbol{W}. \tag{6.43}$$

Hence, we obtained the partial derivatives of the total transmit power with respect to the per-stream target rates by using the properties of partial differentiation, even though no explicit relationship has been found. As a consequence, note that $\boldsymbol{J_\xi}(\boldsymbol{\varrho})$ contains the partial derivatives necessary for the gradient step in (6.33). Therefore,

$$\frac{\partial P_T(\boldsymbol{\varrho})}{\partial \varrho_{k,i}} = -\ln(2)\mathbf{1}^{\text{T}}\boldsymbol{J_f}(\boldsymbol{\xi})^{-1}\boldsymbol{W}\boldsymbol{e}_{\sum_{m=1}^{k-1} d_m + i}, \tag{6.44}$$

where $\mathbf{1}$ and $\boldsymbol{e}_i$ are the all ones and the canonical vectors, respectively.

We now prove that the matrix $-\boldsymbol{J_f}(\boldsymbol{\xi})$ is a Z-matrix, i.e., the diagonal is positive and the off-diagonal ones are negative. We start defining a diagonal matrix containing the power allocation $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{\xi})$. Now, for the matrix $-\boldsymbol{J_f}(\boldsymbol{\xi})\boldsymbol{D}$, we compare the sum of the absolute values for the off-diagonal elements with the diagonal element at the row $a = \sum_{j=1}^{k-1} d_j + i$, corresponding to the stream i of user j, i.e.

$$\begin{aligned}
&\sum_{b\neq a} \left| [-\boldsymbol{J_f}(\boldsymbol{\xi})\boldsymbol{D}]_{a,b} \right| \\
&= \frac{\frac{\xi_{k,i}}{M^2} \left| \boldsymbol{g}_{k,i}^{\mathrm{H}} \boldsymbol{\Theta}_k \boldsymbol{\Phi}_{k,i} \mathbf{1} \right|^2 \frac{1}{M} \boldsymbol{g}_{k,i}^{\mathrm{H}} \sum_{\substack{l=1 \\ l,j\neq k,i}}^{K} \sum_{j=1}^{d_l} \xi_{l,j} \boldsymbol{\Theta}_l \boldsymbol{\Phi}_{l,j} \boldsymbol{\Phi}_{l,j}^{\mathrm{H}} \boldsymbol{\Theta}_l^{\mathrm{H}} \boldsymbol{g}_{k,i}}{y_{k,i}^2} \\
&< \frac{\frac{\xi_{k,i}}{M^2} \left| \boldsymbol{g}_{k,i}^{\mathrm{H}} \boldsymbol{\Theta}_k \boldsymbol{\Phi}_{k,i} \mathbf{1} \right|^2 \left(\frac{1}{M} \boldsymbol{g}_{k,i}^{\mathrm{H}} \sum_{\substack{l=1 \\ l,j\neq k,i}}^{K} \sum_{j=1}^{d_l} \xi_{l,j} \boldsymbol{\Theta}_l \boldsymbol{\Phi}_{l,j} \boldsymbol{\Phi}_{l,j}^{\mathrm{H}} \boldsymbol{\Theta}_l^{\mathrm{H}} \boldsymbol{g}_{k,i} + \| \boldsymbol{g}_{k,i} \|_2^2 \right)}{y_{k,i}^2} \\
&= [-\boldsymbol{J_f}(\boldsymbol{\xi})\boldsymbol{D}]_{a,a} \,. \qquad (6.45)
\end{aligned}$$

Since the inequality holds for all a, $\boldsymbol{J_f}(\boldsymbol{\xi})\boldsymbol{D}$ is strictly diagonally dominant and $-\boldsymbol{J_f}(\boldsymbol{\xi})$ is a non-singular M-matrix with positive inverse [65]. This result meets the intuition that a lower target rate $\varrho_{l,j}$ also leads to lower transmit power $P_T(\boldsymbol{\varrho})$.

It is important to note that by means of the target updating (6.33), i.e., $\varrho'_{k,i} = \varrho_{k,i} - s\frac{\partial P_T(\boldsymbol{\varrho})}{\partial \varrho_{k,i}}$, the per-stream targets are reduced. Since no additional restrictions are imposed, the reduction is individually performed for each of the streams. This way, the updated per-stream target rates do not fulfill the constraints $\sum_{i=1}^{d_k} \varrho'_{k,i} = \rho_k$ anymore. Therefore, a projection to the set of the kth user's feasible target rates can be performed to force the updated targets satisfying the given per-user restrictions. We propose to perform such a projection by minimizing the following Euclidean distance

$$\min_{\boldsymbol{\varrho}_k} \sum_{i=1}^{d_k} (\varrho_{k,i} - \varrho'_{k,i})^2 \quad \text{subject to} \quad \sum_{i=1}^{d_k} \varrho_{k,i} = \rho_k,\ \varrho_{k,i} \geq 0,\ \forall i. \qquad (6.46)$$

This minimization problem can be solved using the Lagrangian function

$$L\left(\boldsymbol{\varrho}_k, \boldsymbol{\lambda}_k, \mu_k\right) = \sum_{i=1}^{d_k} \left(\varrho_{k,i} - \varrho'_{k,i}\right)^2 - \lambda_{k,i}\varrho_{k,i} + \mu_k \left(\sum_{i=1}^{d_k} \varrho_{k,i} - \rho_k \right), \qquad (6.47)$$

with the Lagrange multipliers $\lambda_{k,i}$ and μ_k. Note that (6.47) includes the constraints. Hence, the minimization can be perform by computing the derivative with respect to the targets, i.e.

$$\frac{\partial L\left(\boldsymbol{\varrho}_k, \boldsymbol{\lambda}_k, \mu\right)}{\partial \varrho_{k,i}} = 2\left(\varrho_{k,i} - \varrho'_{k,i}\right) - \lambda_{k,i} + \mu_k. \qquad (6.48)$$

Observe that μ_k is common to all the partial derivatives. It is the factor that takes into account the per-user rate constraints. The *Karush-Kuhn-Tucker* (KKT) conditions are then as follows (see Appendix D)

$$2\left(\varrho_{k,i} - \varrho'_{k,i}\right) - \lambda_{k,i} + \mu_k = 0, \tag{6.49}$$

$$\lambda_{k,i} \geq 0, \tag{6.50}$$

$$\mu_k \geq 0, \tag{6.51}$$

$$\sum_{i=1}^{d_k} \varrho_{k,i} = \rho_k. \tag{6.52}$$

The KKT conditions are sufficient for the optimum of (6.46) as shown in Section F.3 of Appendix F. Accordingly, we find the projection

$$\begin{aligned} \varrho_{k,i} &= \max\left\{\varrho'_{k,i} - \mu_k, 0\right\}, \\ \mu_k &= \frac{1}{d_k}\left(\sum_{i=1}^{d_k} \varrho'_{k,i} - \rho_k\right). \end{aligned} \tag{6.53}$$

Note that after the projection (6.53) only some of the per-stream targets of the user k, but never all of them, could be switched off (i.e. $\varrho_{k,i} = 0$). This way, we obtain the closest per-stream targets to the ones resulting from the gradient step fulfilling the per-user rate restrictions.

6.1.4. Algorithmic Solution

The algorithm referred to as Algorithm 6.1: PM.Pr.Gradient implements the method proposed in this section to solve the power minimization in the multiple stream MIMO BC. In the line 1 both the precoders and the per-stream target rates are randomly initialized. Furthermore, the corresponding spatial decorrelation precoders are computed. Note that considering such spatial decorrelation precoders the transmit power, the average rates, and the MMSEs remain unchanged. Hence, the power minimization (6.23) can be solved by exploiting the Algorithm 5.1: PM.MIMO since every stream is treated as a virtual user (see the line 2).

The algorithm performs an steepest descent method, where we move towards the direction of negative gradients. That way, every iteration reduces the total transmit power or remains unchanged by employing two nested loops. In the outer loop, the gradient is computed in the line 5. Also, the step size is initialized prior to entering into the inner loop. The line 7 updates the per-stream target rates $\varrho_{k,i}$ according to (6.33). Observe that the given per-user constraints are no longer fulfilled after that updating. As a consequence,

Algorithm 6.1: PM.Pr.Gradient. Power Minimization via Projected-Gradient

$\ell \leftarrow 0$, random initialization: $\boldsymbol{P}_k$ and $\varrho_{k,i}^{(0)}$, $\boldsymbol{P}_k \leftarrow \boldsymbol{P}'_k, \forall k$
find the optimum $\boldsymbol{t}_{k,i}^{(0)}$, $\boldsymbol{g}_{k,i}^{(0)}$, and $\xi_{k,i}^{(0)}$ $\forall k, i$. Solve (6.23) via Algorithm 5.1
repeat
  $\ell \leftarrow \ell + 1$
  $\frac{\partial P_T(\boldsymbol{\varrho}^{(\ell)})}{\partial \varrho_{k,i}^{(\ell)}} \leftarrow$ compute the gradient of (6.44), $b_{\text{exit}} \leftarrow 0$, $s \leftarrow s_0$
  repeat
    $\varrho'_{k,i} \leftarrow \varrho_{k,i}^{(\ell-1)} - s\frac{\partial P_T(\boldsymbol{\varrho}^{(\ell)})}{\partial \varrho_{k,i}^{(\ell)}}$, $\forall k, i$. Perform the gradient step (6.33)
    $\varrho_{k,i}^{(\ell)} \leftarrow \max\left\{\varrho'_{k,i} - \mu_k^{(\ell)}, 0\right\}$ $\forall k, i$. Projection to the feasible targets set (6.53)
    find the optimum $\boldsymbol{t}_{k,i}^{(\ell)}$, $\boldsymbol{g}_{k,i}^{(\ell)}$, and $\xi_{k,i}^{(\ell)}$, $\forall k, i$. Solve (6.23) via Algorithm 5.1
    if $P_T^{(\ell-1)} - P_T^{(\ell)} > 0$ **then**
      $b_{\text{exit}} \leftarrow 1$
    else
      $s \leftarrow \frac{s}{2}$. Step size update
    end if
  until $b_{\text{exit}} = 1$
until $\sum_{k=1}^{K}\sum_{i=1}^{d_k} \xi_{k,i}^{(\ell-1)} - \sum_{k=1}^{K}\sum_{i=1}^{d_k} \xi_{k,i}^{(\ell)} \leq \delta$

the projection to the set of feasible solutions is implemented in the line 8 (see the proposed solution for (6.46)).

Next, the power minimization of (6.23) is updated (see the line 9). Then, if the BC total power is smaller than that achieved in the previous iteration, the per-stream target rates and the corresponding transmit and receive filters are updated. If not, the step size s is reduced in the line 13 and the inner loop is repeated until more appropriate per-stream targets are found. If the initial QoS constraints are feasible (see Section 5.2 of Chapter 5), the convergence to a local minimum is guaranteed since in every iteration the power decreases or remains unchanged. Consider the boundary (5.59) and the interpretation of every stream as a virtual user. Then, if the initial constraints are feasible any update of the per-stream targets satisfying the per-user targets will also be feasible. Finally, we set the threshold δ (line 16) to check whether we have reached convergence or not. With the resulting filters for the virtual users in the MAC $\boldsymbol{t}_{k,i}^{(l)}$, $\boldsymbol{g}_{k,i}^{(l)}$, and $\xi_{k,i}^{(l)}$, $\forall k, i$, the reconstruction of the BC transmit and receive filters $\boldsymbol{P}_k$, $\boldsymbol{F}_k$, $\forall k$, is done via the BC/MAC duality.

6.1.5. Simulation Results

In this section we have carried out a computer experiment to illustrate the performance of Algorithm 6.1: PM.Pr.Gradient. We have considered a multiple stream MIMO BC with $K = 2$ users, $R = 8$ antennas per user, and $N = 8$ transmit antennas. The number of streams allocated per user is $d_1 = d_2 = 4$. The AWGN is zero mean with $\boldsymbol{C}_{\boldsymbol{\eta}} = \mathbf{I}_R$, and the per-user target rates are set to $\rho_1 = 14$ and $\rho_2 = 11$ bits per channel use. The initial step size is $s_0 = 1$ and the stop threshold is fixed to $\delta = 10^{-4}$. The partial CSIT v contains information regarding the channel first and second order moments according to $[\mathrm{E}[\boldsymbol{H}_k|v]]_{1:N,r} = \boldsymbol{u}_{k,r}$, for each $r \in \{1, \ldots, R\}$ with $u_{k,r,n} = \mathrm{e}^{j(n-1)\varphi_k}$ and $\varphi_k \sim \mathcal{U}(0, 2\pi)$, and $\boldsymbol{C}_{\boldsymbol{H}_k|v} = R\mathbf{I}_N$, $\forall k$, respectively. That is, a Vandermonde matrix for each of the users. We consider $M = 1000$ channel realizations generated according to the channel statistics above mentioned.

Figure 6.4 depicts the initial per-stream target rates for the execution example of Algorithm 6.1: PM.Pr.Gradient. Note that the initial values satisfy the per-user constraints. The evolution of such targets can be observed in Fig. 6.2, which shows the obtained results after executing the algorithm for the previous scenario. Observe that at every iteration the sum of all the per-stream targets for user k is kept constant and equal to ρ_k. The total power needed to achieve those targets is shown in Fig. 6.3. In such a figure, it can be seen how the power is gradually reduced by modifying the targets using the projected-gradient method until reaching the convergence at a local minimum. The target rates after 28 iterations are depicted in Fig. 6.5.

Although the projected-gradient convergence is slow, notice that it is reached for both the power and the per-streams targets.

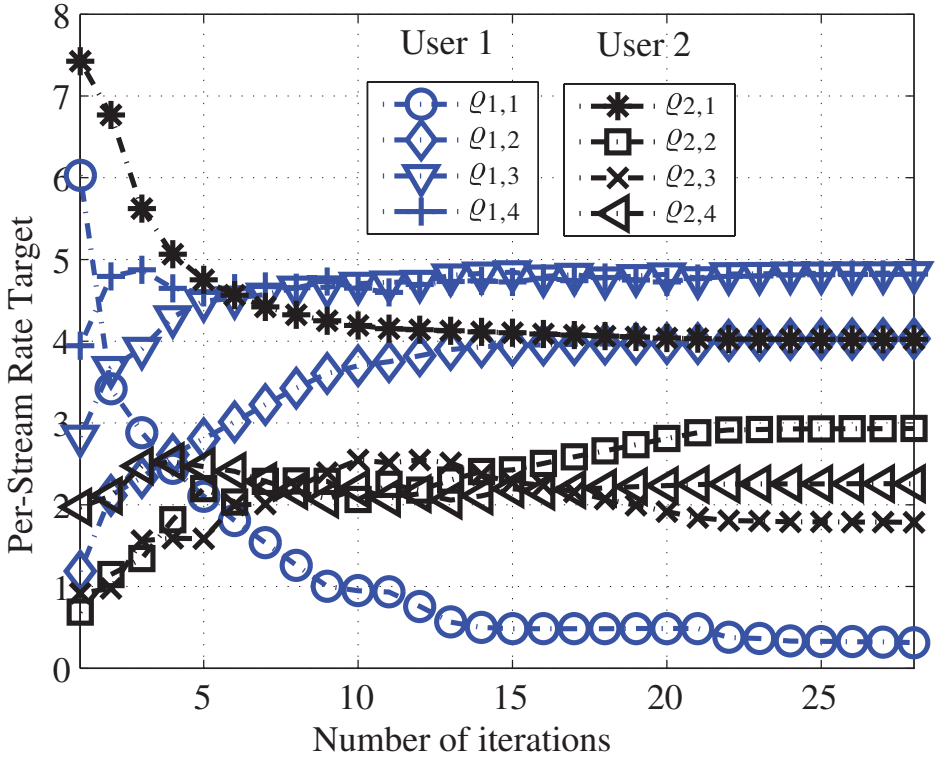


Figure 6.2: Per Stream Rate Targets vs. Number of Iterations.

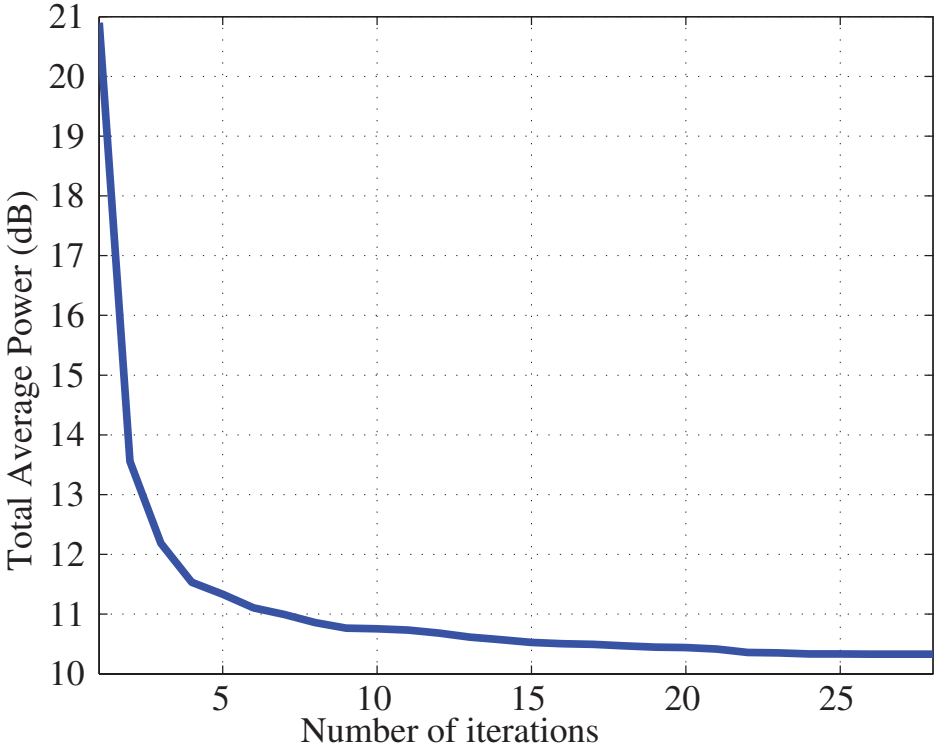


Figure 6.3: Total Transmit Power vs. Number of Iterations.

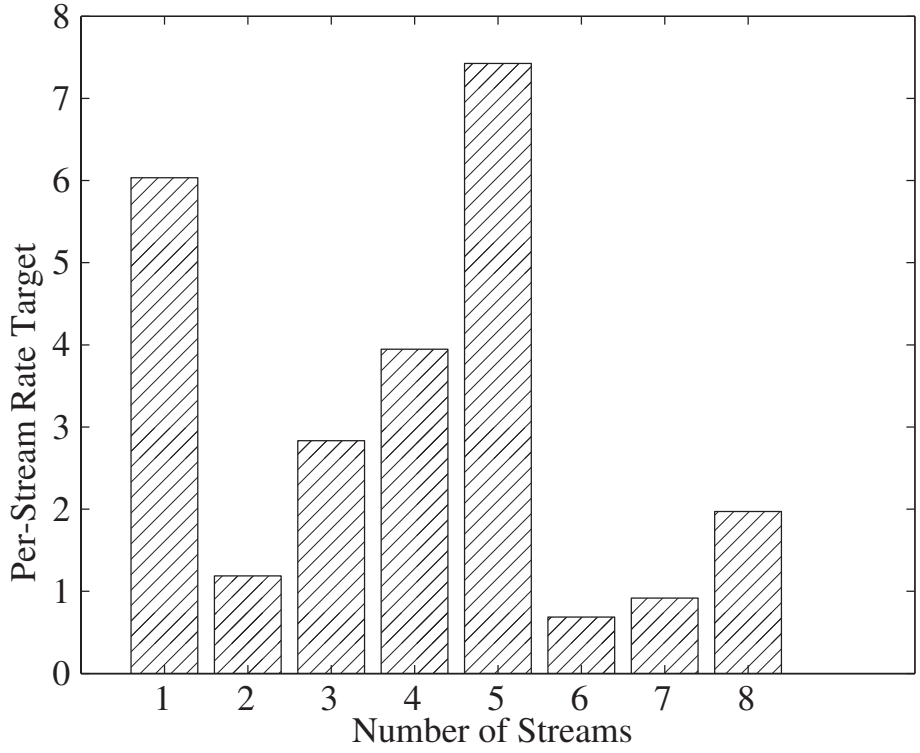


Figure 6.4: Initial Per Stream Average Rate Targets.

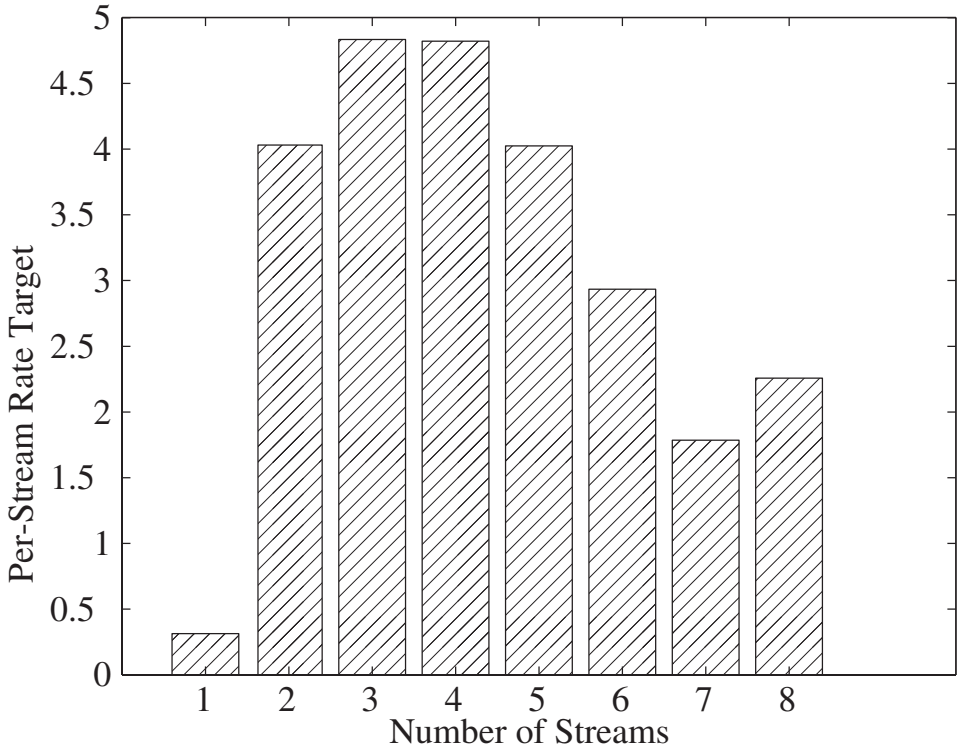


Figure 6.5: Per Streams Average Rate Targets after Convergence.

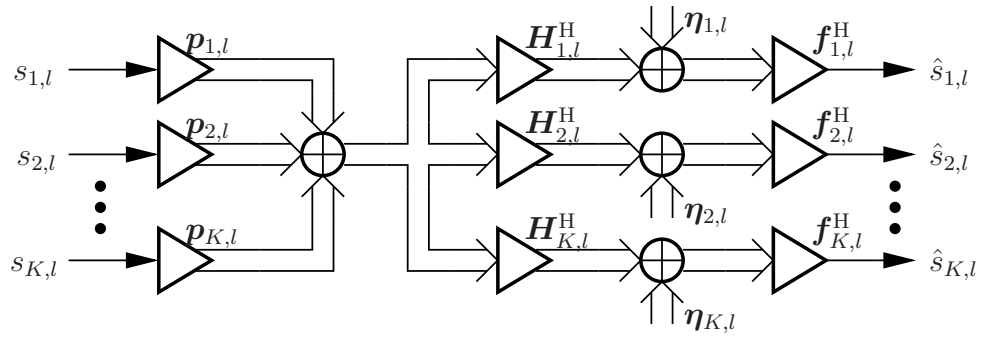


Figure 6.6: Per-Subcarrier Discrete-Time Equivalent Model of a MIMO-OFDM BC.

6.2. Power Minimization in the MIMO-OFDM BC

The optimization of the MIMO-OFDM BC according to some performance metric has been previously considered in the literature. For example, weighted sum-rate maximization algorithms have been proposed in [86, 87]. On the other hand, optimal precoders and receivers minimizing the sum-MSE have also been studied in [88, 89]. The power minimization with QoS constraints has been also considered in [86, 90, 91]. However, the case of imperfect CSIT has never been considered before in these works.

The imperfect CSI assumption is considered in some works on MIMO-OFDM. The design of linear precoders for single-user MIMO-OFDM assuming a certain channel estimation error model was studied in [92]. Also a different error model is used in [93], where data are transmitted only to the best user for each subcarrier.

In this section, we study the power minimization of the MIMO-OFDM BC. To do that, the methods proposed in Chapter 5 for the MIMO BC are adapted to such a scenario. We observe that the algorithmic solution obtained in Section 6.1 is appropriate also for OFDM-MIMO when certain particularities are taken into account [94].

In Section 2.20 of Chapter 2 we described the system model when the OFDM modulation is applied. As we have discussed, the *Inverse Discrete Fourier Transform* (IDFT) and the *Discrete Fourier Transform* (DFT) are performed at the transmitter and the receiver to modulate and demodulate the OFDM symbol, respectively. Moreover, a cyclic prefix is added to remove the *Inter-Symbol Interference* (ISI). Due to these facts, the equivalent channel matrix in the MIMO-OFDM can be obtained as a block diagonal matrix (cf. Section 2.3.2 of Chapter 2).

Accordingly, Fig. 6.6 plots the discrete-time equivalent model corresponding to the lth subcarrier of a MIMO-OFDM BC. We assume that a centralized transmitter sends the data signal $s_{k,l}$ to the user $k \in \{1, \ldots, K\}$ over the subcarrier $l \in \{1, \ldots, L\}$, where L is the number of subcarriers. Data signals corresponding to different users and/or subcarriers are mutually independent and zero-mean Gaussian normalized to unit power

(i.e. $\mathrm{E}[|s_{k,l}|^2] = 1$). We also assume that the transmitter is equipped with N antennas, while all the receivers are equipped with the same number R of receive antennas. Hence, each data signal $s_{k,l}$ is precoded with $\boldsymbol{p}_{k,l} \in \mathbb{C}^N$ and propagates over the MIMO channel $\boldsymbol{H}_{k,l} \in \mathbb{C}^{N\times R}$. The received signal is perturbed by the AWGN $\boldsymbol{\eta}_{k,l} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{\eta}_{k,l}})$ and filtered with the linear receiver $\boldsymbol{f}_{k,l} \in \mathbb{C}^R$.

Along this section we will assume perfect CSIR and imperfect CSIT modeled by v. Additionally, the conditional pdfs $f_{\boldsymbol{H}_{k,l}|v}(\boldsymbol{H}_{k,l}|v)$ associated to each v are known at the transmitter. That is, the assumption of the statistical error model considered in this work, i.e.

$$\boldsymbol{H}_{k,l} = \bar{\boldsymbol{H}}_{k,l} + \tilde{\boldsymbol{H}}_{k,l}, \tag{6.54}$$

with $\bar{\boldsymbol{H}}_{k,l} = \mathrm{E}[\boldsymbol{H}_{k,l}|\, v]$ and $\tilde{\boldsymbol{H}}_{k,l}$ being the imperfect CSI error, with $\tilde{\boldsymbol{H}}_{k,l} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\tilde{\boldsymbol{H}}_{k,l}})$. Moreover, we assume that the cyclic prefix of the OFDM modulation is large enough to avoid ISI.

Let us now collect the L symbols transmitted during one OFDM symbol to the user k into $\boldsymbol{s}_k = [s_{k,1}, \ldots, s_{k,N}]^{\mathrm{T}}$. We next define the following matrices to represent the MIMO-OFDM precoder, channel, receive filter, and channel noise, respectively, corresponding to the user k

$$\begin{aligned}
\boldsymbol{P}_k &= \mathrm{blockdiag}\left(\boldsymbol{p}_{k,1}, \ldots, \boldsymbol{p}_{k,L}\right), \\
\boldsymbol{H}_k &= \mathrm{blockdiag}\left(\boldsymbol{H}_{k,1}, \ldots, \boldsymbol{H}_{k,L}\right), \\
\boldsymbol{F}_k &= \mathrm{blockdiag}\left(\boldsymbol{f}_{k,1}, \ldots, \boldsymbol{f}_{k,L}\right), \\
\boldsymbol{\eta}_k &= \mathrm{blockdiag}\left(\boldsymbol{\eta}_{k,1}, \ldots, \boldsymbol{\eta}_{k,L}\right).
\end{aligned} \tag{6.55}$$

Accordingly, the output of the kth user receive filter (i.e. the kth user estimated symbols) is as follows

$$\hat{\boldsymbol{s}}_k = \boldsymbol{F}_k^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{P}_i \boldsymbol{s}_i + \boldsymbol{F}_k^{\mathrm{H}} \boldsymbol{\eta}_k. \tag{6.56}$$

Note that the former expression matches that obtained for the multiple stream MIMO scenario (6.1). However, for the OFDM system model all the matrices are block diagonal.

In this section, we address the joint optimization of the linear precoders $\{\boldsymbol{P}_k\}_{k=1}^K$ and equalizers $\{\boldsymbol{F}_k\}_{k=1}^K$ to minimize the sum transmit power $\sum_{i=1}^K \|\boldsymbol{P}_i\|_{\mathrm{F}}^2$ ensuring certain QoS restrictions for all the users. Such constraints are given as average rates and we can equivalently focus on the conditional average rates instead, as discussed in Section 6.1 (see also (3.16) of Chapter 3)

$$\mathrm{E}[R_k|\, v] = \mathrm{E}\left[\log_2 \det\left(\mathbf{I}_L + \boldsymbol{P}_k^{\mathrm{H}} \boldsymbol{H}_k \left(\boldsymbol{C}_{\boldsymbol{\eta}_k} + \boldsymbol{H}_k^{\mathrm{H}} \sum_{i\neq k} \boldsymbol{P}_i \boldsymbol{P}_i^{\mathrm{H}} \boldsymbol{H}_k\right)^{-1}\right) \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{P}_k |\, v\right]. \tag{6.57}$$

Observe that the matrix inside (6.57), denoted as $\boldsymbol{\Sigma}_k^{-1}$ in the following, is diagonal. Therefore, the k-th user rate can be expressed as the sum of the per-subcarrier rates

$$\mathrm{E}[R_k|\, v] = \sum_{l=1}^{L} \mathrm{E}[R_{k,l}|\, v] = \sum_{l=1}^{L} \mathrm{E}\left[\log_2 \det\left(\left[\boldsymbol{\Sigma}_k^{-1}\right]_{l,l}\right) |\, v\right]. \tag{6.58}$$

Hence, the per-user rate constraints can be satisfied by imposing a set of more stringent per-subcarrier rate constraints $\boldsymbol{\varrho}_k = [\varrho_{k,1}, \ldots, \varrho_{k,L}]^{\mathrm{T}}$ such that $\sum_{l=1}^{L} \varrho_{k,l} = \rho_k$. Note that the determinant of the diagonal matrix turns into $L-1$ scalar products.

Now, we avoid the complexity of the optimization with average rate constraints following similar steps to that from (6.1.1). Substituting $\boldsymbol{F}_k^{\mathrm{MMSE}}$ into the average MSE expression, and applying the matrix inversion lemma (see Section C.4 of Appendix C), we get

$$\overline{\mathrm{MMSE}}_k^{\mathrm{BC}} = \mathrm{E}\left[\mathrm{tr}\left(\mathbf{I}_L + \boldsymbol{P}_k^{\mathrm{H}}\boldsymbol{H}_k\left(\boldsymbol{C}_{\boldsymbol{\eta}_k} + \boldsymbol{H}_k^{\mathrm{H}}\sum_{i\neq k}\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}}\boldsymbol{H}_k\right)^{-1}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{P}_k\right)^{-1} |\, v\right], \tag{6.59}$$

which is easily related to the average rate using the matrices $\boldsymbol{\Sigma}_k$. Notice that the previous MMSE is computed as the trace of a diagonal matrix. Hence, the MMSE of the user k is computed as the sum of the MMSEs for each subcarrier, that is

$$\overline{\mathrm{MMSE}}_k^{\mathrm{BC}} = \mathrm{E}\left[\mathrm{tr}\left(\boldsymbol{\Sigma}_k\right) |\, v\right] = \mathrm{E}\left[\sum_{l=1}^{L}\left[\boldsymbol{\Sigma}\right]_{l,l} |\, v\right]. \tag{6.60}$$

Recall that in the multiple stream MIMO BC we force the matrix $\mathrm{E}[\boldsymbol{\Sigma}_k|\, v]$ to be diagonal by means of spatial decorrelation precoders (cf. Section 6.1.1). Nevertheless, for the MIMO-OFDM BC the matrices $\boldsymbol{\Sigma}_k$ are diagonal for all the users due to the properties of the OFDM modulation. Accordingly, we arrive at identical problem formulations using the per-subcarrier MMSE based constraints

$$\min_{\{\boldsymbol{P}_k, \boldsymbol{F}_k\}_{k=1}^{K}} \sum_{i=1}^{K} \|\boldsymbol{P}_i\|_{\mathrm{F}}^2 \;\text{ subject to }\; \overline{\mathrm{MMSE}}_{k,l}^{\mathrm{BC}} \leq 2^{-\varrho_{k,l}},\ \forall k, l. \tag{6.61}$$

The former problem is rather difficult since the per-subcarrier targets allowing to get lower values of the total transmit power are not known beforehand. Following similar steps to those proposed in Section 6.1.3, (6.61) is split up into a nested optimization problem. On the one hand, the per-subcarrier targets allowing to minimize the total transmit power have to be found as

$$\min_{\{\boldsymbol{\varrho}_k\}_{k=1}^{K}} P_T(\boldsymbol{\varrho}) \quad \text{subject to} \quad \mathbf{1}^{\mathrm{T}}\boldsymbol{\varrho}_k = \rho_k, \quad \text{and} \quad \boldsymbol{\varrho}_k \geq \mathbf{0},\ \forall k, \tag{6.62}$$

with $\boldsymbol{\varrho} = [\boldsymbol{\varrho}_1^{\mathrm{T}}, \ldots, \boldsymbol{\varrho}_K^{\mathrm{T}}]^{\mathrm{T}}$, and $\boldsymbol{\varrho}_k = [\varrho_{k,1}, \ldots, \varrho_{k,L}]^{\mathrm{T}}$. On the other hand, we address the minimization of the total transmit power for given per-subcarrier average target rates

$$P_T(\boldsymbol{\varrho}) = \min_{\{\boldsymbol{P}_k, \boldsymbol{F}_k\}_{k=1}^K} \sum_{k=1}^{K} \|\boldsymbol{P}_k\|_{\mathrm{F}}^2 \quad \text{subject to} \quad \overline{\mathrm{MMSE}}_{k,l}^{\mathrm{BC}} \leq 2^{-\varrho_{k,l}},\ \forall k, l. \tag{6.63}$$

The Algorithm 6.1: PM.Pr.Gradient proposed in Section 6.1.4 for the multiple stream MIMO is also suitable for the MIMO-OFDM BC. We should remark, however, some differences with respect to the derivative obtained in (6.34). Remember that we have distinguished two cases for the multiple stream MIMO system model, i.e. $(m = k, n = i)$ and $(m \neq k \,\text{or}\, n \neq i)$, corresponding to the expressions (6.35) and (6.36), respectively. For the MIMO-OFDM scenario, the derivative (6.35) holds for $(m = k, n = i)$. Nevertheless, we split the case $(m \neq k \,\text{or}\, n \neq i)$ into $(m \neq k, n = i)$, where (6.36) applies, and $n \neq i$, for every value of m and k, which results in zero due to the orthogonality between the different subcarriers.

Having in mind the previous consideration for the derivative, the per-subcarrier target rates are treated as the per-stream target rates in the Algorithm 6.1: PM.Pr.Gradient. In the following section we will evaluate the performance of such an algorithm for the scenario considered in this section.

6.2.1. Simulation Results

In this section we present the results of simulation experiments carried out to evaluate the performance of the proposed power minimization algorithm. Simulations consider the *Intelligent Multi-Element Transmit and Receive Antennas* (I-METRA) case D channel model for the point-to-point links in the BC [95]. Assuming proper cyclic insertion [96], the I-METRA MIMO-OFDM channel model is given by

$$\boldsymbol{H}_l = \sum_{t=1}^{T} \boldsymbol{R}^{1/2} \boldsymbol{H}(t) \boldsymbol{T}^{1/2} \exp\left(\frac{-j2\pi lt}{L}\right),\ l \in \{1, \ldots, L\}, \tag{6.64}$$

where $\boldsymbol{H}(t),\ t \in \{1, \ldots, T\}$ is a sequence of spatially uncorrelated time-domain $N \times R$ MIMO channel matrices, and $\boldsymbol{T}$ and $\boldsymbol{R}$ represent the transmit and receive spatial-correlation matrices, respectively. The I-METRA model assumes Rayleigh fading, i.e. the entries to $\boldsymbol{H}(t)$ are complex valued zero-mean circularly-symmetric Gaussian random variables. In I-METRA case D, the power delay profile is that of the *International Telecommunication Union* (ITU) Pedestrian B channel model, whereas the matrices $\boldsymbol{T}$ and $\boldsymbol{R}$ for $N = R = 4$ are specified (see [95] for further details). Finally, the noise covariance matrix is $\boldsymbol{C}_{\boldsymbol{\eta}_{k,l}} = \mathbf{I},\ \forall k, l$. We have considered that $K = 2$ independent MIMO-OFDM channels were generated using the above described channel model $\boldsymbol{H}_{k,l},\ \forall k, l$.

The CSI available at the transmitter is obtained by estimating the MIMO-OFDM channel responses, $\boldsymbol{H}_{k,l}$, i.e.

$$\hat{\boldsymbol{H}}_{k,l}^{(m)} = \boldsymbol{H}_{k,l} + \tilde{\boldsymbol{H}}_{k,l}^{(m)}, \tag{6.65}$$

where $\tilde{\boldsymbol{H}}_{k,l}^{(m)}$ is the channel estimation error and $\hat{\boldsymbol{H}}_{k,l}^{(m)}$ is the channel estimate for the realization m. We assume that $\tilde{\boldsymbol{H}}_{k,l}^{(m)} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, 0.1\mathbf{I})$ and $m = 1, \ldots, M$. The gradient initial step is set to $s_0 = 1$ and the threshold δ is fixed to 10^{-4}.

Figures 6.7 and 6.8 depict the evolution of the algorithm for a given BC channel realization $\boldsymbol{H}_{k,l}$, with $M = 500$, $K = 2$ users, and $L = 4$ subcarriers. The per-user target rates were set to $\rho_1 = 5 \times L$, and $\rho_2 = 3 \times L$. In particular, Fig. 6.8 shows how the total sum power diminishes with the number of iterations. Moreover, the evolution of the per-subcarrier target rates with the number of iterations is displayed in Fig. 6.7. Observe that both the per-subcarrier and the total transmit power converge to a locally optimum value. Note also that the sum of all the per-subcarrier target rates gives the per-user target rates ρ_1 and ρ_2 at all the iterations.

Figure 6.9 represents the per-subcarrier transmit power after convergence by averaging over 100 channel realizations of $\boldsymbol{H}_{k,l}$ and $M = 100$ estimates for each realization. Average transmit power per-subcarrier, P_T/L, is logarithmically expressed in dB for $N = 8$, 16, 32, and 64 subcarriers. The results presented in Fig. 6.9 are obtained considering the per-user target rates $\rho_1 = 5 \times L$, and $\rho_2 = 3 \times L$ bits per channel use. Note that the per-user target rates increase proportionally with the number of subcarriers in order to keep constant the system spectral efficiency.

It is apparent from Fig. 6.9 that the proposed power minimization algorithm produces similar results irrespective of the number of subcarriers. Since including new subcarriers does not increase the interference experienced by the other subcarriers, this indicates that the algorithm splits the rates among the subcarriers in a smart way independently of the problem size.

6.3. Feedback and Filter Design in the MISO BC

So far, we have considered the minimization of the total transmit power subject to average rate constraints. Thanks to the Bayes' rule, we have been able to show that we can equivalently focus on the average rates conditioned to the partial CSIT. In this section we address the challenge of designing such an imperfect CSIT. Our proposal is to employ the Lloyd's algorithm to jointly design the filters and the feedback [97].

In this section, we consider a MISO BC where the CSI of the users is jointly quantized into $L = 2^b$ regions. Here, b denotes the overall number of bits available to represent the CSI and L the number of quantizer levels. The CSI quantizer is defined by the

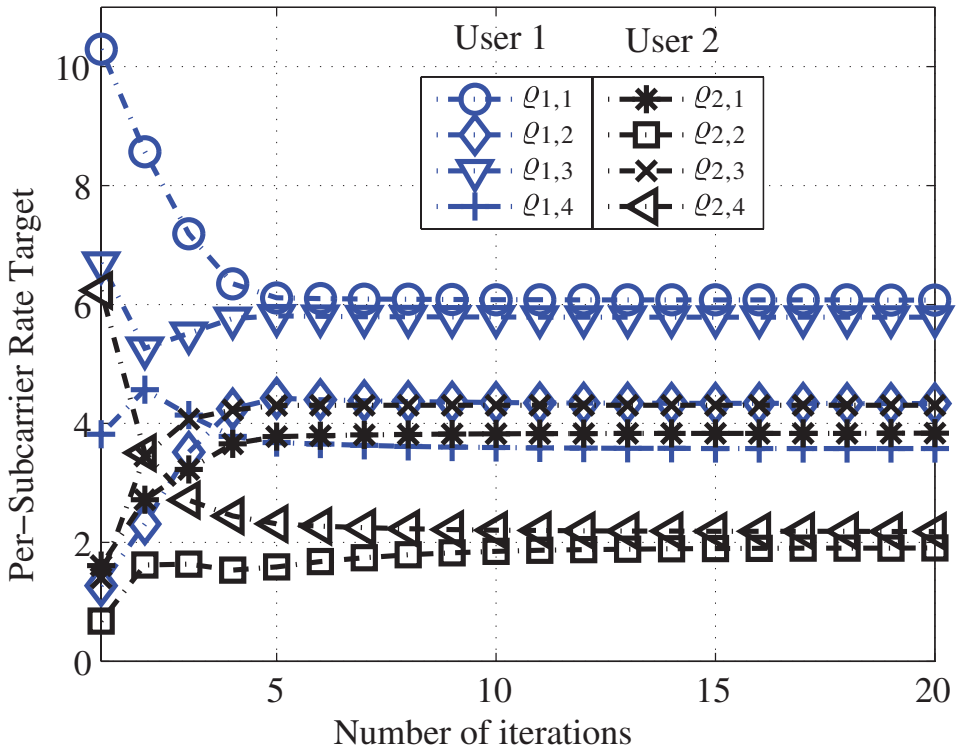


Figure 6.7: Execution Example of Algorithm 6.1 in a MIMO-OFDM Scenario: Per-Subcarrier Target Rate vs. Number of Iterations.

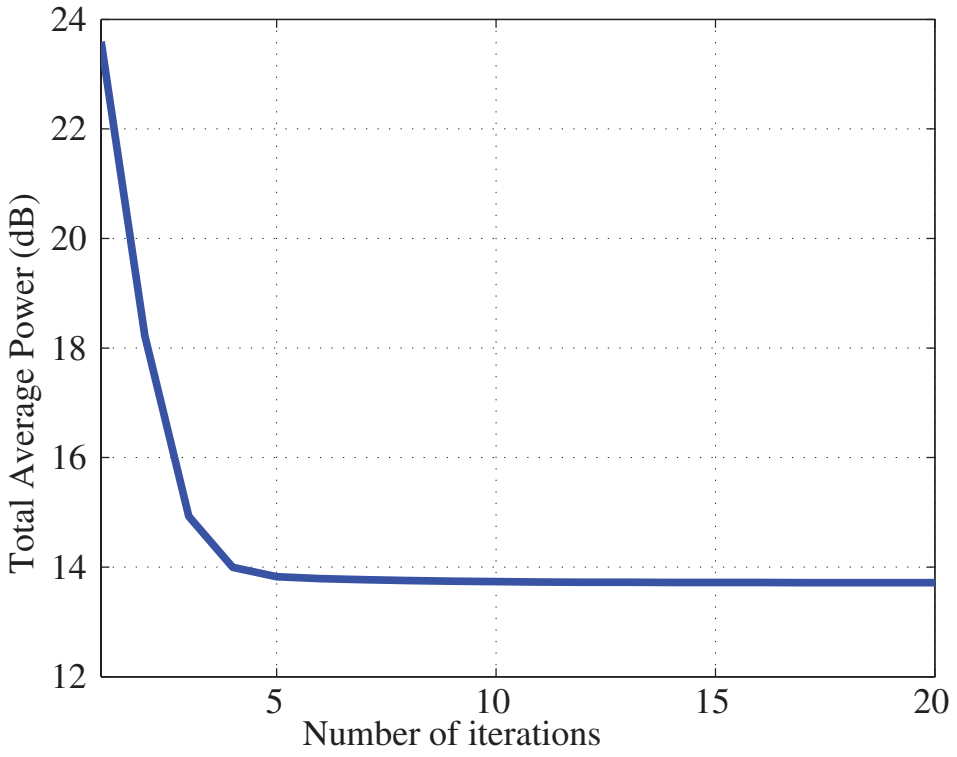


Figure 6.8: Execution Example of Algorithm 6.1 in a MIMO-OFDM Scenario: Total Per-Subcarrier Transmit Power vs. Number of Iterations.

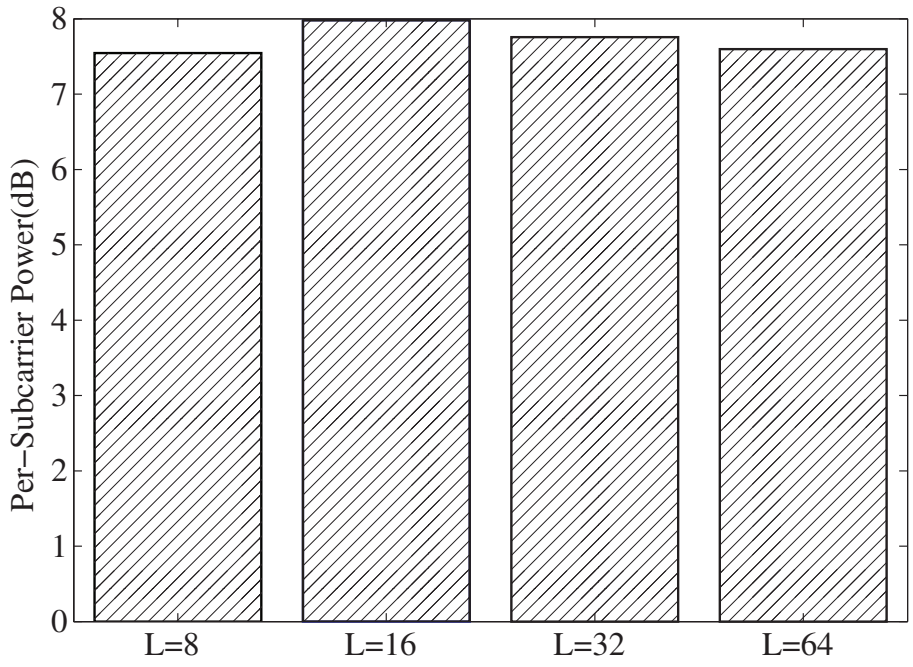


Figure 6.9: Resulting Average Per-Subcarrier Power. L is the Number of Subcarriers.

quantizer regions and the corresponding representatives, i.e., the precoders in our case. The addressed problem is the joint optimization of the linear precoders and the quantizer regions. The optimization criterion is the minimization of the transmit power under rate constraints.

Consequently, we develop a Lloyd's algorithm [75] to find both optimal quantizer and codebook by means of the methods proposed in Section 4.2 of Chapter 4. That is, the Algorithm 4.3: PM.MISO.ICSI.2 based on the duality between MAC and BC with respect to the average MSE, and the properties of *standard* interference functions.

Recall the MISO BC system model presented in Section 4.2, where the BS sends a Gaussian data signal $s_k \sim \mathcal{N}_{\mathbb{C}}(0,1)$, with $k \in \{1,\ldots,K\}$, to the K single-antenna users. Each of the mutually independent data signals s_k, i.e. $\mathrm{E}[s_k s_j] = 0$ for $k \neq j$, is linearly precoded by $\boldsymbol{p}_k \in \mathbb{C}^N$. Then, it propagates over the vector channel $\boldsymbol{h}_k \in \mathbb{C}^N$, and is perturbed by the AWGN $\eta_k \sim \mathcal{N}_{\mathbb{C}}(0,\sigma_{\eta_k}^2)$ which is independent of s_k. Filtering with the scalar receive filter f_k leads to the data signal estimate

$$\hat{s}_k = f_k \boldsymbol{h}_k^{\mathrm{H}} \sum_{i=1}^{K} \boldsymbol{p}_i s_i + f_k \eta_k. \tag{6.66}$$

Recall that perfect CSIR is assumed but the CSIT is only partial. In particular, we assume that the receivers cooperate to decide what information is fed back to the transmitter. That way, the CSIT is represented by certain quantizer region $\mathcal{R}_i$.

With the matrix $\boldsymbol{H}$ collecting the channel vectors for all the users, i.e.

$$\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K], \tag{6.67}$$

the quantizer can be written as

$$\mathrm{Q}(\boldsymbol{H}) = \sum_{i=1}^{L} \boldsymbol{P}_i \, \mathrm{S}_i(\boldsymbol{H}), \tag{6.68}$$

where $L = 2^b$ is the number of quantizer regions, $\mathcal{R}_i$, $i \in \{1,\ldots,L\}$, and b is the number of bits fed back to the transmitter. Likewise, we define the total precoder

$$\boldsymbol{P}_i = [\boldsymbol{p}_{i,1}, \ldots, \boldsymbol{p}_{i,K}], \tag{6.69}$$

as the ith codebook entry which contains the precoders for all the users. Additionally, let us introduce the set of selection functions $\mathrm{S}_i(\boldsymbol{H})$, with $i \in \{1,\ldots,L\}$, defined as follows

$$\mathrm{S}_i(\boldsymbol{H}) = \begin{cases} 1 & \text{if } \boldsymbol{H} \in \mathcal{R}_i, \\ 0 & \text{otherwise.} \end{cases} \tag{6.70}$$

Note that (6.68) implies that the channels $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K$ are jointly quantized because no structural properties for $\mathcal{R}_i$ are assumed.

Let us introduce the distortion measurement d. Such a metric indicates the quality of the quantizer and the codebook, and it is defined as the mean of the average power required to fulfill the rate constraints for each quantizer region $\mathcal{R}_i$, i.e.

$$d = \sum_{i=1}^{L} P_T\left(\mathcal{R}_i\right) p_i, \tag{6.71}$$

where i is the index of the region where the channel $\boldsymbol{H}$ lies with a probability $p_i = \Pr\{\boldsymbol{H} \in \mathcal{R}_i\} = \mathrm{E}[\mathrm{S}_i(\boldsymbol{H})]$, and

$$P_T(\mathcal{R}_i) = \sum_{k=1}^{K} \mathrm{E}\left[\|\boldsymbol{p}_{i,k} s_k\|_2^2\right] = \|\boldsymbol{P}_i\|_{\mathrm{F}}^2. \tag{6.72}$$

Note that $\boldsymbol{P}_i$ is the codebook entry corresponding to the channel realizations belonging to the ith region, i.e. $\boldsymbol{H} \in \mathcal{R}_i$ (cf. (6.68)), that is, the set of precoders to be employed when the CSIT is i. In the following, the problem of jointly designing the L regions,

Bits	Region	Codebook
00	$\mathcal{R}_1$	$\boldsymbol{P}_1$
01	$\mathcal{R}_2$	$\boldsymbol{P}_2$
10	$\mathcal{R}_3$	$\boldsymbol{P}_3$
11	$\mathcal{R}_4$	$\boldsymbol{P}_4$

Table 6.2: Example of Partition Cells and Codebook for $b = 2$ bits.

$\mathcal{R}_i$, $\forall i \in \{1, \ldots, L\}$, and the corresponding precoders $\boldsymbol{P}_i$ for each region is addressed. To this end, the Lloyd's algorithm [75] is used to alternatively optimize the partition cells and the corresponding centroids using the distance measurement induced by the distortion metric (e.g., [98]). Every iteration of the AO can be split up into two steps, viz., the computation of $\boldsymbol{P}_i$ for every region $\mathcal{R}_i$ (centroid condition) and the update of all the regions $\mathcal{R}_i$ (nearest neighbor condition). These two steps are repeated until convergence is reached when that distance measurement falls below a preset threshold.

6.3.1. Centroid Condition

The centroid condition consists on finding the representatives of each region. In our particular case, such representatives are not elements of the region but a set of precoders that are suitable for all of them. Therefore, the entries of the codebook are the above mentioned linear precoders.

For the precoder design, we consider the criterion assumed throughout this work, that is, for given partition cells, minimizing the total transmit power, while ensuring minimum average rates for all the users ρ_k, $\forall k$. The ergodic rates come from the partial CSIT i, the index of the region which is fed back from the users. That is

$$\mathrm{E}[R_k|\,\boldsymbol{H}\in\mathcal{R}_i]=\mathrm{E}\left[\log_2\left(1+\left|\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{p}_{i,k}\right|^2\left(\sum_{j\neq k}\left|\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{p}_{i,j}\right|^2+\sigma_{\eta_k}^2\right)^{-1}\right)|\,\boldsymbol{H}\in\mathcal{R}_i\right]. \tag{6.73}$$

Observe that we can indistinctly use the conditioned expectation due to the Bayes' rule, as shown in Section 3.1.1 of Chapter 3. Moreover, (6.73) is obtained from (3.13) when we consider the MISO system model previously presented. Therefore, the optimization problem for the centroid condition is given as

$$\boldsymbol{P}_i=\mathrm{argmin}_{\boldsymbol{P}}\,\|\boldsymbol{P}\|_{\mathrm{F}}^2\quad\text{subject to}\quad\mathrm{E}\left[R_k|\,\boldsymbol{H}\in\mathcal{R}_i\right]\geq\rho_k,\,\forall k. \tag{6.74}$$

However, the optimization of the average rates is difficult. Hence, thanks to the concavity of the logarithm function in (6.73), and by using the Jensen's inequality, we find an approximation to (6.74). Let us first introduce the average MMSE, i.e., the result of plugging the optimal receive filters, f_k^{MMSE}, into the average MSE expression. Recall that we consider perfect CSIR. Hence, the average MMSE reads as

$$\overline{\mathrm{MMSE}}_k^{\mathrm{BC}}=\mathrm{E}\left[1-\boldsymbol{p}_{i,k}^{\mathrm{H}}\boldsymbol{h}_k\left(\boldsymbol{h}_k^{\mathrm{H}}\sum_{j=1}^{K}\boldsymbol{p}_{i,j}\boldsymbol{p}_{i,j}^{\mathrm{H}}\boldsymbol{h}_k+\sigma_{\eta_k}^2\right)^{-1}\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{p}_{i,k}|\,\boldsymbol{H}\in\mathcal{R}_i\right]. \tag{6.75}$$

Further details regarding the MSE and the optimal filters f_k^{MMSE} can be found in Section 4.2.1 of Chapter 4.

Instead of the average rate constraints of (6.74), we now propose to use conservative average MMSE based restrictions (see Section 4.2.1). In other words, when ensuring an average MMSE, a minimum average rate is guaranteed, i.e. $\mathrm{E}[R_k|\,\boldsymbol{H}\in\mathcal{R}_i]\geq\rho_k=-\log_2\left(\overline{\mathrm{MMSE}}_k^{\mathrm{BC}}\right)$ follows from $\overline{\mathrm{MMSE}}_k^{\mathrm{BC}}\leq 2^{-\rho_k}$. With these conservative bounds, we minimize the total transmit power under QoS constraints expressed as the maximum average MMSEs. Therefore, the problem formulation of (6.74) is rewritten as follows

$$\boldsymbol{P}_i=\mathrm{argmin}_{\boldsymbol{P}}\,\|\boldsymbol{P}\|_{\mathrm{F}}^2\quad\text{subject to}\quad\min_{\{\boldsymbol{g}_k,f_k\}_{k=1}^K}\overline{\mathrm{MMSE}}_k^{\mathrm{BC}}\leq 2^{-\rho_k},\,\forall k. \tag{6.76}$$

Note that the previous problem is similar to that of Section 4.2 of Chapter 4, (4.44). Then, it is possible to solve the optimization (6.76) with the methods already explained in Section 4.2. In particular, we propose to use again the algorithm referred to as Algorithm 4.3: PM.MISO.ICSI.2.

Although the details of the solution to (6.76) shall not be repeated here, we want to highlight some considerations from Section 4.2.3. Notice that we have to calculate the expectation in (6.75). Therefore, we propose to perform Monte Carlo numerical integration to approximate such expectation as in Section 4.2.3. This approximation is suitable when the channel ergodicity holds and the number of realizations M is large enough. To generate the channel realizations, the conditional pdfs $f_{\boldsymbol{h}_k|\boldsymbol{H}\in\mathcal{R}_i}(\boldsymbol{h}_k|\boldsymbol{H}\in\mathcal{R}_i)$ available at the transmitter are used for all the regions $\mathcal{R}_i$. Accordingly, every possible channel state $\boldsymbol{H}$ is contained in one of the quantizer regions

$$\boldsymbol{H}\in\bigcup_{i=1}^{L}\mathcal{R}_i,\ \forall\boldsymbol{H}, \tag{6.77}$$

or equivalently

$$\sum_{i=1}^{L}p_i=\sum_{i=1}^{L}\Pr\left\{\boldsymbol{H}\in\mathcal{R}_i\right\}=1. \tag{6.78}$$

This assumption allows to numerically compute the expectations of (6.75). Additionally, it is important to establish the nearest neighbor condition, as will made in the following section.

6.3.2. Nearest Neighbor Condition

After considering the design of the codebook entries $\boldsymbol{P}_i$, we now focus on the update of the quantizer regions $\mathcal{R}_i$ (see (6.68)). For given precoders $\boldsymbol{P}_i$, the nearest neighbor condition reallocates the M Monte Carlo channel realizations to the best fitting region. That is to say, the precoder $\boldsymbol{P}_i$ is selected to minimize the maximum ratio between the instantaneous and the target MMSE for all the users. Hence, the joint quantization of $\boldsymbol{H}^{(m)}=[\boldsymbol{h}_1^{(m)},\ldots,\boldsymbol{h}_K^{(m)}]$ reads as

$$i_{\text{nearest-neighbor}}(\boldsymbol{H}^{(m)})=\min_{\{i\}_1^L}\max_{\{k\}_1^K}2^{\rho_k}\text{MMSE}_k^{\text{BC}}\left(\boldsymbol{P}_i,\boldsymbol{H}^{(m)}\right), \tag{6.79}$$

where $\text{MMSE}_k^{\text{BC}}(\boldsymbol{P}_i,\boldsymbol{H}^{(m)})$ is the instantaneous MMSE for the channel realization $\boldsymbol{H}^{(m)}$ using the precoders collected in $\boldsymbol{P}_i$, i.e.

$$\text{MMSE}_k^{\text{BC}}(\boldsymbol{P}_i,\boldsymbol{H}^{(m)})=1-\boldsymbol{p}_{i,k}^{\text{H}}\boldsymbol{h}_k^{(m)}\left(\boldsymbol{h}_k^{(m),\text{H}}\sum_{j=1}^{K}\boldsymbol{p}_{i,j}\boldsymbol{p}_{i,j}^{\text{H}}\boldsymbol{h}_k^{(m)}+\sigma_{\eta_k}^2\right)^{-1}\boldsymbol{h}_k^{(m),\text{H}}\boldsymbol{p}_{i,k}. \tag{6.80}$$

Based on (6.79), each Monte Carlo channel realization is assigned to its new region $\mathcal{R}_{i_{\text{nearest-neighbor}}}$. This way, the second step of the iteration of the Lloyd's algorithm is established.

6.3.3. Algorithmic Solution

Algorithm 6.3: PM.Lloyds. Power Minimization: Lloyd's Algorithm

1: $\ell \leftarrow 0$, initialize: $d^{(0)} = \infty$, $\mathcal{R}_i$, $\forall i$
2: **repeat**
3: $\quad \ell \leftarrow \ell + 1$
4: $\quad$ **for** $l = 1$ to L **do**
5: $\quad\quad \boldsymbol{P}_l^{(\ell)} \leftarrow$ find optimum precoders with Algorithm 4.3 (See Section 6.3.1)
6: $\quad$ **end for**
7: $\quad$ **for** $m = 1$ to M **do**
8: $\quad\quad$ find $i_{\text{nearest-neighbor}}(\boldsymbol{H}^{(m)})$ via (6.79) using $\boldsymbol{P}_l^{(\ell)}, \forall l$ (See Section 6.3.2)
9: $\quad\quad$ move $\boldsymbol{H}^{(m)}$ to region $\mathcal{R}_{i_{\text{nearest-neighbor}}}$
10: $\quad$ **end for**
11: $\quad d^{(\ell)} \leftarrow$ compute distortion with (6.71)
12: **until** $|d^{(\ell)} - d^{(\ell-1)}| \leq \delta$

The algorithm referred to as Algorithm 6.3: PM.Lloyds summarizes the discussion in previous sections. In the proposed Lloyd's algorithm, the line 1 initializes the distortion measurement and the regions. That is, the channel realizations are distributed between the regions following some criterion, e.g. an uniform distribution.

On the one hand, the lines 4-6 implement the centroid condition (see Section 6.3.1) using the Algorithm 4.3: PM.MISO.ICSI.2. On the other hand, the lines 8-11 perform the nearest neighbor condition (see Section 6.3.2), where M denotes the total number of Monte Carlo channel realizations. The performance of the quantizer and the codebook is evaluated based on the distortion measurement (6.71), i.e., the average transmit power (see the line 11). Note that the considered QoS problem of (6.76) only has a solution if the QoS constraints are feasible which can be tested with the method presented in Section 5.2 of Chapter 5. To check the algorithm convergence, a difference measurement based on the distortion of consecutive iterations is compared with a threshold, δ, in the line 12.

6.3.4. Simulation Results

We carried out some computer simulations to evaluate Algorithm 6.3: PM.Lloyds. We have considered a MISO BC with $K = 3$ single antenna users and $N = 3$ transmit antennas. The target rates are set to $\rho_1 = 0.62$, $\rho_2 = 0.42$, and $\rho_3 = 0.52$ bits per channel use, respectively. The noise is additive white and Gaussian, with $\sigma_{\eta_k}^2 = 1$, $\forall k$, and the stop threshold is set to $\delta = 10^{-3}$. We generated 2000 realizations of a Rayleigh fading channel with zero mean and covariance matrix $\boldsymbol{C}_{\boldsymbol{h}_k} = \mathbf{I}_N$, $\forall k$, i.e. $\boldsymbol{h}_k^{(m)} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{h}_k})$ for

$m = \{1, \ldots, M\}$. For the initialization of $\mathcal{R}_i$, we have distributed the channel realizations uniformly among the different regions.

Figure 6.10 shows an example of the evolution of the distortion measurement (6.71) throughout the execution of Algorithm 6.3: PM.Lloyds for the proposed scenario. Observe that the distortion quickly decreases for the first iterations, whereas the improvement is marginal in the latter ones. This is in accordance to the initialization performed where any criterion for the assignment of the channels to a certain region is applied.

Figure 6.11 depicts a comparative of the total average power after convergence. We consider scenarios without feedback (0 bits), and scenarios including quantized feedback of $b = 1$, 2, 3, and 4 bits. Recall that, for all cases, the QoS constraints given by the average target rates ρ_k are fulfilled . As it can be appreciated from the figure, the quantized feedback dramatically reduces the necessary total average power. Additionally, the amount of power employed decreases with higher feedback bits, as expected.

6.4. Conclusions

In this chapter the algorithmic solution presented in prior chapters was applied to sole some additional issues to the power minimization in the BCs. Some difficulties arose from the more sophisticated system models, but were circumvented allowing to adapt the previously proposed methods to the new scenarios.

We first considered the total transmit power minimization in the MIMO BC with multiple streams to be allocated for each user. QoS constraints expressed as per-user average rates, due to the imperfect CSIT, have to be fulfilled. With spatial decorrelation precoders the matrices included in the average rate and MMSE expressions were diagonalized. Taking advantage of this fact, and thanks to the Jensen's inequality, it was possible to split up the per-user average target rates into more stringent per-stream average MMSE constraints. That way, we defined a nested optimization problem. On the one hand, the inner problem was solved via the algorithmic solution proposed in the previous chapter, considering each of the streams as virtual single stream MIMO users. On the other hand, the per-stream MMSE targets minimizing the total transmit power were studied. Our proposal used a projected-gradient algorithm with two basic steps. The first one consisted on updating the per-stream average MMSE targets in the gradient direction. The second one, denoted as projection, found the per-stream targets closest to the updated ones satisfying the original per-user rate constraints minimizing the Euclidean distance. The convergence to a local minimum was guaranteed when the initial per-user constraints were feasible. Some simulation experiments carried out to exhibit the good performance of the algorithmic solution were also presented in this chapter.

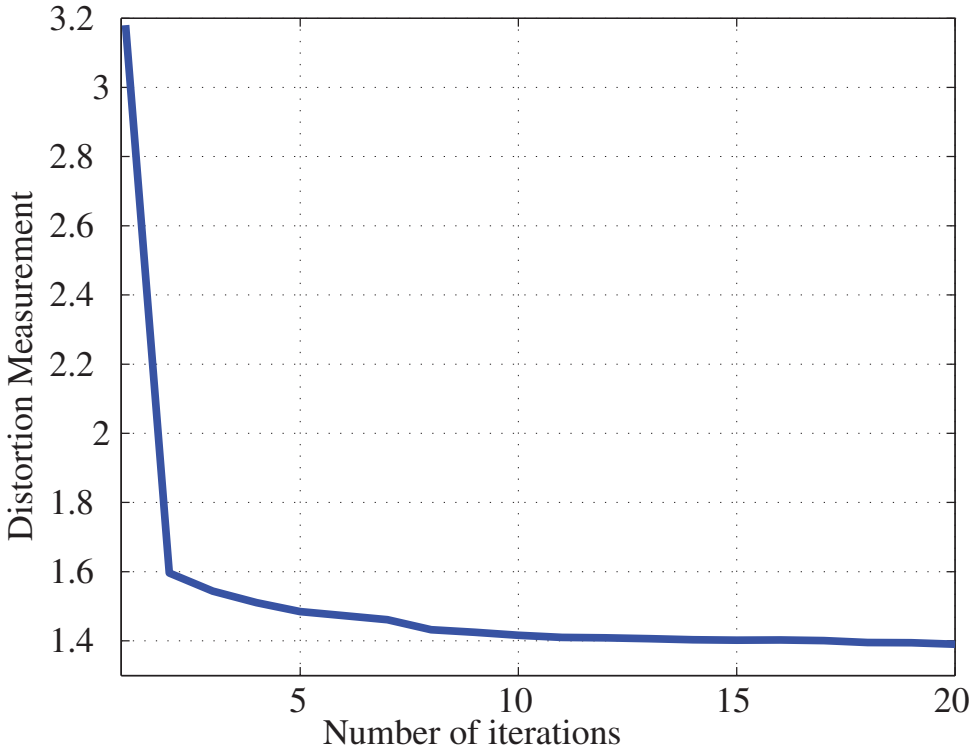


Figure 6.10: Example of Convergence of Algorithm 6.3: PM.Lloyds for a Quantizer of L=4 Levels: Distortion vs. Number of Iterations.

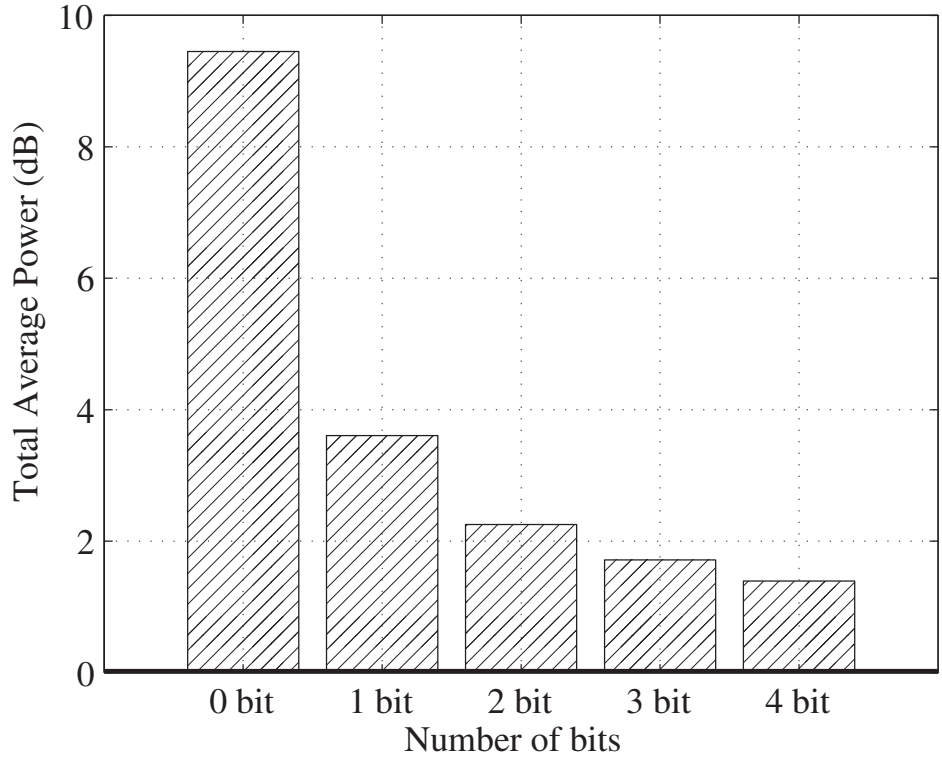


Figure 6.11: Resulting Total Average Transmit Power for Quantizers of L = 0, 2, 4, 8, and 16 Levels.

The OFDM MIMO BC has been also considered in this chapter. Our goal was to minimize the total transmit power subject to per-user average rate constraints. Again, imperfect CSIT was considered in the system model. Also, the assumption of proper length cyclic prefix was made. Therefore, thanks to the properties of the OFDM modulation, we have been able to show that the equivalent channel model results in a block diagonal matrix. Correspondingly, the signal model was rewritten so that both the transmit and the receive filters, and also the noise, were diagonal. Such a system model leaded us to diagonal matrices for the expressions of MMSE and rate performance metrics, allowing us to employ the same solution as for the multiple stream MIMO scenario. In the computer experiments we have considered realistic channel models, in particular, the I-METRA case D with the ITU Pedestrian B power delay profile.

The design of the information fed back from the users to the BS in the MISO BC has been also considered. In particular, we have proposed to jointly optimize the precoders and the available partial CSIT. That way, the possible states of the channel are quantized into a scalar depending on a partition cell. Such information is used in the BS as the entries of the codebook that contain the optimal precoders for every quantization region. To perform the joint optimization we have used the Lloyd's algorithm. Such algorithm consists on two basic steps. The first one addresses the optimization of the codebook for given partition cells, whereas the second one establishes the regions depending on the current codebook. In such a way, the algorithm finally converges to a solution. We have presented a numerical comparison of the total transmit power needed depending on the number of bits that are fed back to the transmitter using the realistic channel model aforementioned. It was shown that better the information fed back to the transmitter is, the larger the system performance improvement is.

Chapter 7

Conclusions and Future Work

In this chapter we present the concluding remarks of this dissertation and some future work lines.

7.1. Conclusions

This work focused on the design of linear precoders and receivers to minimize the transmit power in the *Multiple-Input Multiple-Output* (MIMO) *Broadcast Channel* (BC) fulfilling a set of per-user *Quality-of-Service* (QoS) constraints expressed in terms of per-user average rate requirements. Taking advantage of the Jensen's inequality, we have explained that the QoS constraints can be substituted by more manageable average *Minimum Mean Square Error* (MMSE) restrictions. We have next exploited the *Mean Square Error* (MSE) BC/*Multiple Access Channel* (MAC) duality to jointly determine the optimum transmit and receive filters by means of an *Alternate Optimization* (AO) algorithm. Such duality has been shown to be appropriate for both perfect and imperfect *Channel State Information at the Transmitter* (CSIT) in the BC. Additionally, the optimum transmit power allocation was found using the so-called *standard Interference Function* (IF) framework. We proposed an algorithmic solution that provides the optimal filters and power allocation if the problem constraints are feasible. Two possible implementations of the proposed algorithm are evaluated. The first one is discarded because convergence problems arise for some iterations of the algorithm even when the QoS constraints are feasible. However, with the second proposed implementation, the convergence to the minimum total transmit power is guaranteed.

Contrarily to certain works in the literature, where assumptions about the bounded uncertainty models such as rectangular, ellipsoidal, or spherical are made, we have only considered a statistical error model. Moreover, we do not approximate the ergodic rate, which means that the true rates are not known in advance. This second approach has been

commonly used, as discussed previously. Therefore, our system model is suitable for a large number of scenarios. Furthermore, we have carried out computer experiments with the purpose of comparing the results obtained with the Jensen's inequality based on the lower bound, with those resulting from the ergodic rate approximation above mentioned. The comparison demonstrates that using the lower bound the original QoS constraints are fulfilled. On the contrary, the QoS restrictions can be violated when the approximation is used. Moreover, such experiments also show that the gap between the lower bound and the true average rate is small.

We have also analyzed the problem feasibility to ensure convergence of the proposed power minimization algorithm. To do that, we have extended the existing studies of the sum-MSE feasibility region, where only *Multiple-Input Single-Output* (MISO) BC considering perfect CSIT and *Channel State Information at the Receiver* (CSIR) was analyzed. As a result, we obtained an expression dependent on the transmit filters. Moreover, for the particular case of Rayleigh channel realizations, we have observed a particular behavior of the feasibility region that was shown from computer experiments also presented here. Since the transmit filters are not known beforehand, a gradient step algorithm has been developed to find the optimal sum-MMSE filters and establish a lower bound. By using such algorithm, the impact of the partial CSIT quality has been evaluated, showing that it has a major influence on the feasibility of QoS restrictions.

The feasibility of the power minimization problem subject to QoS constraints is usually not known in advance. For such reason, a different formulation where the goal is to provide the best possible service for a given power arises. This optimization problem is known as rate balancing. In such a formulation, the average rates were manipulated in a way that preserved the equilibrium between the rates corresponding to every user, while the total transmit power had to be limited to a given value. We have shown that this problem can also be addressed using the algorithm employed to solve the power minimization. However, contrarily to the power minimization problem, the rate balancing optimization is always feasible, since the system designer is able to relax the restrictions until the desired total power is used. This adaptation of the rate targets leaded us to the bisection search. That way, we proved that convergence to a solution holds true. Additionally, we carried out simulation experiments to show the good performance of the proposed algorithm and compare it to other existing methods in the literature.

Finally, we have applied the proposed methods considering additional issues such as more complex system models or feedback design.

More specifically, we have investigated the minimization of the transmit power in a multiple stream MIMO BC with imperfect CSIT to accomplish certain user rate constraints. We developed a gradient-projection iterative algorithm to determine the optimal distribution of each user target rate among the different per-stream target rates. This way the multiple stream MIMO BC is interpreted as a single-stream MIMO BC

with virtual users. Two nested optimization problems arise from this interpretation. In the inner one, the goal is to minimize the total transmit power for given QoS restrictions. This optimization problem has been already solved with the algorithm previously proposed. For the outer one, a projected-gradient was proposed. It consisted of performing an update in the gradient direction and, afterwards, the new targets were mapped to the closest ones laying in the feasibility region in the projection procedure. The convergence to a local minimum was discussed using the projected-gradient algorithm.

The projected-gradient algorithm was shown to be also appropriate to minimize the transmit power in a MIMO *Orthogonal Frequency Division Multiplexing* (OFDM) BC fulfilling a given set of target user rates. The algorithm distributed each user target rate between the different subcarriers and, jointly, the linear transmit and receive filters were calculated.

In order to provide a more complete approach of a practical system, the design of the feedback together with the linear transmit and receive filters was studied for the MISO BC. This joint design meant that the quantizer that divides the channel states into partition cells and the codebook have to be implemented. Additionally, the transmit and receive filters had to be found as in prior scenarios. For such purpose, we have used a version of the Lloyd's algorithm. Two basic steps were computed in such an algorithm. The centroid condition was shown to be the step which finds the centroids for each region, i.e., the precoders that become part of the codebook. To do that, the power minimization algorithm was employed. Finally, the second proposed step distributed the channel states to the best fitting region according to the current codebook.

7.2. Future Work

The advantages of employing multiple antennas at transmission and reception make MIMO technologies fundamental for the high speed wireless communications demanded nowadays. Accordingly, this translates to the introduction of MIMO into modern wireless communication standards, e.g. IEEE 802.11n, IEEE 802.11ac (*WiFi*) [44], 4G, *Third Generation Partnership Project* (3GPP) *Long-Term Evolution* (LTE), *Worldwide Interoperability for Microwave Access* (WiMAX), or *High-Speed Packet Access* (HSPA)+. Downlink system models and beamforming techniques are usual in recent standards, e.g. IEEE 802.11ac. Therefore, in current wireless communication systems a common scenario is to have a transmitter that sends data to several independent users simultaneously. Due to the benefits of considering the spatial dimension, such transmissions are possible. As we have previously mentioned, the design of the precoders is of major relevance in all this kind of systems since it makes possible to separate the data signals at the user end with low complexity terminals.

The cellular systems constitute another example where the system model considered in this work can be applied. Such systems have to deal with the problem of the interference which does not even exist in point-to-point communications. Moreover, interference is a factor which limits the throughput in wireless communication networks. A traditional solution to achieve better throughput is to increase the transmit power. However, in systems with interference that translates into stronger interference levels at reception. Although such interference can be mitigated using precoding techniques, the efficient use of the power has become a key feature in the system design. Therefore, the optimization problem studied throughout this work that minimizes the transmit power guaranteeing a certain QoS fits current practical necessities.

Additionally, perfect CSIT is a rather unrealistic assumption. Then, the study of models considering such uncertainty is fundamental in the design of future standards in the field of wireless communications.

In the ensuing sections we describe some future work lines to continue the research contained in this dissertation.

7.2.1. Jensen's Inequality Lower Bound

The approximation of the average rate by means of the average MMSE is a common tool. Due to the combination of the logarithm and the expectation in the average rate expression, the relationship with the average MMSE is established by using the Jensen's inequality. Therefore, the average MMSE becomes a lower bound for the average rate. The gap between the two measurements has been shown to be small in our simulation results. Although the analytical expression for the gap is provided in Appendix A for the *Single-Input Single-Output* (SISO) scenario, the extension to the general case is still unknown. Thus, a possible future work could address the characterization of the gap between the average rate and the average MMSE in a more general scenario.

7.2.2. Feasibility Region

The feasibility region has been characterized in Section 5.2 of Chapter 5. This study provides an insightful view of the optimization problem, allowing to determine if the desired QoS rates are reachable or not. In order to do that, we resort again to the relationship between the MMSE and the average rate. Thus, the average sum-MMSE feasibility region was found. Such a region has been shown to be a polytope where the bounds are reached when the transmit power is increased without restriction. Nevertheless, the expression obtained for the aforementioned bounds depends on the channel and the MAC precoders (or corresponding BC receivers). This fact constitutes an inconvenience since the MAC precoders are not known in advance. Therefore, we

propose to acquire a better comprehension of the bounds corresponding to the feasible region as a possible future work.

7.2.3. Interference Channel

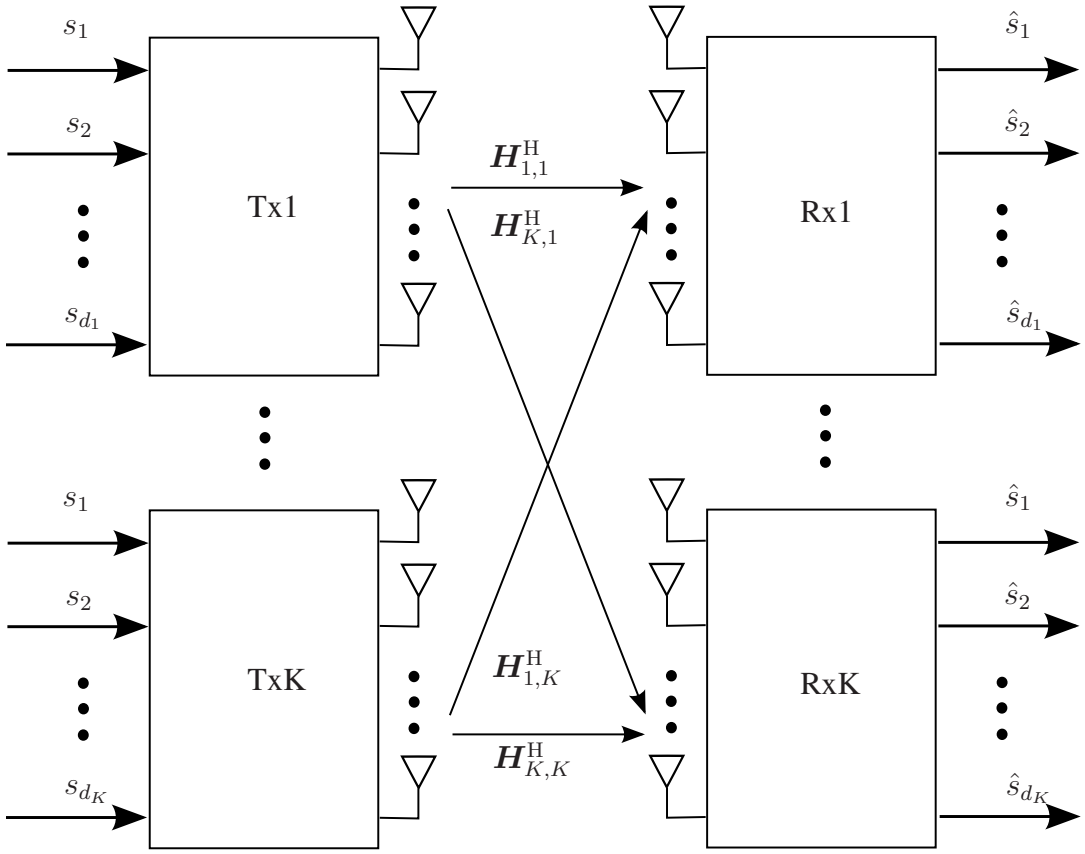


Figure 7.1: MIMO Interference Channel.

In cellular systems the designers have to deal with the problem of intercell interference, different from the intracell interference considered in this work. That behavior shows up when exist several transmitter-receiver pairs. Moreover, as we have demonstrated during this work, the interference has been identified as the major impairment limiting the throughput in wireless communication networks. Indeed, lots of recent standards include some sort of interference coordination to mitigate such a disadvantage. One example of an scenario where the intracell interference plays a role in the system performance comes from the MIMO *Interference Channel* (IFC). In a K user MIMO-IFC system model, there exist K transmitter-receiver pairs where each transmitter communicates to its respective receiver, as it can be seen in Fig. 7.1. Consequently,

each transmitter generates interference at all other receivers. Typical solutions for this scenario are given by the interference alignment signaling technique, where the channels are assumed to be perfectly known [99, 100].

Considering such communication scheme, a proposal for future work is to apply the methods studied in this work to this network.

7.2.4. Power Constraints

Different power restrictions can be applied in som scenarios depending on the network characteristics. For example, power restrictions for a set of users, for a certain user, for an antenna array or for a single antenna. In this work, the aim was minimizing the total transmit power. However, this could lead to situations where a large amount of power is allocated to one or several antennas. In practical systems, a limitation over the maximum power per antenna is a reasonable system restriction. Due to that, an interesting future work consists on studying the power minimization fulfilling QoS constraints in a system where there exists per antenna power limitations.

7.2.5. Non-Linear Precoding

In this work linear transmit and receive filters are used to mitigate the interferences produced in the BC when different data are transmitted to several users at the same time. The use of non-linear filters has been shown to improve the channel capacity, e.g. using *Dirty Paper Coding* (DPC) schemes in the BC. Moreover, the use of non-linear filters combined with the realistic assumption of imperfect CSIT has been considered in previous works, e.g. [81]. Therefore, a extension to the methods here proposed could be the use of non-linear precoders if we want to preserve the computational simplicity of the receive filters in the BC.

Appendix A

Jensen's Inequality

In this appendix, we briefly introduce the Jensen's inequality [36]. It is an important result employed in many fields, e.g. information theory, and also useful in our derivations.

Considering the *probability density function* (pdf) $f_{\boldsymbol{X}}$ and the convex function g (see Appendix F), the Jensen's inequality applied to probabilistic theory states

$$\begin{aligned} g\left(\int_{\mathbb{C}^{M\times N}} \boldsymbol{X} f_{\boldsymbol{X}}\left(\boldsymbol{X}\right) d\boldsymbol{X}\right) &\leq \int_{\mathbb{C}^{M\times N}} g\left(\boldsymbol{X}\right) f_{\boldsymbol{X}}\left(\boldsymbol{X}\right) d\boldsymbol{X}, \\ g\left(\mathrm{E}\left[\boldsymbol{X}\right]\right) &\leq \mathrm{E}\left[g\left(\boldsymbol{X}\right)\right], \end{aligned} \tag{A.1}$$

where g is a convex function. Note that if the function h is concave, i.e. $-h$ is convex, the Jensen's inequality can be applied changing the direction of the inequality as follows

$$h\left(\mathrm{E}\left[\boldsymbol{X}\right]\right) \geq \mathrm{E}\left[h\left(\boldsymbol{X}\right)\right]. \tag{A.2}$$

In this work, we are interested in determining the gap between the average rate and the average MMSE lower bound. Here, the SISO system model is studied. In the mentioned system, the transmit power is given by p, the noise variance is σ^2, and the Rayleigh distributed channel is $h \sim \mathcal{N}_{\mathbb{C}}(0,1)$. The MMSE, computed as $\mathrm{MMSE} = 1 - f^{\mathrm{MMSE}}hp$, reads as

$$\mathrm{MMSE} = 1 - \frac{|hp|^2}{|hp|^2+\sigma^2} = \frac{\sigma^2}{|hp|^2+\sigma^2}. \tag{A.3}$$

The conditional expectation of the MMSE is computed by

$$\mathrm{E}[\mathrm{MMSE}|\, v] = \mathrm{E}\left[\left.\frac{\sigma^2}{|hp|^2+\sigma^2}\right|\, v\right]. \tag{A.4}$$

To calculate such a gap, we consider a noise variance of $\sigma^2 = 10$ and a power $|p|^2 = 1$. In that way, the *Cumulative Distribution Function* (CDF) of the MMSE, $F_{h|v}(\mathrm{MMSE})$,

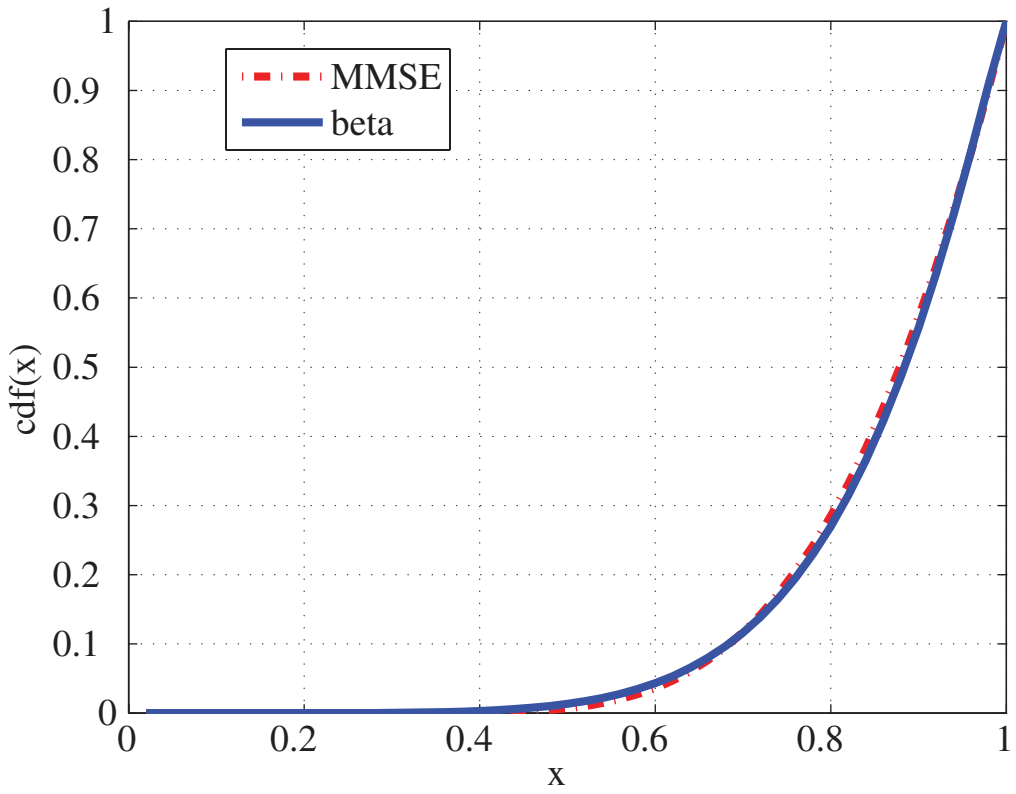


Figure A.1: MMSE Cumulative Distribution vs. Beta Cumulative Distribution.

can be approximated using the CDF of the beta distribution, with $\alpha = 6.54162$, and $\beta = 1.12133$. This approximation is accurate, as it can be observed in Fig. A.1.

We introduce the pdf $f_{h|v}(\text{MMSE})$, and the auxiliary variable $x = \text{MMSE}$ for notational brevity. Now, the expectation of the logarithm is computed as

$$\mathrm{E}[\ln(x)|\, v] = \int_0^1 f_{h|v}(x)\ln(x)dx. \tag{A.5}$$

Employing the approximated beta CDF, we get

$$\begin{aligned}
\mathrm{E}[\ln(x)|\, v] &= F_{h|v}(x)\ln(x)\big]_0^1 - \int_0^1 \frac{1}{x}F_{h|v}(x)dx && \text{(A.6)}\\
&= \cancelto{0}{F_{h|v}(x)\ln(x)\big]_0^1} - \int_0^1 \frac{1}{x}F_{h|v}(x)dx && \text{(A.7)}\\
&= -\sum_{j=\alpha}^{\alpha+\beta-1}\binom{\alpha+\beta-1}{j}\int_0^1 x^{j-1}(1-x)^{\alpha+\beta-j-1}dx && \text{(A.8)}\\
&= -\sum_{j=\alpha}^{\alpha+\beta-1}\binom{\alpha+\beta-1}{j}B(j,\alpha+\beta-j)\int_0^1 f_{h|v}(x)dx && \text{(A.9)}\\
&= -\sum_{j=\alpha}^{\alpha+\beta-1}\frac{(\alpha+\beta-1)!}{j!(\alpha+\beta-1-j)!}\frac{(j-1)!(\alpha+\beta-j-1)!}{(\alpha+\beta-1)!} && \text{(A.10)}\\
&= -\sum_{j=\alpha}^{\alpha+\beta-1}\frac{1}{j}. && \text{(A.11)}
\end{aligned}$$

Then, the average MMSE lower bound is the following

$$\begin{aligned}
-\mathrm{E}[\log_2(x)] &= \frac{1}{\ln(2)}\sum_{j=\alpha}^{\alpha+\beta-1}\frac{1}{j} = \frac{1}{\ln(2)}\left[H_{\alpha+\beta-1} - H_{\alpha-1}\right] && \text{(A.12)}\\
&\approx \frac{1}{\ln(2)}\left[\ln(\alpha+\beta-1) - \ln(\alpha-1)\right] = \frac{1}{\ln(2)}\ln\left(1+\frac{\beta}{\alpha-1}\right), && \text{(A.13)}
\end{aligned}$$

where H_n is the nth harmonic number.

Considering the expectation of the beta distribution, $\mathrm{E}[x|\, v] = \frac{\alpha}{\alpha+\beta}$, the average MMSE lower bound is $-\log_2(\mathrm{E}[x]) = \frac{1}{\ln(2)}\ln(1+\frac{\beta}{\alpha})$. Thus, the gap between the average

rate $\mathrm{E}[R|\, v]$ and the lower bound is as follows

$$\mathrm{E}\left[R|\, v\right] - \left[-\log_2\left(\mathrm{E}\left[\mathrm{MMSE}|\, v\right]\right)\right] = \tag{A.14}$$

$$= \frac{1}{\ln(2)} \sum_{j=\alpha}^{\alpha+\beta-1} \frac{1}{j} - \frac{1}{\ln(2)} \ln\left(1+\frac{\beta}{\alpha}\right) \tag{A.15}$$

$$\approx \frac{1}{\ln(2)} \left[\ln\left(1+\frac{\beta}{\alpha-1}\right) - \ln\left(1+\frac{\beta}{\alpha}\right)\right] \tag{A.16}$$

$$= \frac{1}{\ln(2)} \ln\left(\frac{\alpha}{\alpha-1}\left(1-\frac{1}{\alpha+\beta}\right)\right) \tag{A.17}$$

$$\approx \log_2\left(\left(1+\frac{1}{\alpha}\right)\left(1-\frac{1}{\alpha+\beta}\right)\right) = \log_2\left(1+\frac{\beta-1}{\alpha\left(\alpha+\beta\right)}\right). \tag{A.18}$$

For the previous example, where the MMSE pdf for $|p|^2 = 1$ and $\sigma^2 = 10$ is approximated by a beta distribution, with $\alpha = 6.54162$ and $\beta = 1.12133$, the gap value is $\log_2(1 + 0.0024) = 0.0035$.

Appendix B

Standard Interference Function Framework

The *standard* IF framework characterizes a family of functions and provides a solution to distribute the available power between a set of users, which interfere each other. It was proposed by Yates in [38] to solve the power control algorithm

$$\boldsymbol{\xi}(n+1) = \boldsymbol{I}\left(\boldsymbol{\xi}\left(n\right)\right), \tag{B.1}$$

where $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_K]^{\mathrm{T}}$ contains the powers for the K users, and $\boldsymbol{I}(\boldsymbol{\xi}) = [I_1(\boldsymbol{\xi}), \ldots, I_K(\boldsymbol{\xi})]^{\mathrm{T}}$ is the vector containing the interferences seen at each of the users. In the following, some of the lemmas proposed in [38] that we find useful for our optimization problem will be explained in more detail.

An IF $\boldsymbol{I}(\boldsymbol{\xi})$ is feasible when a feasible solution $\boldsymbol{\xi}$ exists, i.e,

$$\boldsymbol{\xi} \geq \boldsymbol{I}(\boldsymbol{\xi}). \tag{B.2}$$

In other words, if $\boldsymbol{I}(\boldsymbol{\xi})$ is feasible, a solution for the problem can be found. Another important procedure to solve the power control problem is the observation of whether IF is *standard* or not. An IF $\boldsymbol{I}(\boldsymbol{\xi})$ is said to be *standard* when the following properties are satisfied for $\boldsymbol{\xi} \geq \mathbf{0}$

- positivity, $\boldsymbol{I}(\boldsymbol{\xi}) > \mathbf{o}$;
- monotonicity, $\boldsymbol{I}(\boldsymbol{\xi}') \geq \boldsymbol{I}(\boldsymbol{\xi}), \quad \text{for} \quad \boldsymbol{\xi}' \geq \boldsymbol{\xi}$; and
- scalability, $a\boldsymbol{I}(\boldsymbol{\xi}) > \boldsymbol{I}(a\boldsymbol{\xi}), \quad \text{for} \quad a > 1$.

When an IF is *standard* and feasible, the power control algorithm of (B.1) converges to the optimal solution of (B.2). The optimum would be reached when the power is

minimum and satisfies (B.2), that is, (B.2) would be fulfilled with equality in the optimum $\boldsymbol{\xi}^{\text{opt}} = \boldsymbol{I}(\boldsymbol{\xi}^{\text{opt}})$. Moreover, the optimum $\boldsymbol{\xi}^{\text{opt}}$ is unique, as we will show in the following.

Let us consider a feasible solution of $\boldsymbol{I}(\boldsymbol{\xi})$, $\boldsymbol{\xi}$. Then $\boldsymbol{\xi}(1) = \boldsymbol{I}(\boldsymbol{\xi}(0)) \leq \boldsymbol{\xi}(0) = \boldsymbol{\xi}$, and $\boldsymbol{I}(\boldsymbol{\xi}(0)) \geq \boldsymbol{I}(\boldsymbol{\xi}(1))$ due to the monotonicity property. Thus, for every iteration $\boldsymbol{\xi}(n+1) = \boldsymbol{I}(\boldsymbol{\xi}(n)) \leq \boldsymbol{\xi}(n)$ holds. Since $\boldsymbol{\xi}(n)$ is lower bounded by $\mathbf{0}$, the decreasing sequence converges to the optimum $\boldsymbol{\xi}^{\text{opt}}$. Note that no assumption but feasibility is made for $\boldsymbol{\xi}(0)$.

Regarding optimum uniqueness, suppose that there exists an additional optimum $\boldsymbol{\xi}' = \boldsymbol{I}(\boldsymbol{\xi}')$ such that $\boldsymbol{\xi}^{\text{opt}} = \boldsymbol{\Gamma}\boldsymbol{\xi}'$, for $\boldsymbol{\Gamma} = \text{diag}(\gamma_1, \ldots, \gamma_K) > \mathbf{0}$. We consider that for some user k, $\gamma_k \geq \gamma_j, \forall j$, and $\gamma_k > 1$, without loss of generality. Hence, $\gamma_k \boldsymbol{\xi}' \geq \boldsymbol{\xi}^{\text{opt}}$, and due to the monotonicity and the scalability properties above explained, we get the following contradiction

$$\xi_k^{\text{opt}} = I_k\left(\boldsymbol{\xi}^{\text{opt}}\right) \leq I_k\left(\gamma_k \boldsymbol{\xi}'\right) < \gamma_k I_k\left(\boldsymbol{\xi}'\right) = \gamma_k \boldsymbol{\xi}'_k. \tag{B.3}$$

Consider now an IF that not only depends on the power $\boldsymbol{\xi}$ but also on a matrix $\boldsymbol{X}$, i.e., $\boldsymbol{I}_{\boldsymbol{X}}(\boldsymbol{\xi}, \boldsymbol{X})$. It is not surprising that the selection over all the matrices $\boldsymbol{X}$ preserves the properties of the *standard* IFs if the optimum choice for $\boldsymbol{X}$ is the one that minimizes the IF. That is to say, $\boldsymbol{I}(\boldsymbol{\xi}) = \min_{\boldsymbol{X}} \boldsymbol{I}_{\boldsymbol{X}}(\boldsymbol{\xi}, \boldsymbol{X})$ is a *standard* IF.

The *standard* IF framework is a powerful tool that has been successfully used in previous related works (e.g. [34, 70]).

Appendix C

Matrix Properties

In this appendix, we present some algebraic properties of the matrices exploited in the developments presented throughout this work.

C.1. Eigenvalue Decomposition

Consider the square matrix $\boldsymbol{A} \in \mathbb{C}^{n \times n}$. A vector $\boldsymbol{x} \in \mathbb{C}^n$ is an eigenvector of $\boldsymbol{A}$, and λ its corresponding eigenvalue, if

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}. \tag{C.1}$$

We can equivalently rewrite the previous equation to get

$$\left(\boldsymbol{A} - \mathbf{I}_n \lambda\right) \boldsymbol{x} = \mathbf{0}, \tag{C.2}$$

which allows us to define the characteristic polynomial of $\boldsymbol{A}$. That is, the polynomial of degree n resulting from

$$\det\left(\boldsymbol{A} - \mathbf{I}_n \lambda\right) = 0. \tag{C.3}$$

Note that $\boldsymbol{M}\boldsymbol{x} = \mathbf{0}$ for the non-zero vector $\boldsymbol{x}$ only if $\boldsymbol{M}$ is singular. As a consequence, every eigenvalue of the matrix $\boldsymbol{A}$ is a root of the characteristic polynomial.

The eigenvalue decomposition of $\boldsymbol{A}$ comes from the equality [101]

$$\boldsymbol{A}\begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_n\end{bmatrix} = \begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_n\end{bmatrix}\begin{pmatrix}\lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n\end{pmatrix}, \tag{C.4}$$

and it is given by the factorization

$$\boldsymbol{\Lambda} = \boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U}, \tag{C.5}$$

with $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. From the previous expression it is straightforward to see that not all the square matrices have an eigenvalue decomposition. A matrix $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ is called defective if the number of linearly independent eigenvectors is less than n. In such a case, the eigenvalue decomposition can not be performed.

C.2. Determinant Properties

The determinant of a 2×2 matrix $\boldsymbol{A}$ is defined as $[\boldsymbol{A}]_{1,1}[\boldsymbol{A}]_{2,2} - [\boldsymbol{A}]_{1,2}[\boldsymbol{A}]_{2,1}$. For the more general case where $\boldsymbol{A} \in \mathbb{C}^{n \times n}$, it can be defined as a recursive function as follows

$$\det(\boldsymbol{A}) = \sum_{j=1}^{n} (-1)^{i+j} [\boldsymbol{A}]_{i,j} M_{i,j}, \tag{C.6}$$

for any row $i \in [1, \ldots, n]$, and where $M_{i,j}$ is the determinant of the matrix resulting from removing the ith row and jth column of $\boldsymbol{A}$.

The determinant operator satisfies the following properties:

1. The determinant is commutative with respect to the matrix product. Given the matrices $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ and $\boldsymbol{B} \in \mathbb{C}^{n \times n}$, it means that

$$\det(\boldsymbol{AB}) = \det(\boldsymbol{A}) \det(\boldsymbol{B}) = \det(\boldsymbol{BA}). \tag{C.7}$$

2. The determinant of a scalar a times a matrix $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ is

$$\det(a\boldsymbol{A}) = a^n \det(\boldsymbol{A}). \tag{C.8}$$

3. The inverse of the determinant of a matrix $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ is equal to the determinant of its inverse, i.e.,

$$\frac{1}{\det(\boldsymbol{A})} = \det\left(\boldsymbol{A}^{-1}\right). \tag{C.9}$$

4. The determinant of an square matrix $\boldsymbol{A} \in \mathbb{C}^{n \times n}$, with an eigenvalue decomposition $\boldsymbol{A} = \boldsymbol{U\Lambda U}^{-1}$, where $\boldsymbol{U}$ is the basis and the diagonal matrix $\boldsymbol{\Lambda}$ contains the eigenvalues, is

$$\det(\boldsymbol{A}) = \det\left(\boldsymbol{U\Lambda U}^{-1}\right) = \det\left(\boldsymbol{U}^{-1}\boldsymbol{U\Lambda}\right) = \det(\boldsymbol{\Lambda}) = \prod_{i=1}^{n} [\boldsymbol{\Lambda}]_{i,i}, \tag{C.10}$$

   i.e., the determinant of $\boldsymbol{A}$ is equivalent to the product of its eigenvalues.

5. Suppose the matrices $\boldsymbol{A} \in \mathbb{C}^{n\times m}$, $\boldsymbol{B} \in \mathbb{C}^{n\times m}$, $\boldsymbol{C} \in \mathbb{C}^{m\times n}$, and $\boldsymbol{D} \in \mathbb{C}^{m\times m}$. Then, the determinant of a block matrix is

$$\det \begin{pmatrix} \boldsymbol{A} & \mathbf{0} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix} = \det \begin{pmatrix} \boldsymbol{A} & \mathrm{B} \\ \mathbf{0} & \boldsymbol{D} \end{pmatrix} = \det(\boldsymbol{A}) \det(\boldsymbol{D}). \tag{C.11}$$

Furthermore, when $\boldsymbol{A}$ is invertible,

$$\begin{aligned} \det \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix} &= \det \begin{pmatrix} \boldsymbol{A} & \mathbf{0} \\ \boldsymbol{C} & \mathbf{I}_m \end{pmatrix} \det \begin{pmatrix} \mathbf{I}_n & \boldsymbol{A}^{-1}\boldsymbol{B} \\ \mathbf{0} & \boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B} \end{pmatrix} \\ &= \det(\boldsymbol{A}) \det(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}). \end{aligned} \tag{C.12}$$

6. Considering $\boldsymbol{A} \in \mathbb{C}^{n\times m}$, and $\boldsymbol{B} \in \mathbb{C}^{m\times n}$, the Sylvester's theorem states that

$$\det\left(\mathbf{I}_n + \boldsymbol{A}\boldsymbol{B}\right) = \det\left(\mathbf{I}_m + \boldsymbol{B}\boldsymbol{A}\right). \tag{C.13}$$

This equality can be shown using two different block decompositions of the same matrix, and applying the previous statement, i.e.

$$\begin{aligned} \det \begin{pmatrix} \mathbf{I}_n & -\boldsymbol{A} \\ \boldsymbol{B} & \mathbf{I}_m \end{pmatrix} &= \det \begin{pmatrix} \mathbf{I}_n & \mathbf{0} \\ \boldsymbol{B} & \mathbf{I}_m \end{pmatrix} \det \begin{pmatrix} \mathbf{I}_n & -\boldsymbol{A} \\ \mathbf{0} & \mathbf{I}_m + \boldsymbol{B}\boldsymbol{A} \end{pmatrix} = \det\left(\mathbf{I}_m + \boldsymbol{B}\boldsymbol{A}\right), \\ \det \begin{pmatrix} \mathbf{I}_n & -\boldsymbol{A} \\ \boldsymbol{B} & \mathbf{I}_m \end{pmatrix} &= \det \begin{pmatrix} \mathbf{I}_n + \boldsymbol{A}\boldsymbol{B} & -\boldsymbol{A} \\ \mathbf{0} & \mathbf{I}_m \end{pmatrix} \det \begin{pmatrix} \mathbf{I}_n & \mathbf{0} \\ \boldsymbol{B} & \mathbf{I}_m \end{pmatrix} = \det\left(\mathbf{I}_n + \boldsymbol{A}\boldsymbol{B}\right). \end{aligned}$$

C.3. Trace Properties

The trace operator satisfies the following properties:

1. In linear algebra, the trace of an $n \times n$ square matrix $\boldsymbol{A}$ is defined as the sum of the elements on the main diagonal, which gives

$$\mathrm{tr}\left(\boldsymbol{A}\right) = \sum_{i=1}^{n} [\boldsymbol{A}]_{i,i} = \sum_{i=1}^{n} [\boldsymbol{A}^{\mathrm{T}}]_{i,i} = \mathrm{tr}\left(\boldsymbol{A}^{\mathrm{T}}\right). \tag{C.14}$$

2. The trace is a linear operator, and then it satisfies

$$\mathrm{tr}\left(a\boldsymbol{A} + b\boldsymbol{B}\right) = \mathrm{tr}\left(a\boldsymbol{A}\right) + \mathrm{tr}\left(b\boldsymbol{B}\right) = a\,\mathrm{tr}\left(\boldsymbol{A}\right) + b\,\mathrm{tr}\left(\boldsymbol{B}\right). \tag{C.15}$$

3. The trace of a product of matrices verifies

$$\mathrm{tr}\left(\boldsymbol{A}\boldsymbol{B}\right) = \sum_{i=1}^{n} [\boldsymbol{A}\boldsymbol{B}]_{i,i} = \sum_{i=1}^{n} [\boldsymbol{B}\boldsymbol{A}]_{i,i} = \mathrm{tr}\left(\boldsymbol{B}\boldsymbol{A}\right). \tag{C.16}$$

4. The trace of an square matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, with an eigenvalue decomposition $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1}$, where $\boldsymbol{U}$ is the basis and the diagonal matrix $\boldsymbol{\Lambda}$ contains the eigenvalues, is

$$\operatorname{tr}\left(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1}\right) = \operatorname{tr}\left(\boldsymbol{U}^{-1}\boldsymbol{U}\boldsymbol{\Lambda}\right) = \operatorname{tr}\left(\boldsymbol{\Lambda}\right) = \sum_{i=1}^{n}[\boldsymbol{\Lambda}]_{i,i}, \tag{C.17}$$

   i.e., the trace of $\boldsymbol{A}$ is equal to the sum of its eigenvalues.

C.4. Matrix Inversion Lemma

The matrix inversion lemma for squared matrices, $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$, states that

$$(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} = \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})\boldsymbol{C}\boldsymbol{A}^{-1}, \tag{C.18}$$

assuming that $\boldsymbol{A}$ and $\boldsymbol{D}$ are invertible. The direct proof is as follows,

$$\begin{aligned}
&(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})\left[\boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}\right] \\
&\quad = \mathbf{I} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} + \boldsymbol{B}\left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} \\
&\quad - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} \\
&\quad = \mathbf{I} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} + \left(\mathbf{I} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}\right)\boldsymbol{B}\left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} \\
&\quad = \mathbf{I} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} + \left(\boldsymbol{B} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)\left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} \\
&\quad = \mathbf{I} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} + \left(\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{D} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)\left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} \\
&\quad = \mathbf{I} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} + \boldsymbol{B}\boldsymbol{D}^{-1}\left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)\left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} \\
&\quad = \mathbf{I} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} + \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} \\
&\quad = \mathbf{I}.
\end{aligned}$$

Another insightful proof is provided employing block matrices

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix},$$

where it is assumed that $\boldsymbol{M}$ is invertible. The matrix $\boldsymbol{M}$ can be decomposed as follows

$$\boldsymbol{M} = \boldsymbol{X}\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{A} & \mathbf{0} \\ \boldsymbol{C} & \mathbf{I} \end{pmatrix}\begin{pmatrix} \mathbf{I} & \boldsymbol{A}^{-1}\boldsymbol{B} \\ \mathbf{0} & \boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B} \end{pmatrix}.$$

Now, to compute the inverse of $\boldsymbol{M}$ we can calculate the product $\boldsymbol{Y}^{-1}\boldsymbol{X}^{-1}$. Thanks to the structure of the matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ including the blocks $\mathbf{0}$ and $\mathbf{I}$, the inverses can be easily

obtained, i.e.

$$\boldsymbol{X}^{-1} = \begin{pmatrix} \boldsymbol{A}^{-1} & \mathbf{0} \\ -\boldsymbol{C}\boldsymbol{A}^{-1} & \mathbf{I} \end{pmatrix},$$
$$\boldsymbol{Y}^{-1} = \begin{pmatrix} \mathbf{I} & -\boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \\ \mathbf{0} & \left(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \end{pmatrix}.$$

And the inverse of the matrix $\boldsymbol{M}$ is the product of both of them, that is

$$\begin{aligned} \boldsymbol{M}^{-1} &= \boldsymbol{Y}^{-1}\boldsymbol{X}^{-1} \\ &= \begin{pmatrix} \mathbf{I} & -\boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \\ \mathbf{0} & \left(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{A}^{-1} & \mathbf{0} \\ -\boldsymbol{C}\boldsymbol{A}^{-1} & \mathbf{I} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{A}^{-1}+\boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \\ -\left(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & \left(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \end{pmatrix}. \end{aligned} \tag{C.19}$$

Consider now an alternative decomposition for the matrix $\boldsymbol{M}$, where the blocks $\mathbf{0}$ and $\mathbf{I}$ are placed in rows instead of columns as follows

$$\boldsymbol{M} = \boldsymbol{Z}\boldsymbol{W} = \begin{pmatrix} \boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C} & \boldsymbol{B}\boldsymbol{D}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}.$$

Exploiting again the structure of $\boldsymbol{Z}$ and $\boldsymbol{W}$, the respective inverses are computed via

$$\boldsymbol{Z}^{-1} = \begin{pmatrix} \left(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)^{-1} & -\left(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix},$$
$$\boldsymbol{W}^{-1} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\boldsymbol{D}^{-1}\boldsymbol{C} & \boldsymbol{D}^{-1} \end{pmatrix}.$$

That way, we obtain a different expression for the inverse of $\boldsymbol{M}$ as $\boldsymbol{W}^{-1}\boldsymbol{Z}^{-1}$, which leads to

$$\boldsymbol{M}^{-1} = \begin{pmatrix} \left(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)^{-1} & -\left(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \\ -\boldsymbol{D}^{-1}\boldsymbol{C}\left(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)^{-1} & \boldsymbol{D}^{-1}+\boldsymbol{D}^{-1}\boldsymbol{C}\left(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \end{pmatrix}. \tag{C.20}$$

By comparing the upper left elements of (C.19) and (C.20) we find the desired equality [102] (cf. (C.18)).

C.5. Positive Definite Matrices

Consider the Hermitian matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, such that $\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{A}$. The matrix $\boldsymbol{A}$ is positive definite if

$$\boldsymbol{z}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{z} > 0, \tag{C.21}$$

for any $\boldsymbol{z} \in \mathbb{C}^n \neq \mathbf{0}$ [103]. In the particular case of real symmetric matrices, the property holds if $\boldsymbol{z}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{z} > 0$. When the inequality is weakened to $\geq$, the matrices fulfilling such an inequality are called positive semidefinite. Of course, any positive definite matrix is also positive semidefinite.

Some useful properties of the positive (semi)definite matrices are:

1. Consider the positive (semi)definite matrices $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, and $\boldsymbol{B} \in \mathbb{C}^{n\times n}$, and the scalars $a \geq 0$, and $b \geq 0$. Then, the linear combination of the two matrices results into a positive semidefinite matrix $\boldsymbol{C} \in \mathbb{C}^{n\times n} = a\boldsymbol{A} + b\boldsymbol{B}$

$$\boldsymbol{z}^{\mathrm{H}}\boldsymbol{C}\boldsymbol{z} = \boldsymbol{z}^{\mathrm{H}}\left(a\boldsymbol{A} + b\boldsymbol{B}\right)\boldsymbol{z} = a\boldsymbol{z}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{z} + b\boldsymbol{z}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{z} \geq 0. \tag{C.22}$$

   Note that if $\boldsymbol{A}$ and $\boldsymbol{B}$ are positive definite and the scalars satisfy $a > 0$, and $b > 0$, the inequality is strict, and $\boldsymbol{C}$ is also positive definite.

2. If a matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ is positive (semi)definite, $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{A}$, and λ its corresponding eigenvalue, then

$$\boldsymbol{u}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{u} = \boldsymbol{u}^{\mathrm{H}}\lambda\boldsymbol{u} = \lambda\left\|\boldsymbol{u}\right\|_2^2. \tag{C.23}$$

   Therefore, $\lambda = \boldsymbol{u}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{u}/||\boldsymbol{u}||_2^2$. Since $\boldsymbol{A}$ is positive (semi)definite, the product in the numerator is greater (or equal) to 0, and then $\lambda > 0$ for positive definite $\boldsymbol{A}$ and $\lambda \geq 0$ for positive semidefinite $\boldsymbol{A}$.

3. The inverse of a positive (semi)definite matrix $\boldsymbol{A}$, $\boldsymbol{A}^{-1}$, is also positive (semi)definite. To prove that, let us introduce the vector $\boldsymbol{x}$ such that $\boldsymbol{A}\boldsymbol{z} = \boldsymbol{x}$. That way, we obtain the product

$$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{-1}\boldsymbol{x} = \boldsymbol{z}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{z} = \boldsymbol{z}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{z}. \tag{C.24}$$

   Since $\boldsymbol{A}$ is positive (semi)definite, $\boldsymbol{z}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{z}$ is greater than (or equal to) 0.

4. The Cholesky decomposition is a factorization of a Hermitian positive definite matrix $\boldsymbol{A}$ of the form

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{H}}, \tag{C.25}$$

   where $\boldsymbol{L}$ is a lower triangular matrix with real diagonal entries. Moreover, for every Hermitian positive definite $\boldsymbol{A}$ the decomposition is unique [101].

5. The Gram of a matrix $\boldsymbol{A} \in \mathbb{C}^{n\times m}$, $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$, is a positive semidefinite matrix. Consider some vector $\boldsymbol{z} \in \mathbb{C}^m$. Then

$$\boldsymbol{z}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{z} = \boldsymbol{w}^{\mathrm{H}}\boldsymbol{w} = \left\|\boldsymbol{w}\right\|_2^2 \geq 0. \tag{C.26}$$

Using a similar reasoning, for a positive definite matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ and some matrix $\boldsymbol{B} \in \mathbb{C}^{n\times m}$, the product $\boldsymbol{B}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{B}$ is positive semidefinite. Thus, for any vector $\boldsymbol{z} \in \mathbb{C}^m$

$$\boldsymbol{z}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{B}\boldsymbol{z} = \boldsymbol{w}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{w} \geq 0. \tag{C.27}$$

Note that (C.26) can be seen as a particular case of (C.27) if $\boldsymbol{A} = \mathbf{I}_n$.

C.6. Complex Derivatives

Let us first introduce the derivative of the function $f(\boldsymbol{X})$, whose argument is a real matrix $\boldsymbol{X} \in \mathbb{R}^{m\times n}$, $f : \mathbb{R}^{m\times n} \to \mathbb{R}$, as

$$\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}} = \begin{pmatrix} \frac{\partial f(\boldsymbol{X})}{\partial [\boldsymbol{X}]_{1,1}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial [\boldsymbol{X}]_{1,n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\boldsymbol{X})}{\partial [\boldsymbol{X}]_{m,1}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial [\boldsymbol{X}]_{m,n}} \end{pmatrix} \in \mathbb{R}^{m\times n}. \tag{C.28}$$

Given a matrix $\boldsymbol{X} \in \mathbb{C}^{m\times n}$, and $f : \mathbb{C}^{m\times n} \to \mathbb{C}$, it has to be satisfied [104, 105]

$$\frac{\partial f(\boldsymbol{X})}{\partial \Im\{\boldsymbol{X}\}} = j\frac{\partial f(\boldsymbol{X})}{\partial \Re\{\boldsymbol{X}\}}.$$

Since the last equality is not fulfilled in general, a generalized definition of complex derivatives is used instead [104, 105] in this way

$$\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}} = \frac{1}{2}\begin{pmatrix} \frac{\partial f(\boldsymbol{X})}{\partial \Re\{[\boldsymbol{X}]_{1,1}\}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial \Re\{[\boldsymbol{X}]_{1,n}\}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\boldsymbol{X})}{\partial \Re\{[\boldsymbol{X}]_{m,1}\}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial \Re\{[\boldsymbol{X}]_{m,n}\}} \end{pmatrix} - \frac{j}{2}\begin{pmatrix} \frac{\partial f(\boldsymbol{X})}{\partial \Im\{[\boldsymbol{X}]_{1,1}\}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial \Im\{[\boldsymbol{X}]_{1,n}\}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\boldsymbol{X})}{\partial \Im\{[\boldsymbol{X}]_{m,1}\}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial \Im\{[\boldsymbol{X}]_{m,n}\}} \end{pmatrix},$$

and the complex conjugate derivative is

$$\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}^*} = \frac{1}{2}\begin{pmatrix} \frac{\partial f(\boldsymbol{X})}{\partial \Re\{[\boldsymbol{X}]_{1,1}\}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial \Re\{[\boldsymbol{X}]_{1,n}\}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\boldsymbol{X})}{\partial \Re\{[\boldsymbol{X}]_{m,1}\}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial \Re\{[\boldsymbol{X}]_{m,n}\}} \end{pmatrix} + \frac{j}{2}\begin{pmatrix} \frac{\partial f(\boldsymbol{X})}{\partial \Im\{[\boldsymbol{X}]_{1,1}\}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial \Im\{[\boldsymbol{X}]_{1,n}\}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\boldsymbol{X})}{\partial \Im\{[\boldsymbol{X}]_{m,1}\}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial \Im\{[\boldsymbol{X}]_{m,n}\}} \end{pmatrix}.$$

Taking into account the complex derivatives above described, the following results are used throughout this work

- $\frac{\partial \operatorname{tr}(\boldsymbol{A}\boldsymbol{X})}{\partial \boldsymbol{X}^*} = \mathbf{0}$,
- $\frac{\partial \operatorname{tr}(\boldsymbol{A}\boldsymbol{X}^{\mathrm{H}})}{\partial \boldsymbol{X}^*} = \boldsymbol{A}$,

- $\frac{\partial(\mathrm{tr}(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}))}{\partial \boldsymbol{X}^{*}} = \boldsymbol{A}\boldsymbol{X}$,
- $\frac{\partial(\mathrm{tr}(\Re\{\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\}))}{\partial \boldsymbol{X}^{*}} = \frac{1}{2}\frac{\partial(\mathrm{tr}(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}))}{\partial \boldsymbol{X}^{*}} + \frac{1}{2}\frac{\partial(\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{*}))}{\partial \boldsymbol{X}^{*}} = \frac{1}{2}\boldsymbol{A}$,
- $\frac{\partial\,\mathrm{tr}(\boldsymbol{A}\boldsymbol{X}^{-1})}{\partial \boldsymbol{X}^{\mathrm{T}}} = -\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X}^{-1}$,
- $\frac{\partial f(\boldsymbol{X})}{\partial [\boldsymbol{X}]_{i,j}} = \sum_{k,l}\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}_{k,l}}\frac{\partial \boldsymbol{X}_{k,l}}{\partial \boldsymbol{X}_{i,j}} = \mathrm{tr}\left(\left(\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}}\right)^{\mathrm{T}}\frac{\partial \boldsymbol{X}}{\partial \boldsymbol{X}_{i,j}}\right)$, where $f(\boldsymbol{X})$ is a scalar function,
- $\frac{\partial \det(\boldsymbol{A})}{\partial x} = \det(\boldsymbol{A})\,\mathrm{tr}\left(\boldsymbol{A}^{-1}\frac{\partial \boldsymbol{A}}{\partial x}\right)$.

C.7. Jacobian Matrix

Consider a function $\boldsymbol{f}(\boldsymbol{x}) : \mathbb{R}^n \mapsto \mathbb{R}^m$, which is given by m real valued component functions such that $f_i(\boldsymbol{x}) : \mathbb{R}^n \mapsto \mathbb{R}$, for $i \in \{1, \ldots, m\}$. The partial derivatives of the functions $f_i(\boldsymbol{x})$ with respect to the vector of variables $\boldsymbol{x}$ can be collected into the $m \times n$ Jacobian Matrix

$$\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{x}) = \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}}. \tag{C.29}$$

The former expression lead us to

$$\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\boldsymbol{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\boldsymbol{x})}{\partial x_n} \end{pmatrix}. \tag{C.30}$$

The Jacobian matrix generalizes the derivative of a scalar valued function of a single variable. Moreover, similarly to the Taylor series for a scalar function of a single variable, the Jacobian matrix evaluated at the point $\boldsymbol{x}_0$ is the best linear approximation of $\boldsymbol{f}(\boldsymbol{x})$ in the neighborhood of $\boldsymbol{x}_0$, that is

$$\boldsymbol{f}(\boldsymbol{x}) \approx \boldsymbol{f}(\boldsymbol{x}_0) + \boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{x}_0)\,(\boldsymbol{x} - \boldsymbol{x}_0)\,. \tag{C.31}$$

Appendix D

Karush-Kuhn-Tucker Conditions

The optimization problems can be described in general by a function to minimize and several side restrictions, leading to the following problem statement [14, 106–109]

$$\begin{aligned} \boldsymbol{X}^{\text{opt}} = \operatorname{argmin}_{\boldsymbol{X}} f(\boldsymbol{X}) \quad \text{subject to} \quad & g_i(\boldsymbol{X}) \leq 0,\, \forall i \in \{1,\ldots,l\} \\ & h_j(\boldsymbol{X}) = 0,\, \forall j \in \{1,\ldots,m\}, \end{aligned} \tag{D.1}$$

where $\boldsymbol{X}$ and $\boldsymbol{X}^{\text{opt}} \in \mathbb{C}^{m\times n}$. The functions $f(\boldsymbol{X})$, $g_i(\boldsymbol{X})$, $i = 1,\ldots,l$, and $h_j(\boldsymbol{X})$, $j = 1,\ldots,m$, are real-valued with complex-valued arguments, i.e.

$$\begin{aligned} f &: \mathbb{C}^{m\times n} \to \mathbb{R}, \\ g_i &: \mathbb{C}^{m\times n} \to \mathbb{R}, \quad i = 1,\ldots,l \\ h_j &: \mathbb{C}^{m\times n} \to \mathbb{R}, \quad j = 1,\ldots,m. \end{aligned}$$

The function to be minimized is $f(\boldsymbol{X})$, whereas $g_i(\boldsymbol{X})$ and $h_j(\boldsymbol{X})$ are the constraints. This problem optimization (D.1) can be solved via Lagrangian functions in most cases [14, 108]

$$L\left(\boldsymbol{X}, \boldsymbol{\lambda}, \boldsymbol{v}\right) = f(\boldsymbol{X}) + \sum_{i=1}^{l} \lambda_i g_i(\boldsymbol{X}) + \sum_{j=1}^{m} v_j h_j(\boldsymbol{X}),$$

with $\lambda_i \in \mathbb{R}^{0,+}$, for $i = 1,\ldots,l$, and $v_j \in \mathbb{R}$, for $j = 1,\ldots,m$. This allows to rewrite the problem formulation (D.1) as a new optimization problem without constraints

$$\boldsymbol{X}^{\text{opt}} = \max_{\boldsymbol{\lambda},\boldsymbol{v}} \operatorname{argmin}_{\boldsymbol{X}} L\left(\boldsymbol{X}, \boldsymbol{\lambda}, \boldsymbol{v}\right). \tag{D.2}$$

A solution to (D.2) is also a solution of (D.1).

The *Karush-Kuhn-Tucker* (KKT) have to be fulfilled for any solution of the problem statement (D.1) [106, 107]. We can equivalently focus on (D.2) instead of (D.1) to get

such conditions as the following first order derivative

$$\begin{aligned} \frac{\partial L\left(\boldsymbol{X}, \boldsymbol{\lambda}, \boldsymbol{v}\right)}{\partial \boldsymbol{X}} &= \mathbf{0}, \\ g_i(\boldsymbol{X}) &\leq 0 \quad i = 1, \ldots, l, \\ \lambda_i g_i(\boldsymbol{X}) &= 0 \quad i = 1, \ldots, l, \\ \lambda_i &\geq 0 \quad i = 1, \ldots, l, \\ v_j(\boldsymbol{X}) &= 0 \quad j = 1, \ldots, m, \end{aligned} \tag{D.3}$$

which are also necessary to get the optimum of (D.1).

Since the conditions in (D.3) are not sufficient in general, and there is a dependence on the functions $f(\boldsymbol{X})$, $h_j(\boldsymbol{X})$, and $g_i(\boldsymbol{X})$, we are able to conclude the optimality of the KKT conditions. If the functions $f(\boldsymbol{X})$ and $g_i(\boldsymbol{X})$, $\forall i$, are convex, and $h_j(\boldsymbol{X})$, $\forall j$, are affine [14], the KKT conditions are necessary and sufficient for global optimality. However, if the mentioned functions are locally convex or affine, the KKT conditions only guarantee local optimality.

Appendix E

Multivariate Normal Distribution

The multivariate Normal distribution, also called multivariate Gaussian distribution, arises as the generalization of the one dimension Normal distribution. Its importance derives from several reasons, e.g. from the multivariate Central Limit Theorem. Let us define the vector $\boldsymbol{x} = [x_1, \ldots, x_n]^{\mathrm{T}} \in \mathbb{C}^n$. The mean value of $\boldsymbol{x}$ is then

$$\boldsymbol{\mu_x} = \mathrm{E}[\boldsymbol{x}] = [\mu_1, \ldots, \mu_n]^{\mathrm{T}} \in \mathbb{C}^n, \tag{E.1}$$

where $\mu_i = \mathrm{E}[x_i]$. Consequently, the components of the vector $\boldsymbol{\mu_x}$ are means themselves. Similarly, the covariance matrix of $\boldsymbol{x}$, $\boldsymbol{C_x} \in \mathbb{C}^{n \times n}$, is given by

$$\boldsymbol{C_x} = \mathrm{E}[(\boldsymbol{x} - \boldsymbol{\mu_x})(\boldsymbol{x} - \boldsymbol{\mu_x})^{\mathrm{H}}]. \tag{E.2}$$

Note that the entries of $\boldsymbol{C_x}$ are the covariances of the components of $\boldsymbol{x}$, that is $[\boldsymbol{C_x}]_{i,j} = \mathrm{E}[(x_i - \mu_i)(x_j - \mu_j)^*]$. Therefore, the matrix $\boldsymbol{C_x}$ can be considered Hermitian and positive definite for the regular cases, where $\boldsymbol{C_x}$ is full rank.

Considering the mean and the covariance previously introduced, the random vector $\boldsymbol{x}$ is said to be zero-mean circularly symmetric complex Gaussian if its pdf is given by [73]

$$f(\boldsymbol{x}) = \frac{1}{\pi^n \det(\boldsymbol{C_x})} \mathrm{e}^{\left(-(\boldsymbol{x}-\boldsymbol{\mu_x})^{\mathrm{H}} \boldsymbol{C_x}^{-1} (\boldsymbol{x}-\boldsymbol{\mu_x})\right)}, \tag{E.3}$$

with $\boldsymbol{\mu_x} = \boldsymbol{0}$.

E.1. Joint Probability of two Gaussian

Let us define the circularly symmetric complex Gaussian random vectors $\boldsymbol{x} \in \mathbb{C}^n$ and $\boldsymbol{y} \in \mathbb{C}^m$. Consider now that $\boldsymbol{x}$ and $\boldsymbol{y}$ are jointly circularly symmetric complex Gaussian distributed, i.e.

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \sim \mathcal{N}_{\mathbb{C}} \left(\mathrm{E} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}, \boldsymbol{C} \right), \tag{E.4}$$

with the covariance and the mean as follows

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_{\boldsymbol{x}} & \boldsymbol{C}_{\boldsymbol{x},\boldsymbol{y}} \\ \boldsymbol{C}_{\boldsymbol{y},\boldsymbol{x}} & \boldsymbol{C}_{\boldsymbol{y}} \end{pmatrix}, \quad \mathrm{E}\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} = \begin{bmatrix} \mathrm{E}[\boldsymbol{x}] \\ \mathrm{E}[\boldsymbol{y}] \end{bmatrix}, \tag{E.5}$$

where $\boldsymbol{C}_{\boldsymbol{x}}$ and $\boldsymbol{C}_{\boldsymbol{y}}$ are the covariances of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, and $\boldsymbol{C}_{\boldsymbol{x},\boldsymbol{y}} = \mathrm{E}[(\boldsymbol{x} - \mathrm{E}[\boldsymbol{x}])(\boldsymbol{y} - \mathrm{E}[\boldsymbol{y}])^{\mathrm{H}}]$. Then, the pdf of the Gaussian random vector $[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}^{\mathrm{T}}]^{\mathrm{T}}$ reads

$$p(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{\pi^{n+m}\det(\boldsymbol{C})}\,\mathrm{e}^{-\begin{bmatrix} \boldsymbol{x}-\mathrm{E}[\boldsymbol{x}] \\ \boldsymbol{y}-\mathrm{E}[\boldsymbol{y}] \end{bmatrix}^{\mathrm{H}} \boldsymbol{C}^{-1} \begin{bmatrix} \boldsymbol{x}-\mathrm{E}[\boldsymbol{x}] \\ \boldsymbol{y}-\mathrm{E}[\boldsymbol{y}] \end{bmatrix}}. \tag{E.6}$$

The conditional pdf can be obtained via Bayes' rule

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{p(\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{x})}, \tag{E.7}$$

and accordingly

$$\frac{p(\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{x})} = \frac{\frac{1}{\pi^{n+m}\det(\boldsymbol{C})}\,\mathrm{e}^{-\begin{bmatrix} \tilde{\boldsymbol{x}} \\ \tilde{\boldsymbol{y}} \end{bmatrix}^{\mathrm{H}} \boldsymbol{C}^{-1} \begin{bmatrix} \tilde{\boldsymbol{x}} \\ \tilde{\boldsymbol{y}} \end{bmatrix}}}{\frac{1}{\pi^{n}\det(\boldsymbol{C}_x)}\,\mathrm{e}^{-\tilde{\boldsymbol{x}}^{\mathrm{H}}\boldsymbol{C}_x^{-1}\tilde{\boldsymbol{x}}}}, \tag{E.8}$$

where we have defined $\tilde{\boldsymbol{x}} = \boldsymbol{x} - \mathrm{E}[\boldsymbol{x}]$ and $\tilde{\boldsymbol{y}} = \boldsymbol{y} - \mathrm{E}[\boldsymbol{y}]$.

To simplify the previous expression, we first decompose the covariance matrix $\boldsymbol{C}$ as follows

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_{\boldsymbol{x}} & \mathbf{0} \\ \boldsymbol{C}_{\boldsymbol{y}\boldsymbol{x}} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \boldsymbol{C}_{\boldsymbol{x}}^{-1}\boldsymbol{C}_{\boldsymbol{x}\boldsymbol{y}} \\ \mathbf{0} & \boldsymbol{C}_{\boldsymbol{y}} - \boldsymbol{C}_{\boldsymbol{y}\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\boldsymbol{C}_{\boldsymbol{x}\boldsymbol{y}} \end{pmatrix}. \tag{E.9}$$

That way, the determinant of the covariance matrix can be easily computed as

$$\det(\boldsymbol{C}) = \det(\boldsymbol{C}_{\boldsymbol{x}})\det(\boldsymbol{B}), \tag{E.10}$$

with $\boldsymbol{B}$ the Schur complement of $\boldsymbol{C}$

$$\boldsymbol{B} = \boldsymbol{C}_{\boldsymbol{y}} - \boldsymbol{C}_{\boldsymbol{y},\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\boldsymbol{C}_{\boldsymbol{x},\boldsymbol{y}}. \tag{E.11}$$

Now, we rewrite the inverse of the covariance matrix as follows

$$\boldsymbol{C}^{-1} = \begin{pmatrix} \boldsymbol{C}_{\boldsymbol{x}}^{-1} + \boldsymbol{C}_{\boldsymbol{x}}^{-1}\boldsymbol{C}_{\boldsymbol{x},\boldsymbol{y}}\boldsymbol{B}^{-1}\boldsymbol{C}_{\boldsymbol{y},\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1} & -\boldsymbol{C}_{\boldsymbol{x}}^{-1}\boldsymbol{C}_{\boldsymbol{x},\boldsymbol{y}}\boldsymbol{B}^{-1} \\ -\boldsymbol{B}^{-1}\boldsymbol{C}_{\boldsymbol{y},\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1} & \boldsymbol{B}^{-1} \end{pmatrix} \tag{E.12}$$

$$= \begin{pmatrix} \mathbf{I} & -\boldsymbol{C}_{\boldsymbol{x}}^{-1}\boldsymbol{C}_{\boldsymbol{x},\boldsymbol{y}} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{C}_{\boldsymbol{x}}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\boldsymbol{C}_{\boldsymbol{y},\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1} & \mathbf{I} \end{pmatrix}. \tag{E.13}$$

Note that we have applied the matrix inversion lemma (see (C.19)) in (E.12), and decomposed the resulting matrix to obtain the expression of (E.13) (see Section C.4 of Appendix C). Employing (E.13), the exponent of the joint pdf (E.6) reads as

$$-\begin{bmatrix} \tilde{\boldsymbol{x}} \\ \tilde{\boldsymbol{y}} - \boldsymbol{C}_{\boldsymbol{y},\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\tilde{\boldsymbol{x}} \end{bmatrix}^{\mathrm{H}} \begin{pmatrix} \boldsymbol{C}_{\boldsymbol{x}}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B}^{-1} \end{pmatrix} \begin{bmatrix} \tilde{\boldsymbol{x}} \\ \tilde{\boldsymbol{y}} - \boldsymbol{C}_{\boldsymbol{y},\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\tilde{\boldsymbol{x}} \end{bmatrix}. \tag{E.14}$$

Computing the matrix product, we get

$$-\left(\tilde{\boldsymbol{x}}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\tilde{\boldsymbol{x}} + \left(\tilde{\boldsymbol{y}} - \boldsymbol{C}_{\boldsymbol{y},\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\tilde{\boldsymbol{x}}\right)^{\mathrm{H}} \boldsymbol{B}^{-1} \left(\tilde{\boldsymbol{y}} - \boldsymbol{C}_{\boldsymbol{y},\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\tilde{\boldsymbol{x}}\right)\right). \tag{E.15}$$

Correspondingly, from (E.15) and (E.10) is easy to see that (E.8) can be calculated as follows

$$p\left(\boldsymbol{y}|\boldsymbol{x}\right) = \frac{1}{\pi^m \det\left(\boldsymbol{B}\right)} \mathrm{e}^{-\left[\tilde{\boldsymbol{y}} - \boldsymbol{C}_{\boldsymbol{y},\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\tilde{\boldsymbol{x}}\right]^{\mathrm{H}}\boldsymbol{B}^{-1}\left[\tilde{\boldsymbol{y}} - \boldsymbol{C}_{\boldsymbol{y},\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\tilde{\boldsymbol{x}}\right]}. \tag{E.16}$$

Then, we conclude from the previous expression that for $\boldsymbol{y}$ conditioned to $\boldsymbol{x}$, the mean and the covariance matrix are as follows

$$\mathrm{E}\left[\boldsymbol{y}|\boldsymbol{x}\right] = \mathrm{E}[\boldsymbol{y}] + \boldsymbol{C}_{\boldsymbol{y},\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\left(\boldsymbol{x} - \mathrm{E}\left[\boldsymbol{x}\right]\right), \tag{E.17}$$

$$\boldsymbol{C}_{\boldsymbol{y}|\boldsymbol{x}} = \boldsymbol{C}_{\boldsymbol{y}} - \boldsymbol{C}_{\boldsymbol{y},\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\boldsymbol{C}_{\boldsymbol{x},\boldsymbol{y}}. \tag{E.18}$$

Appendix F

Convexity Proofs

Consider the function $g(\boldsymbol{X}) : \mathbb{C}^{M\times N} \rightarrow \mathbb{R}$. The function g is said to be convex if $\forall \boldsymbol{X}_1, \boldsymbol{X}_2 \in \mathbb{C}^{M\times N}$, $\forall t \in [0,1]$, satisfies

$$g(t\boldsymbol{X}_1 + (1-t)\boldsymbol{X}_2) \leq tg(\boldsymbol{X}_1) + (1-t)g(\boldsymbol{X}_2). \tag{F.1}$$

F.1. Sum-MSE Lower Bound

Let us consider the MAC MSE expression when we let the power of all the users increase unlimitedly. In other words, we neglect the noise to obtain

$$\text{MSE}_k^{\text{MAC}} = 1 - 2\Re\left\{\boldsymbol{g}_k^{\text{H}}\boldsymbol{H}_k\boldsymbol{t}_k\right\} + \sum_{i=1}^{K}\boldsymbol{g}_k^{\text{H}}\boldsymbol{H}_i\boldsymbol{t}_i\boldsymbol{t}_i^{\text{H}}\boldsymbol{H}_i^{\text{H}}\boldsymbol{g}_k. \tag{F.2}$$

In particular, we are interested in showing that is a convex function with respect to the receive filter $\boldsymbol{g}_k$. Now, taking into account that the function $h(\boldsymbol{g}_k) = 1 - 2\Re\left\{\boldsymbol{g}_k^{\text{H}}\boldsymbol{H}_k\boldsymbol{t}_k\right\}$ is affine, we focus on $\sum_{i=1}^{K}\boldsymbol{g}_k^{\text{H}}\boldsymbol{H}_i\boldsymbol{t}_i\boldsymbol{t}_i^{\text{H}}\boldsymbol{H}_i^{\text{H}}\boldsymbol{g}_k$. Using the property of convex functions $c_i(\boldsymbol{x})$ which states that $\sum_i \lambda_i c_i(\boldsymbol{x})$ is convex for $\lambda_i \geq 0$, we only have to show that

$$f(\boldsymbol{g}_k) = \boldsymbol{g}_k^{\text{H}}\sum_{i=1}^{K}\boldsymbol{H}_i\boldsymbol{t}_i\boldsymbol{t}_i^{\text{H}}\boldsymbol{H}_i^{\text{H}}\boldsymbol{g}_k = \boldsymbol{g}_k^{\text{H}}\boldsymbol{M}\boldsymbol{g}_k, \tag{F.3}$$

is convex. We will drop the subindex k for the shake of notational brevity in the following. Notice that the matrix $\boldsymbol{M}$ is positive semidefinite.

A function is convex if and only if it is convex when restricted to any line that intersects its domain [14]. Therefore, we introduce the function $h(t) = f(t\boldsymbol{g}_1 + (1-t)\boldsymbol{g}_2)$

with $t \in [0,1]$, and study the convexity of $h(t)$ instead

$$\frac{\partial h(t)}{\partial t} = \frac{\partial \left(t\boldsymbol{g}_1 + (1-t)\boldsymbol{g}_2\right)^{\mathrm{H}} \boldsymbol{M} \left(t\boldsymbol{g}_1 + (1-t)\boldsymbol{g}_2\right)}{\partial t}$$
$$= \left(\boldsymbol{g}_1 - \boldsymbol{g}_2\right)^{\mathrm{H}} \boldsymbol{M} \left(t\boldsymbol{g}_1 + (1-t)\boldsymbol{g}_2\right) + \left(t\boldsymbol{g}_1 + (1-t)\boldsymbol{g}_2\right)^{\mathrm{H}} \boldsymbol{M} \left(\boldsymbol{g}_1 - \boldsymbol{g}_2\right), \quad \text{(F.4)}$$
$$\frac{\partial^2 h(t)}{\partial t} = 2\left(\boldsymbol{g}_1 - \boldsymbol{g}_2\right)^{\mathrm{H}} \boldsymbol{M} \left(\boldsymbol{g}_1 - \boldsymbol{g}_2\right). \quad \text{(F.5)}$$

The last expression is greater than or equal to zero since $\boldsymbol{M}$ is positive semidefinite. Hence, $\mathrm{MSE}_k^{\mathrm{MAC}}$ and also the sum-MSE are convex. Moreover, when partial *Channel State Information* (CSI) is considered the conditioned expectation does not change the convexity due to the positive-weighted summation of the expectation operator.

F.2. Logarithm of Determinant Function

Let us define the function $f(\boldsymbol{X}) = \log(\det(\boldsymbol{X}))$, and the positive semidefinite matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$. Thus, to check the concavity of f, we introduce an additional function $g(t) = f(t\boldsymbol{X} + (1-t)\boldsymbol{Y})$ with the scalar $t \in [0,1]$. Recall that the concavity of a function can be checked when restricted to any line that intersects its domain [14]. First, we compute the derivative of $g(t)$ with respect to t

$$\frac{\partial g(t)}{\partial t} = \frac{\partial}{\partial t} \log\left(\det\left(t\boldsymbol{X} + (1-t)\boldsymbol{Y}\right)\right)$$
$$= \frac{1}{\det\left(t\boldsymbol{X} + (1-t)\boldsymbol{Y}\right)} \det\left(t\boldsymbol{X} + (1-t)\boldsymbol{Y}\right) \mathrm{tr}\left(\left(t\boldsymbol{X} + (1-t)\boldsymbol{Y}\right)^{-1} \left(\boldsymbol{X} - \boldsymbol{Y}\right)\right)$$
$$= \mathrm{tr}\left(\left(t\boldsymbol{X} + (1-t)\boldsymbol{Y}\right)^{-1} \left(\boldsymbol{X} - \boldsymbol{Y}\right)\right). \quad \text{(F.6)}$$

Now, we compute the second derivative as

$$\frac{\partial \,\mathrm{tr}\left(\left(t\boldsymbol{X} + (1-t)\boldsymbol{Y}\right)^{-1} \left(\boldsymbol{X} - \boldsymbol{Y}\right)\right)}{\partial t} =$$
$$-\mathrm{tr}\left(\left(\boldsymbol{X} - \boldsymbol{Y}\right)\left(t\boldsymbol{X} + (1-t)\boldsymbol{Y}\right)^{-1} \left(\boldsymbol{X} - \boldsymbol{Y}\right)\left(t\boldsymbol{X} + (1-t)\boldsymbol{Y}\right)^{-1}\right)$$
$$= -\mathrm{tr}\left(\left(\boldsymbol{X} - \boldsymbol{Y}\right)\boldsymbol{C}\left(\boldsymbol{X} - \boldsymbol{Y}\right)\boldsymbol{L}^{\mathrm{H}}\boldsymbol{L}\right)$$
$$= -\mathrm{tr}\left(\boldsymbol{L}\left(\boldsymbol{X} - \boldsymbol{Y}\right)\boldsymbol{C}\left(\boldsymbol{X} - \boldsymbol{Y}\right)^{\mathrm{H}}\boldsymbol{L}^{\mathrm{H}}\right), \quad \text{(F.7)}$$

where we introduced the matrices $\boldsymbol{C}$ and $\boldsymbol{L}$ such that

$$\boldsymbol{C} = \boldsymbol{L}^{\mathrm{H}}\boldsymbol{L} = \left(t\boldsymbol{X} + (1-t)\boldsymbol{Y}\right)^{-1}. \quad \text{(F.8)}$$

Since the matrix $\boldsymbol{C}$ is positive semidefinite, we can perform the Cholesky decomposition to obtain $\boldsymbol{L}$ (see Section C.5 of Appendix C). Note, moreover, that the matrix inside the

trace in (F.7) can be written in the form $\boldsymbol{BCB}^{\mathrm{H}}$. Therefore, due to the fact that the trace of a positive semidefinite matrix satisfies $-\operatorname{tr}\left(\boldsymbol{BCB}^{\mathrm{H}}\right) \leq 0$, we conclude that $g(t)$ is concave.

F.3. Euclidean Distance

Consider the problem formulation of the Euclidean distance minimization subject to certain constraints

$$\min_{\boldsymbol{\varrho}_k} \sum_{i=1}^{d_k} (\varrho_{k,i} - \varrho'_{k,i})^2 \quad \text{subject to} \quad \sum_{i=1}^{d_k} \varrho_{k,i} - \rho_k,\ \varrho_{k,i} \geq 0,\ \forall i. \tag{F.9}$$

Observe that the constraints are affine. Thus, the convexity of the problem depends on the Euclidean distance $\|\boldsymbol{\varrho} - \boldsymbol{\varrho}'\|_2^2$. Let us define the vector $\boldsymbol{a} = \boldsymbol{\varrho} - \boldsymbol{\varrho}' \in \mathbb{R}^{d_k}$. Therefore, we have to show that $f(\boldsymbol{a}) = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ is convex. To that end, we introduce the vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2 \in \mathbb{R}^{d_k}$ and the scalar $\lambda \in [0, 1]$. In such a way, we show that $f(\boldsymbol{a})$ satisfies F.1

$$\begin{aligned}
\lambda f(\boldsymbol{a}_1) + (1-\lambda) f(\boldsymbol{a}_2) &\geq f\left(\lambda \boldsymbol{a}_1 + (1-\lambda)\boldsymbol{a}_2\right) \\
\lambda \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_1 + (1-\lambda)\boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{a}_2 &\geq \left(\lambda \boldsymbol{a}_1^{\mathrm{T}} + (1-\lambda)\boldsymbol{a}_2^{\mathrm{T}}\right)\left(\lambda \boldsymbol{a}_1 + (1-\lambda)\boldsymbol{a}_2\right) \\
\lambda \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_1 + (1-\lambda)\boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{a}_2 &\geq \lambda^2 \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_1 + \lambda(1-\lambda)\left(\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_2 + \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{a}_1\right) \\
&\quad + (1-\lambda)^2 \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{a}_2 \\
(\lambda - \lambda^2)\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_1 + \left(1 - \lambda - \left(1-\lambda^2\right)\right)\boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{a}_2 &\geq \lambda(1-\lambda)\left(\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_2 + \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{a}_1\right) \\
\lambda(1-\lambda)\left(\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_1 + \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{a}_2\right) &\geq \lambda(1-\lambda)\left(\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_2 + \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{a}_1\right) \\
\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_1 + \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{a}_2 &\geq \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_2 + \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{a}_1.
\end{aligned} \tag{F.10}$$

Consequently, the Euclidean distance is convex and the minimization problem is also convex. Therefore, the KKT conditions D are sufficient for optimality.

Appendix G

List of Acronyms

AWGN Additive White Gaussian Noise

AO Alternate Optimization

BC Broadcast Channel

bps bits per second

BS Base Station

CDF Cumulative Distribution Function

CSI Channel State Information

CSIR Channel State Information at the Receiver

CSIT Channel State Information at the Transmitter

dB deciBels

DA Deterministic Annealing

DPC Dirty Paper Coding

DFT Discrete Fourier Transform

FDD Frequency-Division Duplex

FDMA Frequency-Division Multiple Access

3GPP Third Generation Partnership Project

GP Geometric Programing

HSPA High-Speed Packet Access

Hz Hertz

IDFT Inverse Discrete Fourier Transform

IF Interference Function

IFC Interference Channel

I-METRA Intelligent Multi-Element Transmit and Receive Antennas

ISI Inter-Symbol Interference

ITU International Telecommunication Union

KKT Karush-Kuhn-Tucker

LTE Long-Term Evolution

LTI Linear Time-Invariant

MAC Multiple Access Channel

MIMO Multiple-Input Multiple-Output

MISO Multiple-Input Single-Output

MU-MIMO Multiple-User Multiple-Input Multiple-Output

MU-MISO Multiple-User Multiple-Input Single-Output

MMSE Minimum Mean Square Error

MSE Mean Square Error

MU Multi-User

NLOS Non Line-of-Sight

OFDM Orthogonal Frequency Division Multiplexing

pdf probability density function

QoS Quality-of-Service

SAA Sample Averaging Approximation

SINR Signal to Interference-plus-Noise Ratio

SIR Signal to Interference Ratio

SISO Single-Input Single-Output

SIMO Single-Input Multiple-Output

SDP SemiDefinite Program

SNR Signal-to-Noise Ratio

SU-MIMO Single-User Multiple-Input Multiple-Output

TDD Time-Division Duplex

TDMA Time-Division Multiple Access

WiMAX Worldwide Interoperability for Microwave Access

References

[1] C. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, July 1948.

[2] E. Telatar, "Capacity of Multi-antenna Gaussian Channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[3] W. Yu and J. Cioffi, "Trellis precoding for the broadcast channel," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 2, November 2001, pp. 1344–1348.

[4] P. Viswanath and D. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912–1921, August 2003.

[5] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, Achievable Rates, and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2658–2668, October 2003.

[6] M. H. M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[7] M. Schubert and H. Boche, "Solution of the Multiuser Downlink Beamforming Problem with Individual SINR Constraints," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 1, pp. 18–28, January 2004.

[8] R. Hunger, M. Joham, and W. Utschick, "On the MSE Duality of the Broadcast Channel and the Multiple Access Channel," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 698–713, February 2009.

[9] N. Vucic and H. Boche, "Robust QoS-Constrained Optimization of Downlink Multiuser MISO Systems," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 714–725, February 2009.

[10] A. Mutapcic, S. Kim, and S. Boyd, "A Tractable Method for Robust Downlink Beamforming in Wireless Communications," in *Proc. Asilomar Conference on Signals, Systems and Computers (ACSSC)*, November 2007, pp. 1224–1228.

[11] N. Vucic and H. Boche, "Downlink precoding for multiuser MISO systems with imperfect channel knowledge," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2008, pp. 3121–3124.

[12] Y. Bin, L. YongJun, and H. Hanying, "Robust Transceiver Optimization for Downlink Multiuser MIMO Systems with Bounded Channel Uncertainties," in *Proc. Spring Congress on Engineering and Technology (S-CET)*, May 2012, pp. 1–6.

[13] V. Sharma and S. Lambotharan, "Robust Multiuser Downlink Beamforming with Per Antenna Power Constraints using Worst-Case Performance Optimization," in *Proc. International Symposium on Communications and Information Technologies (ISCIT)*, October 2008, pp. 5–8.

[14] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[15] M. B. Shenouda and T. N. Davidson, "On the Design of Linear Transceivers for Multiuser Systems with Channel Uncertainty," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 6, pp. 1015–1024, August 2008.

[16] P. Ubaidulla and A. Chockalingam, "Robust Transceiver Design for Multiuser MIMO Downlink," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, November 2008, pp. 1–5.

[17] T. Bogale and L. Vandendorpe, "Robust Sum MSE Optimization for Downlink Multiuser MIMO Systems With Arbitrary Power Constraint: Generalized Duality Approach," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1862–1875, April 2012.

[18] M. Kobayashi and G. Caire, "Joint Beamforming and Scheduling for a Multi-Antenna Downlink with Imperfect Transmitter Channel Knowledge," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1468–1477, September 2007.

[19] F. Negro, I. Ghauri, and D. T. M. Slock, "Sum Rate maximization in the noisy MIMO interfering broadcast channel with partial CSIT via the expected weighted MSE," in *Proc. International Symposium on Wireless Communication Systems (ISWCS)*, August 2012, pp. 576–580.

[20] M. Razaviyayn, M. Boroujeni, and Z.-Q. Luo, "A stochastic weighted MMSE approach to sum rate maximization for a MIMO interference channel," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2013, pp. 325–329.

[21] M.B. Shenouda and T.N. Davidson, "Probabilistically-constrained approaches to the design of the multiple antenna downlink," in *Proc. Asilomar Conference on Signals, Systems and Computers (ACSSC)*, October 2008, pp. 1120–1124.

[22] N. Vucic, H. Boche, and S. Shi, "Robust Transceiver Optimization in Downlink Multiuser MIMO Systems," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3576–3587, September 2009.

[23] J. Choi and C.-C. Lim, "Multiuser Downlink Beamforming with Limited Feedback," in *Proc. IFIP International Conference on Wireless and Optical Communications Networks (WOCN)*, July 2007, pp. 1–4.

[24] F. She, H. Luo, and W. Chen, "Low Complexity Multiuser Scheduling in MIMO Broadcast Channel with Limited Feedback," in *Proc. IEEE International Conference on Circuits and Systems for Communications (ICCSC)*, May 2008, pp. 125–129.

[25] F. Boccardi, H. Huang, and M. Trivellato, "Multiuser eigenmode transmission for mimo broadcast channels with limited feedback," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2007, pp. 1–5.

[26] G. Li, X. Zhang, X. Liu, and D. Yang, "Joint Combiner and Precoding in MU-MIMO Downlink Systems with Limited Feedback," in *Proc. IEEE Vehicular Technology Conference (VTC)*, September 2011, pp. 1–4.

[27] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO Achievable Rates With Downlink Training and Channel State Feedback," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2845–2866, June 2010.

[28] K. Wu, L. Wang, and L. Cai, "Joint Multiuser Precoding and Scheduling with Imperfect Channel State Information at the Transmitter," in *IEEE Vehicular Technology Conference (VTC)*, May 2008, pp. 265–269.

[29] M. Shenouda and T. Davidson, "Convex Conic Formulations of Robust Downlink Precoder Designs With Quality of Service Constraints," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 714–724, December 2007.

[30] ——, "Nonlinear and Linear Broadcasting With QoS Requirements: Tractable Approaches for Bounded Channel Uncertainties," *IEEE Transactions on Signal Processing*, vol. 57, no. 5, pp. 1936–1947, May 2009.

[31] T. Bogale, B. Chalise, and L. Vandendorpe, "Robust transceiver optimization for downlink multiuser mimo systems," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 446–453, January 2011.

[32] S. Shi, M. Schubert, and H. Boche, "Downlink MMSE Transceiver Optimization for Multiuser MIMO Systems: Duality and Sum-MSE Minimization," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5436 –5446, November 2007.

[33] ——, "Rate Optimization for Multiuser MIMO Systems With Linear Processing," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 4020–4030, August 2008.

[34] ——, "Downlink MMSE Transceiver Optimization for Multiuser MIMO Systems: MMSE Balancing," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3702–3712, August 2008.

[35] S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4792–4799, December 2008.

[36] J. Jensen, "Sur les fonctions convexes et les ingalits entre les valeurs moyennes," *Acta Mathematica*, vol. 30, no. 1, pp. 175–193, 1906.

[37] M. Joham, M. Vonbun, and W. Utschick, "MIMO BC/MAC MSE Duality with Imperfect Transmitter and Perfect Receiver CSI," in *Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2010, pp. 1–5.

[38] R. Yates, "A Framework for Uplink Power Control in Cellular Radio Systems," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1341–1347, September 1995.

[39] H. Boche and M. Schubert, "Multiuser Interference Balancing for General Interference Functions - A Convergence Analysis," in *Proc. IEEE International Conference on Communications (ICC)*, June 2007, pp. 4664–4669.

[40] R. Hunger and M. Joham, "A Complete Description of the QoS Feasibility Region in the Vector Broadcast Channel," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3870–3878, July 2010.

[41] W. Jakes, *Microwave mobile communications*. Wiley, 1974.

[42] J. Proakis, *Digital Communications*. McGraw-Hill, 2001.

[43] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[44] "IEEE Standard for Information technology– Local and metropolitan area networks– Specific requirements– Part 11: Wireless LAN Medium Access Control and Physical Layer Specifications Amendment 5: Enhancements for Higher Throughput," *IEEE Std 802.11n-2009*, pp. 1–565, Oct 2009.

[45] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[46] J. Proakis and M. Salehi, *Digital Communications*, ser. McGraw-Hill higher education. McGraw-Hill Education, 2007.

[47] H. Schulze and C. Lueders, *Theory and Applications of OFDM and CDMA: Wideband Wireless Communications*, ser. Wiley-Interscience online books. Wiley, 2005. [Online]. Available: http://books.google.es/books?id=PzWSwn9gHWEC

[48] J. Choi, *Adaptive and Iterative Signal Processing in Communications*. Cambridge University Press, 2006.

[49] P. Davis, *Circulant Matrices*, ser. Chelsea Publishing Series. Chelsea, 1994.

[50] T. M. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO Wireless Communications*. Cambridge University Press, 2007.

[51] F. A. Dietrich, W. Utschick, and P. Breun, "Linear Precoding Based on a Stochastic MSE Criterion," in *Proc. European Signal Processing Conference (EUSIPCO)*, September 2005.

[52] F. A. Dietrich, P. Breun, and W. Utschick, "Robust Tomlinson-Harashima Precoding for the Wireless Broadcast Channel," *IEEE Transactions on Signal Processing*, vol. 55, no. 2, pp. 631–644, February 2007.

[53] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.

[54] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[55] R. Esmailzadeh and M. Nakagawa, *TDD-CDMA for Wireless Communications*. Artech House, 2002.

[56] ——, "TDD-CDMA for the 4th Generation of Wireless Communications," *IEEE Communications Magazine*, vol. 41, no. 8, pp. 8–15, August 2003.

[57] W. Keusgen, C. M. Walke, and B. Rembold, "A System Model Considering the Influence of Front-End Imperfections on the Reciprocity of Up- and Downlink System Impulse Responses," in *Proc. Aachen Symposium on Signal Theory (ASST)*, September 2001, pp. 243–248.

[58] Philips, *Comparison Between MU-MIMO Codebook-based Channel Reporting Techniques for LTE Downlink*. 3GPP TSG RAN WG1, Tech. Rep. R1-062483, October 2006.

[59] S. Sesia, I. Toufik, and M. Baker, *LTE, The UMTS Long Term Evolution: From Theory to Practice*. John Wiley & Sons, 2009.

[60] J. G. Andrews, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*. Prentice Hall, 2007.

[61] Cover, T.M. and Thomas, J.A., *Elements of Information Theory*, ser. A Wiley-Interscience publication. Wiley, 2006.

[62] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," in *Proc. International Symposium on Information Theory (ISIT)*, June 2004, p. 174.

[63] R. Hunger and M. Joham, "A General Rate Duality of the MIMO Multiple Access Channel and the MIMO Broadcast Channel," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, November 2008, pp. 1–5.

[64] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2646–2660, June 2007.

[65] A. Berman and R. Plemmons, *Nonnegative matrices in the mathematical sciences* . New York: Academic Press, 1979.

[66] P. Viswanath and D. Tse, "Sum Capacity of the Vector Gaussian Broadcast Channel and Uplink-Downlink Duality," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912–1921, August 2003.

[67] G. Zheng, K. Wong, and B. Ottersten, "Robust Cognitive Beamforming With Bounded Channel Uncertainties," *IEEE Transactions on Signal Processing*, vol. 57, no. 12, pp. 4871–4881, December 2009.

[68] K. Wang, T. Chang, W. Ma, and C. Chi, "A Semidefinite Relaxation Based Conservative Approach to Robust Transmit Beamforming with Probabilistic SINR Constraints," in *Proc. European Signal Processing Conference (EUSIPCO 2010)*, August 2010, pp. 407–411.

[69] J. González-Coma, M. Joham, P. Castro, and L. Castedo, "Power minimization in the multiuser downlink under user rate constraints and imperfect transmitter CSI," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 4863–4867.

[70] M. Schubert and H. Boche, "A Generic Approach to QoS-Based Transceiver Optimization," *IEEE Transactions on Communications*, vol. 55, no. 8, pp. 1557–1566, August 2007.

[71] J. González-Coma, M. Joham, P. Castro, and L. Castedo, "Power minimization and QoS feasibility region in the multiuser MIMO broadcast channel with imperfect CSI," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2013, pp. 619–623.

[72] J. Kennan, "Uniqueness of positive fixed points for increasing concave functions on: An elementary result," *Rev. Economic Dynamics*, vol. 4, no. 4, pp. 893–899, 2001.

[73] J. Gubner, *Probability and Random Processes for Electrical and Computer Engineers*. Cambridge University Press, 2006.

[74] X. Huang and V. Lawrence, "Bandwidth-Efficient Bit and Power Loading for Underwater Acoustic OFDM Communication System with Limited Feedback," in *Proc. IEEE Vehicular Technology Conference (VTC)*, May 2011, pp. 1–5.

[75] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, March 1982.

[76] B. Khoshnevis and W. Yu, "Joint power control and beamforming codebook design for {MISO} channels under the outage criterion," *Journal of the Franklin Institute*, vol. 349, no. 1, pp. 140 – 161, 2012.

[77] A. Dabbagh and D. Love, "Precoding for Multiple Antenna Broadcast Channels with Channel Mismatch," in *Proc. Asilomar Conference on Signals, Systems and Computers (ACSSC)*, Oct 2006, pp. 1601–1605.

[78] ——, "Multiple antenna MMSE based downlink precoding with quantized feedback or channel mismatch," *Communications, IEEE Transactions on*, vol. 56, no. 11, pp. 1859–1868, November 2008.

[79] M. Joham, P. Castro, L. Castedo, and W. Utschick, "MMSE optimal feedback of correlated CSI for multi-user precoding," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2008, pp. 3129–3132.

[80] M. Joham, P. Castro, J. Zhen, L. Castedo, and W. Utschick, "Joint Design of limited feedback and multiuser precoding based on a precoding MSE metric," in *Proc. Asilomar Conference on Signals, Systems and Computers (ACSSC)*, October 2008, pp. 1305–1309.

[81] C. P.M., "Design of Limited Feedback for Robust MMSE Precoding in Multiuser MISO Systems," *Ph.D. Dissertation, GTEC, Universidade da Corua*, 2009.

[82] R. Agarwal and J. Cioffi, "Capacity of fading broadcast channels with limited-rate feedback," *Proc. Allerton Conference on Communication, Control, and Computing (ACCC)*, September 2006.

[83] C.-H. Wang and S. Dey, "Distortion outage minimization in Nakagami fading using limited feedback," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, p. 92, 2011. [Online]. Available: http://asp.eurasipjournals.com/content/2011/1/92

[84] M. Trivellato, H. Huang, and F. Boccardi, "Antenna Combining and Codebook Design for the MIMO Broadcast Channel with Limited Feedback," in *Proc. Asilomar Conference on Signals, Systems and Computers (ACSSC)*, Nov 2007, pp. 302–308.

[85] J. González-Coma, M. Joham, P. Castro, and L. Castedo, "Power minimization in the multiple stream MIMO Broadcast Channel with imperfect CSI," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, June 2014, pp. 165–168.

[86] P. Tejera, W. Utschick, J. Nossek, and G. Bauch, "Rate Balancing in Multiuser MIMO OFDM Systems," *IEEE Transactions on Communications*, vol. 57, no. 5, pp. 1370–1380, May 2009.

[87] M. Kobayashi and G. Caire, "A Practical Approach for Weighted Rate Sum Maximization in MIMO-OFDM Broadcast Channels," in *Proc. Asilomar Conference on Signals, Systems and Computers (ACSSC)*, November 2007, pp. 1591–1595.

[88] H. Karaa, R. Adve, and A. Tenenbaum, "Linear Precoding for Multiuser MIMO-OFDM Systems," in *Proc. IEEE International Conference on Communications (ICC)*, June 2007, pp. 2797–2802.

[89] J. Duplicy, J. Louveaux, and L. Vandendorpe, "Optimization of linear pre-decoders for multi-user closed-loop MIMO-OFDM," in *Proc. IEEE International Conference on Communications (ICC)*, vol. 3, May 2005, pp. 2026–2030.

[90] C. Hellings, W. Utschick, and M. Joham, "Power minimization in parallel vector broadcast channels with separate linear precoding," in *Proc. European Signal Processing Conference (EUSIPCO)*, August 2011, pp. 1834–1838.

[91] Y. Zhang and K. Letaief, "An efficient resource-allocation scheme for spatial multiuser access in MIMO/OFDM systems," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 107–116, January 2005.

[92] F. Rey, M. Lamarca, and G. Vazquez, "Robust power allocation algorithms for MIMO OFDM systems with imperfect CSI," *IEEE Transactions on Signal Processing*, vol. 53, no. 3, pp. 1070–1085, March 2005.

[93] Z. Hu, G. Zhu, Y. Xia, and G. Liu, "Multiuser subcarrier and bit allocation for MIMO-OFDM systems with perfect and partial channel information," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, vol. 2, March 2004, pp. 1188–1193 Vol.2.

[94] J. González-Coma, M. Joham, P. Castro, and L. Castedo, "Power Minimization in the Multiuser MIMO-OFDM Broadcast Channel with Imperfect CSI," in *Proc. European Signal Processing Conference (EUSIPCO)*, September 2014.

[95] Y. Cho, J. Kim, W. Yang, and C. Kang, *MIMO-OFDM Wireless Communications with MATLAB*. Wiley, 2010. [Online]. Available: http://books.google.es/books?id=6HwAoeuMr3kC

[96] B. Lu, G. Yue, and X. Wang, "Performance analysis and design optimization of LDPC-coded MIMO OFDM systems," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 348–361, February 2004.

[97] J. González-Coma, M. Joham, P. Castro, and L. Castedo, "Feedback and power optimization in the multiuser downlink under QoS constraints," in *Proc. International Symposium on Communications, Control and Signal Processing (ISCCSP)*, May 2014, pp. 412–415.

[98] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Kluwer, 1993.

[99] S. Peters and R. Heath, "Interference alignment via alternating minimization," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2009, pp. 2445–2448.

[100] I. Santamaria, O. Gonzalez, R. Heath, and S. Peters, "Maximum Sum-Rate Interference Alignment Algorithms for MIMO Channels," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, Dec 2010, pp. 1–6.

[101] L. Trefethen and D. Bau, *Numerical Linear Algebra*. Society for Industrial and Applied Mathematics, 1997.

[102] L. L. Scharf, *Statistical Signal Processing – Detection, Estimation, and Time Series Analysis*. Addison-Wesley, 1991.

[103] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

[104] K. Petersen and M. Pedersen, *The Matrix Cookbook* , November 2008. [Online]. Available: www.matrixcookbook.com

[105] M. Brookes, *The Matrix Reference Manual*, 2011. [Online]. Available: http://www.ee.imperial.ac.uk/hp/staff/dmb/matrix/intro.html

[106] W. Karush, *Minima of Functions of Several Variables with Inequalities as Side Conditions. M.S. Thesis*. The University of Chicago, 1939.

[107] H. W. Kuhn and A. W. Tucker, "Nonlinear Programming," in *Proc. Second Berkeley Symposium on Mathematical Statistics and Probability*. J. Neyman, University of California Press, 1951, pp. 481–492.

[108] R. Fletcher, *Practical Methods of Optimization*. John Wiley & Sons, 1967.

[109] D. G. Luenberger, *Linear and Nonlinear Programming*. Addison-Wesley, 1989.