Space-Time Processing
for MIMO Communications

Editors:

Alex B. Gershman
Dept. of ECE, McMaster University
Hamilton, L8S 4K1, Ontario, Canada;
& Dept. of Communication Systems, Duisburg-Essen
University, Germany

and

Nicholas D. Sidiropoulos
Dept. of ECE, Technical University of Crete
Chania - Crete, 73100 Greece;
& Dept. of ECE and Digital Technology Center,
University of Minnesota, Minneapolis, U.S.A.

Contents

2

Convex Optimization Theory Applied to Joint Transmitter-Receiver Design in MIMO Channels

Daniel Pérez Palomar,[1] Antonio Pascual-Iserte,[2] John M. Cioffi,[3] and Miguel Angel Lagunas[2,4]

[1]Princeton University, [2]Technical University of Catalonia, [3]Stanford University, and [4]Telecommunications Technological Center of Catalonia

2.1   Introduction

Multi-antenna MIMO channels have recently become a popular means to increase the spectral efficiency and quality of wireless communications by the use of spatial diversity at both sides of the link [1, 2, 3, 4]. In fact, the MIMO concept is much more general and embraces many other scenarios such as wireline digital subscriber line (DSL) systems [5] and single-antenna frequency-selective channels [6]. This general modeling of a channel as an abstract MIMO channel allows for a unified treatment using a compact and convenient vector-matrix notation.

MIMO systems are not just mathematically more involved than SISO systems but also conceptually different and more complicated, since several substreams are typically established in MIMO channels (multiplexing property) [7]. The existence of several substreams, each with its own quality, makes the definition of a global measure of the system quality very difficult; as a consequence, a variety of design criteria have been adopted in the literature. In fact, the design of such systems is a multi-objective optimization problem characterized by not having just optimal solutions (as happens in single-objective optimization problems) but a set of Pareto-optimal solutions [8].

The fundamental limits of MIMO communications have been known since 1948, when Shannon, in his ground-breaking paper [9], defined the concept of channel capacity–the maximum reliably achievable rate–and obtained the capacity-achieving signaling strategy. For a given realization of a MIMO channel, such a theoretical limit can be achieved by a Gaussian signaling with a waterfilling power profile over the channel eigenmodes [10, 3, 2]. In real systems, however, rather than with Gaussian codes, the transmission is done with practical signal constellations and coding schemes. To simplify the analysis and design of such systems, it is convenient to divide them into an uncoded part, which transmits symbols drawn from some constellations, and a coded part that builds upon the uncoded system. It is important to bear in mind that the ultimate system performance depends on the combination of both parts (in fact, for some systems, such a division does not apply).

The signaling scheme in a MIMO channel depends on the quantity and the quality of the channel state information (CSI) available at both sides of the communication link. For the case of no CSI at the transmitter, a wide family of techniques—termed space-time coding—have been proposed in the literature [1, 11, 12]. The focus of this chapter is on communication systems with CSI, either perfect (i.e., sufficiently good) or imperfect, and, more specifically, on the design of the uncoded part of the system in the form of linear MIMO transceivers (or transmit-receive beamforming strategies) under the framework of convex optimization theory.

In the last two decades, a number of fundamental and practical results have been obtained in convex optimization theory [13, 14]. It is a well-developed area both in the theoretical and practical aspects. Many convex problems, for example, can be analytically studied and solved using the optimality conditions derived from Lagrange duality theory. In any case, a convex problem can always be solved in practice with very efficient algorithms such as interior-point methods [14]. The engineering community has greatly benefited from these recent advances by finding applications. This chapter describes another application in the design of beamforming for MIMO channels.

This chapter starts with a brief overview of convex optimization theory in Section 2.2, with special emphasis on the art of unveiling the hidden convexity of problems (illustrated with some recent examples). Then, after introducing the system model in Section 2.3, the design of linear MIMO transceivers is considered in Section 2.4 under the framework of convex optimization theory. The derivation of optimal designs focuses on how the originally nonconvex problem is reformulated in convex form, culminating with closed-form expressions obtained from the optimality conditions. This work presents in a unified fashion the results obtained in [15] and [16] (see also [17]) and generalizes some of the results as well. The practical problem of imperfect

CSI is addressed in Section 2.5, deriving a robust design less sensitive to errors in the CSI.

Notation: Boldface upper-case letters denote matrices, boldface lower-case letters denote column vectors, and italics denote scalars. The super-scripts $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$ denote transpose, complex conjugate, and Hermitian operations, respectively. $\mathbb{R}^{m \times n}$ and $\mathbb{C}^{m \times n}$ represent the set of $m \times n$ matrices with real- and complex-valued entries, respectively (the subscript $_+$ is sometimes used to restrict the elements to nonnegative values). $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ denote the real and imaginary part, respectively. $\mathrm{Tr}(\cdot)$ and $\det(\cdot)$ denote the trace and determinant of a matrix, respectively. $\|\mathbf{x}\|$ is the Euclidean norm of a vector $\mathbf{x}$ and $\|\mathbf{X}\|_F$ is the Frobenius norm of a matrix $\mathbf{X}$ (defined as $\|\mathbf{X}\|_F \triangleq \sqrt{\mathrm{Tr}(\mathbf{X}^H \mathbf{X})}$). $[\mathbf{X}]_{i,j}$ (also $[\mathbf{X}]_{ij}$) denotes the (ith, jth) element of matrix $\mathbf{X}$. $\mathbf{d}(\mathbf{X})$ and $\boldsymbol{\lambda}(\mathbf{X})$ denote the diagonal elements and eigenvalues, respectively, of matrix $\mathbf{X}$. A block-diagonal matrix with diagonal blocks given by the set $\{\mathbf{X}_k\}$ is denoted by $\mathrm{diag}(\{\mathbf{X}_k\})$. The operator $(x)^+ \triangleq \max(0, x)$ is the projection on the nonnegative orthant.

2.2 Convex Optimization Theory

In the last two decades, several fundamental and practical results have been obtained in convex optimization theory [13, 14]. The engineering community not only has benefited from these recent advances by finding applications, but has also fueled the mathematical development of both the theory and efficient algorithms. The two classical mathematical references on the subject are [18] and [19]. Two more recent engineering-oriented excellent references are [13] and [14].

Traditionally, it was a common believe that linear problems were easy to solve as opposed to nonlinear problems. However, as stated by Rockafellar in a 1993 survey [20], "the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity" [14]. In a nutshell, convex problems can always be solved optimally either in closed form (by means of the optimality conditions derived from Lagrange duality) or numerically (with very efficient algorithms that exhibit a polynomial convergence). As a consequence, roughly speaking, one can say that once a problem has been expressed in convex form, it has been solved.

Unfortunately, most engineering problems are not convex when directly formulated. However, many of them have a potential hidden convexity that engineers have to unveil in order to be able to use all the machinery of convex optimization theory.

This section introduces the basic ideas of convex optimization (both the theory and practice) and then focuses on the art of reformulating engineering problems in convex form by means of recent real examples.

2.2.1 Definitions and Classes of Convex Problems

Basic Definitions

An optimization problem with arbitrary equality and inequality constraints can always be written in the following standard form [14]:

$$
\begin{array}{ll}
\underset{\mathbf{x}}{\text{minimize}} & f_0(\mathbf{x}) \\
\text{subject to} & f_i(\mathbf{x}) \le 0 \qquad 1 \le i \le m, \\
& h_i(\mathbf{x}) = 0 \qquad 1 \le i \le p,
\end{array}
\tag{2.1}
$$

where $\mathbf{x} \in \mathbb{R}^n$ is the optimization variable, f_0 is the *cost* or *objective function*, $f_1, \cdots, f_m$ are the m inequality constraint functions, and $h_1, \cdots, h_p$ are the p equality constraint functions.

If the objective and inequality constraint functions are convex[1] and the equality constraint functions are linear (or, more generally, affine), the problem is then a *convex optimization problem* (or *convex program*). A point $\mathbf{x}$ in the domain of the problem (set of points for which the objective and all constraint functions are defined) is *feasible* if it satisfies all the constraints $f_i(\mathbf{x}) \le 0$ and $h_i(\mathbf{x}) = 0$. The problem (2.1) is said to be feasible if there exists at least one feasible point and *infeasible* otherwise. The *optimal value* (minimal value) is denoted by $f^\star$ and is achieved at an optimal solution $\mathbf{x}^\star$, i.e., $f^\star = f_0(\mathbf{x}^\star)$.

Classes of Convex Problems

When the functions f_i and h_i in (2.1) are linear (affine), the problem is called a *linear program* (LP) and is much simpler to solve. If the objective function is quadratic and the constraint functions are linear (affine), then it is called a *quadratic program* (QP); if, in addition, the inequality constraints are also quadratic, it is called *quadratically constrained quadratic program* (QCQP). QPs include LPs as special case.

A problem that is closely related to quadratic programming is the *second-order cone program* (SOCP) [21, 14] that includes constraints of the form

$$
\|\mathbf{A}\mathbf{x} + \mathbf{b}\| \le \mathbf{c}^T \mathbf{x} + d
\tag{2.2}
$$

where $\mathbf{A} \in \mathbb{R}^{k \times n}$, $\mathbf{b} \in \mathbb{R}^k$, $\mathbf{c} \in \mathbb{R}^n$, and $d \in \mathbb{R}$ are given and fixed. Note that (2.2) defines a convex set because it is an affine transformation of the second-order cone $\mathcal{C}^n = \{(\mathbf{x}, t) \in \mathbb{R}^n \mid \|\mathbf{x}\| \le t\}$, which is convex since both $\|\mathbf{x}\|$ and $-t$ are convex. If $\mathbf{c} = \mathbf{0}$, then (2.2) reduced to a quadratic constraint (by squaring both sides).

A more general problem than an SOCP is a *semidefinite program* (SDP) [22, 14] that has matrix inequality constraints of the form

$$
x_1 \mathbf{F}_1 + \ldots + x_n \mathbf{F}_n + \mathbf{G} \le \mathbf{0}
\tag{2.3}
$$

where $\mathbf{F}_1, \cdots, \mathbf{F}_n, \mathbf{G} \in \mathcal{S}^k$ ($\mathcal{S}^k$ is the set of Hermitian $k \times k$ matrices) and $\mathbf{A} \ge \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semidefinite.

---

[1]A function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is convex if, for all $\mathbf{x}, \mathbf{y} \in \operatorname{dom} f$ and $\theta \in [0, 1]$, $\theta \mathbf{x} + (1 - \theta)\mathbf{y} \in \operatorname{dom} f$ (i.e., the domain is a convex set) and $f(\theta \mathbf{x} + (1 - \theta)\mathbf{y}) \le \theta f(\mathbf{x}) + (1 - \theta)f(\mathbf{y})$.

A very useful generalization of the standard convex optimization problem (2.1) is obtained by allowing the inequality constraints to be vector valued and using generalized inequalities [14]:

$$
\begin{aligned}
\underset{\mathbf{x}}{\text{minimize}} \quad & f_0(\mathbf{x}) \\
\text{subject to} \quad & \mathbf{f}_i(\mathbf{x}) \preceq_{\mathcal{K}_i} \mathbf{0} \qquad 1 \le i \le m, \\
& \mathbf{h}_i(\mathbf{x}) = \mathbf{0} \qquad 1 \le i \le p,
\end{aligned} \tag{2.4}
$$

where the generalized inequalities[2] $\preceq_{\mathcal{K}_i}$ are defined by the proper cones $\mathcal{K}_i$ ($\mathbf{a} \preceq_{\mathcal{K}} \mathbf{b} \Leftrightarrow \mathbf{b} - \mathbf{a} \in \mathcal{K}$) [14] and f_i are $\mathcal{K}_i$-convex.[3]

Among the simplest convex optimization problems with generalized inequalities are the *cone programs* (CP) (or *conic form problems*), which have a linear objective and one inequality constraint function [23, 14]:

$$
\begin{aligned}
\underset{\mathbf{x}}{\text{minimize}} \quad & \mathbf{c}^T \mathbf{x} \\
\text{subject to} \quad & \mathbf{F}\mathbf{x} + \mathbf{g} \preceq_{\mathcal{K}} \mathbf{0} \\
& \mathbf{A}\mathbf{x} = \mathbf{b}.
\end{aligned} \tag{2.5}
$$

CPs particularize nicely to LPs, SOCPs, and SDPs as follows: i) if $\mathcal{K} = \mathbb{R}^n_+$ (nonnegative orthant), the partial ordering $\preceq_{\mathcal{K}}$ is the usual componentwise inequality between vectors and (2.5) reduces to an LP; ii) if $\mathcal{K} = \mathcal{C}^n$ (second-order cone), $\preceq_{\mathcal{K}}$ corresponds to a constraint of the form (2.2) and the problem (2.5) becomes an SOCP; iii) if $\mathcal{K} = \mathcal{S}^n_+$ (positive semidefinite cone), the generalized inequality $\preceq_{\mathcal{K}}$ reduces to the usual matrix inequality as in (2.3) and the problem (2.5) simplifies to an SDP.

There is yet another very interesting and useful class of problems, the family of *geometric programs* (GP), that are not convex in their natural form but can be transformed into convex problems [14].

2.2.2 Reformulating a Problem in Convex Form

As has been previously said, convex problems can always be solved in practice, either in closed form or numerically. However, the natural formulation of most engineering problems is not convex. In many cases, fortunately, there is a hidden convexity that can be unveiled by properly reformulating the problem. The main task of an engineer is then to cast the problem in convex form and, if possible, in any of the well-known classes of convex problems (so that specific and optimized algorithms can be used).

Unfortunately, there is not a systematic way to reformulate a problem in convex form. In fact, it is rather an art that can only be learned by examples (see §2.2.5). There are two main ways to reformulate a problem in convex form. The main one is to devise a convex problem equivalent to the original nonconvex one by using a series of clever changes of variables. As an example, consider the minimization of $1/\left(1 + x^2\right)$ subject to $x^2 \ge 1$, which is a nonconvex problem (both the cost function and the

---

[2]A generalized inequality is a partial ordering on $\mathbb{R}^n$ that has many of the properties of the standard ordering on $\mathbb{R}$. A common example is the matrix inequality defined by the cone of positive semidefinite $n \times n$ matrices $\mathcal{S}^n_+$.

[3]A function $\mathbf{f} : \mathbb{R}^n \longrightarrow \mathbb{R}^{k_i}$ is $\mathcal{K}_i$-convex if the domain is a convex set and, for all $\mathbf{x}, \mathbf{y} \in \operatorname{dom} f$ and $\theta \in [0, 1]$, $\mathbf{f}\left(\theta\mathbf{x} + (1 - \theta)\mathbf{y}\right) \preceq_{\mathcal{K}_i} \theta\mathbf{f}\left(\mathbf{x}\right) + (1 - \theta)\mathbf{f}\left(\mathbf{y}\right)$.

constraint are nonconvex). The problem can be rewritten in convex form, after the change of variable $y = x^2$, as the minimization of $1/(1+y)$ subject to $y \geq 1$ (and the optimal x can be recovered from the optimal y as $x = \sqrt{y}$). A more realistic example is briefly described in §2.2.5 for robust beamforming. The class of geometric problems is a very important example of nonconvex problems that can be reformulated in convex form by a change of variable [14]. Another example is the beamforming design for MIMO channels treated in detail in §2.4.

Nevertheless, it is not really necessary to devise a convex problem that is exactly equivalent to the original one. In fact, it suffices if they both have the same set of optimal solutions (related by some mapping). In other words, both problems have to be equivalent only within the set of optimal solutions but not otherwise. Of course, the difficulty is how to obtain such a magic convex problem without knowing beforehand the set of optimal solutions. One very popular way to do this is by relaxing the problem (removing some of the constraints) such that it becomes convex, in a way that the "relaxed" optimal solutions can be shown to satisfy the removed constraints as well. A remarkable example of this approach is described in §2.2.5 for multiuser beamforming. Several relaxations are also employed in the beamforming design for MIMO channels in §2.4.

2.2.3 Lagrange Duality Theory and KKT Optimality Conditions

Lagrange duality theory is a very rich and mature theory that links the original minimization problem (2.1), termed primal problem, with a dual maximization problem. In some occasions, it is simpler to solve the dual problem than the primal one. A fundamental result in duality theory is given by the optimality Karush-Kuhn-Tucker (KKT) conditions that any primal-dual solution must satisfy. By exploring the KKT conditions, it is possible in many cases to obtain a closed-form solution to the original problem (see, for example, the iterative waterfilling described in §2.2.5 and the closed-form results obtained in §2.4 for MIMO beamforming). In the following, the basic results on duality theory including the KKT conditions are stated (for details, the reader is referred to [13, 14]).

The basic idea in Lagrange duality is to take the constraints of (2.1) into account by augmenting the objective function with a weighted sum of the constraint functions. The *Lagrangian* of (2.1) is defined as

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \qquad (2.6)$$

where λ_i and ν_i are the *Lagrange multipliers* associated with the ith inequality constraint $f_i(\mathbf{x}) \leq 0$ and with the ith equality constraint $h_i(\mathbf{x}) = 0$, respectively.

The optimization variable $\mathbf{x}$ is called the *primal variable* and the Lagrange multipliers $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ are also termed the *dual variables*. The original objective function $f_0(\mathbf{x})$ is referred to as the *primal objective*, whereas the *dual objective* $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is defined as the minimum value of the Lagrangian over $\mathbf{x}$:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}), \qquad (2.7)$$

which is concave even if the original problem is not convex because it is the pointwise infimum of a family of affine functions of $(\boldsymbol{\lambda}, \boldsymbol{\nu})$. Note that the infimum in (2.7) is with respect all $\mathbf{x}$ (not necessarily feasible points). The dual variables $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ are *dual feasible* if $\boldsymbol{\lambda} \geq \mathbf{0}$.

It turns out that the primal and dual objectives satisfy $f_0(\mathbf{x}) \geq g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ for any feasible $\mathbf{x}$ and $(\boldsymbol{\lambda}, \boldsymbol{\nu})$. Therefore, it makes sense to maximize the dual function to obtain a lower bound on the optimal value $f^\star$ of the original problem (2.1):

$$\begin{array}{cl} \underset{\boldsymbol{\lambda}, \boldsymbol{\nu}}{\text{maximize}} & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{subject to} & \boldsymbol{\lambda} \geq \mathbf{0}, \end{array} \qquad (2.8)$$

which is always a convex optimization problem even if the original problem is not convex. It is interesting to point out that a primal-dual feasible pair $(\mathbf{x}, (\boldsymbol{\lambda}, \boldsymbol{\nu}))$ localizes the optimal value of the primal (and dual) problem in an interval:

$$f^\star \in [g(\boldsymbol{\lambda}, \boldsymbol{\nu}), f_0(\mathbf{x})]. \qquad (2.9)$$

This property can be used in optimization algorithms to provide nonheuristic stopping criteria.

The difference between the optimal primal objective $f^\star$ and the optimal dual objective $g^\star$ is called the *duality gap,* which is always nonnegative $f^\star - g^\star \geq 0$ (weak duality). A central result in convex analysis [19, 18, 13, 14] is that when the problem is convex, under some mild technical conditions (called constraint qualifications[4]), the duality gap reduces to zero at the optimal (i.e., strong duality holds). Hence, the primal problem (2.1) can be equivalently solved by solving the dual problem (2.8) (see, for example, the simultaneous routing and resource allocation described in §2.2.5).

The optimal solutions of the primal and dual problems, $\mathbf{x}^\star$ and $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$, respectively, are linked together through the KKT conditions:

$$h_i(\mathbf{x}^\star) = 0, \qquad f_i(\mathbf{x}^\star) \leq 0, \qquad (2.10)$$

$$\lambda_i^\star \geq 0, \qquad (2.11)$$

$$\nabla_\mathbf{x} f_0(\mathbf{x}^\star) + \sum_{i=1}^m \lambda_i^\star \nabla_\mathbf{x} f_i(\mathbf{x}^\star) + \sum_{i=1}^p \nu_i^\star \nabla_\mathbf{x} h_i(\mathbf{x}^\star) = \mathbf{0}, \qquad (2.12)$$

$$(\text{complementary slackness}) \quad \lambda_i^\star f_i(\mathbf{x}^\star) = 0. \qquad (2.13)$$

The KKT conditions are necessary and sufficient for optimality (when strong duality holds) [13, 14]. Hence, if they can be solved, both the primal and dual problems are implicitly solved.

2.2.4 Efficient Numerical Algorithms to Solve Convex Problems

During the last decade, there has been a tremendous advance in developing efficient algorithms for solving wide classes of convex optimization problems. The most recent

---

[4]One simple version of the constraint qualifications is Slater's condition, which is satisfied when the problem is strictly feasible (i.e., when there exists $\mathbf{x}$ such that $f_i(\mathbf{x}) < 0$ for $1 \leq i \leq m$ and $h_i(\mathbf{x}) = 0$ for $1 \leq i \leq p$) [13, 14].

breakthrough in convex optimization theory is probably the development of interior-point methods for nonlinear convex problems. This was well established by Nesterov and Nemirovski in 1994 [24], where they extended the theory of linear programming interior-point methods (Karmarkar, 1984) to nonlinear convex optimization problems (based on the convergence theory of Newton's method for self-concordant functions).

The traditional optimization methods are based on gradient descent algorithms, which suffer from slow convergence and sensitivity to the algorithm initialization and stepsize selection. The recently developed methods for convex problems enjoy excellent convergence properties (polynomial convergence) and do not suffer from the usual problems of the traditional methods. In addition, it is simple to employ nonheuristic stopping criteria based on a desired resolution, since the difference between the cost value at each iteration and the optimum value can be upper-bounded using duality theory as in (2.9) [13, 14].

Many different software implementations have been recently developed and many of them are publicly available for free. It is worth pointing out that the existing packages not only provide the optimal primal variables of the problem but also the optimal dual variables. Currently, one of the most popular software optimization packages is SeDuMi [25], which is a Matlab toolbox for solving optimization problems over symmetric cones.

In the following, the most common optimization methods are briefly described with emphasis in interior-point methods.

Interior-Point Methods

Interior-point methods solve constrained problems by solving a sequence of smooth (continuous second derivatives are assumed) unconstrained problems, usually using Newton's method [13, 14]. The solutions at each iteration are all strictly feasible (they are in the interior of the domain), hence the name interior-point method. They are also called *barrier methods* since at each iteration a barrier function is used to guarantee that the obtained solution is strictly feasible.

Suppose that the following problem is to be solved:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0 \qquad 1 \leq i \leq m. \end{aligned} \qquad (2.14)$$

(Note that equality constraints can always be eliminated by a reparameterization of the affine feasible set.[5]) An interior-point method is easily implemented, for example, by forming the logarithmic barrier $\phi(\mathbf{x}) = -\sum_i \log(-f_i(\mathbf{x}))$, which is defined only on the feasible set and tends to $+\infty$ as any of the constraint functions goes to 0. At this point, the function $f_0(\mathbf{x}) + \frac{1}{t}\phi(\mathbf{x})$ can be easily minimized for a given t since it is an unconstrained minimization, obtaining the solution $x^\star(t)$, which of course is only an approximation of the solution to the original problem $x^\star$. Interestingly, $x^\star(t)$ as a function of t describes a curve called the central path, with the property that

---

[5]The set $\{\mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b}\}$ is equal to $\{\mathbf{F}\mathbf{z} + \mathbf{x}_0\}$, where $\mathbf{x}_0$ is any solution that satisfies the constraints $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$ and $\mathbf{F}$ is any matrix whose range is the nullspace of $\mathbf{A}$. Hence, instead of minimizing $f(\mathbf{x})$ subject to the equality constraints, one can equivalently minimize the function $\tilde{f}(\mathbf{z}) = f(\mathbf{F}\mathbf{z} + \mathbf{x}_0)$ with no equality constraints [14].

$x^{\star}(t) \to x^{\star}$ as $t \to \infty$. In practice, instead of choosing a large value of t and solving the approximated unconstrained problem (which would be very difficult to minimize since its Hessian would vary rapidly near the boundary of the feasible set), it is much more convenient to start with a small value of t and successively increase it (this way, the unconstrained minimization for some t can use as a starting point the optimal solution obtained in the previous unconstrained minimization). Note that it is not necessary to compute $x^{\star}(t)$ exactly since the central path has no significance beyond the fact that it leads to a solution of the original problem as $t \to \infty$.

It can be shown from a worst-case complexity analysis that the total number of Newton steps grows as $\sqrt{m}$ (polynomial complexity), although in practice this number is between 10 and 50 iterations [14].

Cutting-Plane and Ellipsoid Methods

Cutting-plane methods are based on a completely different philosophy and do not require differentiability of the objective and constraint functions [26, 13]. They start with the feasible space and iteratively divide it into two halfspaces to reject the one that is known not to contain any optimal solution. Ellipsoid methods are related to cutting-plane methods in that they sequentially reduce an ellipsoid known to contain an optimal solution [26]. In general, cutting-plane methods are less efficient for problems to which interior-point methods apply.

Primal-Dual Interior-Point Methods

Primal-dual interior-point methods are similar to (primal) interior-point methods in the sense that they follow the central path, but they are more sophisticated since they solve the primal and dual linear programs simultaneously by generating iterates of the primal-dual variables [13, 14]. For several basic problem classes, such as linear, quadratic, second-order cone, geometric, and semidefinite programming, customized primal-dual methods outperform the barrier method. For general nonlinear convex optimization problems, primal-dual interior-point methods are still a topic of active research, but show great promise.

2.2.5   Applications in Signal Processing and Communications

The number of applications of convex optimization theory has exploded in the last eight years. An excellent source of examples and applications is [14] (see also [27] for an overview of recent applications).

The following is a nonexhaustive list of several illustrative recent results that make a strong use of convex optimization theory, with special emphasis on examples that have successfully managed to reformulate nonconvex problems in convex form.

Filter/Beamforming Design

The design of finite impulse response (FIR) filters and, similarly, of antenna array weighting (beamforming), have greatly benefited from convex optimization theory. Some examples are: [28], where the design of the antenna array weighting to satisfy

some specifications in different directions is formulated as an SOCP; [29], where the design of FIR filters subject to upper and lower bounds on the discrete-frequency response magnitude is formulated in convex form using a change of variables and spectral factorization; and [30], where FIR filters are designed enforcing piecewise constant and piecewise trigonometric polynomial masks in a finite and convex manner via linear matrix inequalities.

Worst-Case Robust Beamforming

A classical approach to design receive beamforming is the Capon's method, also termed minimum variance distortionless response (MVDR) beamformer [31]. Capon's method obtains the beamvector $\mathbf{w}$ as the minimization of the weighted array output power $\mathbf{w}^H \mathbf{R} \mathbf{w}$ subject to a unity-gain constraint in the desired look direction $\mathbf{w}^H \mathbf{s} = 1$, where $\mathbf{R}$ is the covariance matrix of the received signal and $\mathbf{s}$ is the steering vector of the desired signal. Under ideal conditions, this design maximizes the signal to interference-plus-noise ratio (SINR). However, a slight mismatch between the presumed and actual steering vectors, $\hat{\mathbf{s}}$ and $\mathbf{s}$, respectively, can cause a severe performance degradation. Therefore, robust approaches to adaptive beamforming are needed.

A worst-case robust approach essentially models the estimated parameters with an uncertainty region [32, 33, 34, 35]. As formulated in [33], an effective worst-case robust design is obtained by considering that the actual steering vector is close to the estimated one $\mathbf{s} = \hat{\mathbf{s}} + \mathbf{e}$, where $\mathbf{e}$ is an error vector with bounded norm $\|\mathbf{e}\| \leq \varepsilon$ that describes the uncertainty region (more general uncertainty regions and different formulations were considered in [32, 34]). The robust formulation can be formulated by imposing a good response along all directions in the uncertainty region:

$$
\begin{aligned}
&\underset{\mathbf{w}}{\text{minimize}} \quad \mathbf{w}^H \mathbf{R} \mathbf{w} \\
&\text{subject to} \quad \left| \mathbf{w}^H \mathbf{c} \right| \geq 1 \quad \forall \mathbf{c} = \hat{\mathbf{s}} + \mathbf{e}, \ \|\mathbf{e}\| \leq \varepsilon.
\end{aligned}
\tag{2.15}
$$

Such a problem is a semi-infinite nonconvex quadratic problem which needs to be simplified. The semi-infinite set of constraints can be replaced by the single constraint $\min_{\|\mathbf{e}\| \leq \varepsilon} \left| \mathbf{w}^H (\hat{\mathbf{s}} + \mathbf{e}) \right| \geq 1$ and then, by applying the triangle and Cauchy-Schwarz inequalities along with $\|\mathbf{e}\| \leq \varepsilon$, the following is obtained:

$$
|\mathbf{w}^H \hat{\mathbf{s}} + \mathbf{w}^H \mathbf{e}| \geq |\mathbf{w}^H \hat{\mathbf{s}}| - |\mathbf{w}^H \mathbf{e}| \geq |\mathbf{w}^H \hat{\mathbf{s}}| - \varepsilon \|\mathbf{w}\|
\tag{2.16}
$$

where the lower bound is indeed achieved if $\mathbf{e}$ is proportional to $\mathbf{w}$ with a phase such that $\mathbf{w}^H \mathbf{e}$ has opposite direction as $\mathbf{w}^H \hat{\mathbf{s}}$ [33]. Now, since $\mathbf{w}$ admits any arbitrary rotation without affecting the problem, $\mathbf{w}^H \hat{\mathbf{s}}$ can be forced to be real and nonnegative. The problem can be finally formulated in convex form

$$
\begin{aligned}
&\underset{\mathbf{w}}{\text{minimize}} \quad \mathbf{w}^H \mathbf{R} \mathbf{w} \\
&\text{subject to} \quad \mathbf{w}^H \hat{\mathbf{s}} \geq 1 + \varepsilon \|\mathbf{w}\| \\
&\qquad\qquad\ \ \text{Im} \left\{ \mathbf{w}^H \hat{\mathbf{s}} \right\} = 0.
\end{aligned}
\tag{2.17}
$$

In addition, the problem can be further manipulated to be expressed as an SOCP [33].

This type of worst-case robust formulation generally leads to a diagonal loading on $\mathbf{R}$ [32, 33, 34, 35], where the loading factor is optimally calculated as opposed that the more traditional ad-hoc techniques (the computation of the diagonal loading was explicitly characterized in [34] in a simple form). A similar problem was considered in [35] for a general-rank signal model, i.e., by considering the constraint $\mathbf{w}^H \mathbf{R}_s \mathbf{w} = 1$, where $\mathbf{R}_s$ is the covariance matrix of the desired signal with arbitrary rank (in the previous case, $\mathbf{R}_s = \mathbf{s}\mathbf{s}^H$ which is rank one).

Multiuser Beamforming

Beamforming for transmission in a wireless network was addressed in [36] within a convex optimization framework for an scenario with multi-antenna base stations transmitting simultaneously to several single-antenna users. The design problem can be formulated as the minimization of the total transmitted power subject to independent SINR constraints on each user:

$$
\begin{aligned}
&\underset{\{\mathbf{w}_k\}}{\text{minimize}} && \sum_{k=1}^K \mathbf{w}_k^H \mathbf{w}_k \\
&\text{subject to} && \frac{\mathbf{w}_k^H \mathbf{R}_{k,\beta(k)} \mathbf{w}_k}{\sum_{l \neq k} \mathbf{w}_l^H \mathbf{R}_{k,\beta(l)} \mathbf{w}_l + \sigma_k^2} \geq \gamma_k \qquad 1 \leq k \leq K
\end{aligned}
\tag{2.18}
$$

where K is the total number of users and, for the kth user, $\beta(k)$ is the corresponding base station, $\mathbf{w}_k$ is the beamvector, γ_k is the minimum required SINR, $\mathbf{R}_{k,\beta(l)} = \mathbb{E}\left[\mathbf{h}_{k,\beta(l)} \mathbf{h}_{k,\beta(l)}^H\right]$ is the channel correlation matrix of the downlink channel $\mathbf{h}_{k,\beta(l)}$ between the base station $\beta(l)$ and the user k, and σ_k^2 is the noise power. This problem can be easily written as a quadratic optimization problem but with quadratic nonconvex constraints.

The problem can be reformulated in convex form as an SDP by defining the change of variable $\mathbf{W}_k = \mathbf{w}_k \mathbf{w}_k^H$:

$$
\begin{aligned}
&\underset{\{\mathbf{W}_k\}}{\text{minimize}} && \sum_{k=1}^K \text{Tr}(\mathbf{W}_k) \\
&\text{subject to} && \text{Tr}\left(\mathbf{R}_{k,\beta(k)} \mathbf{W}_k\right) - \gamma_k \sum_{l \neq k} \text{Tr}\left(\mathbf{R}_{k,\beta(l)} \mathbf{W}_l\right) \geq \gamma_k \sigma_k^2 \qquad 1 \leq k \leq K \\
& && \mathbf{W}_k = \mathbf{W}_k^H \geq 0.
\end{aligned}
\tag{2.19}
$$

This problem, however, is a relaxation of the original one since it lacks the rank-one constraint $\text{rank}(\mathbf{W}_k) = 1$, which would make the problem nonconvex again. Surprisingly, as was shown in [36], it turns out that the relaxed problem (2.19) always has one solution where all $\mathbf{W}_k$'s have rank one and, as a consequence, it is not just a relaxation but actually an equivalent reformulation of (2.18).

In addition, if each user knows its instantaneous channel, it follows that $\mathbf{R}_{k,\beta(k)} = \mathbf{h}_{k,\beta(k)} \mathbf{h}_{k,\beta(k)}^H$, then $\mathbf{w}_k^H \mathbf{R}_{k,\beta(k)} \mathbf{w}_k = \left|\mathbf{w}_k^H \mathbf{h}_{k,\beta(k)}\right|^2$, and the original problem (2.18) can be expressed as an SOCP. This is achieved, as was done in the previous application of robust beamforming, by imposing without loss of generality $\text{Im}\left\{\mathbf{w}_k^H \mathbf{h}_{k,\beta(k)}\right\} = 0$ and $\mathbf{w}_k^H \mathbf{h}_{k,\beta(k)} \geq 0$ (such that $\left|\mathbf{w}_k^H \mathbf{h}_{k,\beta(k)}\right|^2 = \left(\mathbf{w}_k^H \mathbf{h}_{k,\beta(k)}\right)^2$), and by taking the square root on both sides of the inequality constraints to finally obtain a linear trans-

formation of the second order convex cone:

$$\mathbf{w}_k^H \mathbf{h}_{k,\beta(k)} \geq \sqrt{\gamma_k \sum_{l \neq k} \mathbf{w}_l^H \mathbf{R}_{k,\beta(l)} \mathbf{w}_l + \gamma_k \sigma_k^2}. \tag{2.20}$$

Duality Between Channel Capacity and Rate Distortion

As Shannon himself pointed out in 1959 [37], the two fundamental limits of data transmission and data compression are "somewhat dual". However, such a relation between the two problems is not a "duality by mapping" (in the sense that both problems cannot be related by simple mappings of variables, constant parameters, and mathematical operations). As was unveiled in [38], using convex optimization tools, it turns out that the Lagrange dual formulation of the two problems exhibit a precise "duality by mapping" in the form of two geometric problems, resolving the apparent asymmetry between the two original problems. This is an example of how convex optimization can be used to perform an analytical study of a problem.

Network Optimization Problems

In wireless networks, the optimal routing of data depends on the link capacities which, in turn, are determined by the allocation of communications resources (such as powers and bandwidths) to the links. Traditionally, the link capacities are assumed fixed and the routing problem is often formulated as a convex multicommodity network flow problem. However, the optimal performance of the network can only be achieved by simultaneous optimization of routing and resource allocation. In [39], such a problem was formulated as a convex optimization problem over the network flow variables and the communications variables. In addition, the structure of the problem was exploited to obtain efficient algorithms based on a decomposition approach of the dual problem.

Many existing multihop networks are based on the TCP protocol with some mechanism of congestion control such as Reno or Vegas (which essentially control the transmission rate of each source). Indeed, not only TCP is the predominant protocol in the Internet but it is also being extended to wireless networks. It was recently shown in [40] that this type of congestion control can be interpreted as a distributed primal-dual algorithm carried out by sources and links over the network to solve a global network utility maximization problem (different protocols correspond to different objective functions in the global problem). For example, it turns out that the congestion avoidance mechanism of Vegas can be interpreted as an approximate gradient projection algorithm to solve the dual problem. In [41], such an interpretation was extended to ad-hoc wireless networks with flexible link capacities as a function of the allocated powers and interference, obtaining joint congestion control and power control iterative algorithms.

Iterative Waterfilling

The iterative waterfilling algorithm for the multiple-access channel [42] is an example of a simple resolution of a convex problem based on the KKT conditions. Just to give a flavour of the solution, it turns out that the following convex problem that obtains

the transmit covariance matrices $\{\mathbf{Q}_k\}$ that achieve the sum-capacity for the K-user multiple-access channel with channels $\{\mathbf{H}_k\}$ and noise covariance matrix $\mathbf{R}_n$:

$$\begin{array}{ll} \underset{\{\mathbf{Q}_k\}}{\text{maximize}} & \log\det\left(\sum_{k=1}^{K} \mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^H + \mathbf{R}_n\right) \\ \text{subject to} & \text{Tr}\left(\mathbf{Q}_k\right) \leq P_k \qquad 1 \leq k \leq K \\ & \mathbf{Q}_k \geq \mathbf{0} \end{array} \qquad (2.21)$$

can be solved very efficiently in practice by solving a sequence of simpler problems. In particular, each user k should solve in a sequential order the convex problem

$$\begin{array}{ll} \underset{\mathbf{Q}_k}{\text{maximize}} & \log\det\left(\mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^H + \mathbf{R}_k\right) \\ \text{subject to} & \text{Tr}\left(\mathbf{Q}_k\right) \leq P_k \\ & \mathbf{Q}_k \geq \mathbf{0} \end{array} \qquad (2.22)$$

where $\mathbf{R}_k = \sum_{l\neq k} \mathbf{H}_l\mathbf{Q}_l\mathbf{H}_l^H + \mathbf{R}_n$. This problem has a well-known solution [10, 3] given by a $\mathbf{Q}_k$ with eigenvectors equal to the those of $\mathbf{R}_{H,k} = \mathbf{H}_k^H\mathbf{R}_k^{-1}\mathbf{H}_k$ and eigenvalues given by a waterfilling solution (easily derived from the KKT conditions) of the form $\lambda_i\left(\mathbf{Q}_k\right) = \left(\mu_k - \lambda_i^{-1}\left(\mathbf{R}_{H,k}\right)\right)^+$.

This sequential updating of the $\mathbf{Q}_k$'s is in fact a particular instance of the nonlinear Gauss-Seidel algorithm [43].

Linear MIMO Transceiver Design

The design of linear transceivers for point-to-point MIMO systems was formulated in convex form in [15] and [16] (see also [17]) for a wide family of measures of the system quality, after a change of variable based on majorization theory [44]. This problem is treated in detail in §2.4.

The design of linear transceivers in the multiuser case is even more difficult and a general result is still missing. However, an interesting convex formulation as an SDP was obtained in [45] (see also [27]) for the particular case of minimizing the average of the mean square errors (MSEs) of all substreams and users.

2.3 System Model and Preliminaries

2.3.1 Signal Model

The baseband signal model corresponding to a transmission through a general MIMO communication channel with n_T transmit and n_R receive dimensions is

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n} \qquad (2.23)$$

where $\mathbf{s} \in \mathbb{C}^{n_T \times 1}$ is the transmitted vector, $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$ is the channel matrix, $\mathbf{y} \in \mathbb{C}^{n_R \times 1}$ is the received vector, and $\mathbf{n} \in \mathbb{C}^{n_R \times 1}$ is a zero-mean circularly symmetric complex Gaussian interference-plus-noise vector with arbitrary covariance matrix $\mathbf{R}_n$.

The focus is on systems employing linear transceivers (composed of a linear precoder at the transmitter and a linear equalizer at the receiver), as opposed to nonlinear ones such as those including a maximum likelihood (ML) receiver, for reasons
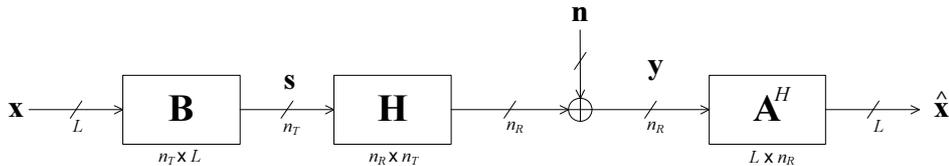
Figure 2.1: Scheme of a general MIMO communication system with a linear transceiver.

of practical complexity (decision-feedback receivers may be an interesting alternative in terms of performance/complexity).

The transmitted vector can be written as (see Figure 2.1)

$$\mathbf{s} = \mathbf{B}\mathbf{x} \tag{2.24}$$

where $\mathbf{B} \in \mathbb{C}^{n_T \times L}$ is the transmit matrix (linear precoder) and $\mathbf{x} \in \mathbb{C}^{L \times 1}$ is the data vector that contains the L symbols to be transmitted (zero-mean,[6] normalized and uncorrelated, i.e., $\mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right] = \mathbf{I}$) drawn from a set of constellations. For the sake of notation, it is assumed that $L \leq \min\left(n_R, n_T\right)$. The total average transmitted power (in units of energy per transmission) is

$$P_T = \mathbb{E}[\|\mathbf{s}\|^2] = \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right). \tag{2.25}$$

Similarly, the estimated data vector at the receiver is (see Figure 2.1)

$$\hat{\mathbf{x}} = \mathbf{A}^H \mathbf{y} \tag{2.26}$$

where $\mathbf{A}^H \in \mathbb{C}^{L \times n_R}$ is the receive matrix (linear equalizer).

It is interesting to observe that the ith column of $\mathbf{B}$ and $\mathbf{A}$, $\mathbf{b}_i$ and $\mathbf{a}_i$ respectively, can be interpreted as the transmit and receive beamvectors associated to the ith transmitted symbol x_i:

$$\hat{x}_i = \mathbf{a}_i^H \left(\mathbf{H}\mathbf{b}_i x_i + \mathbf{n}_i\right) \tag{2.27}$$

where $\mathbf{n}_i = \sum_{j \neq i} \mathbf{H}\mathbf{b}_j x_j + \mathbf{n}$ is the equivalent noise seen by the ith substream, with covariance matrix $\mathbf{R}_{n_i} = \sum_{j \neq i} \mathbf{H}\mathbf{b}_j \mathbf{b}_j^H \mathbf{H}^H + \mathbf{R}_n$. Therefore, the linear MIMO transceiver scheme (see Figure 2.1) can be equivalently interpreted as a multiple beamforming transmission (see Figure 2.2).

The previously introduced complex-valued signal model could have been similarly written with an augmented real-valued notation, simply by augmenting the n-dimensional complex vectors to $2n$-dimensional real vectors (stacking the real and imaginary parts). However, the use of a complex-valued notation is always preferred since it models the system in a simpler and more compact way. Interestingly, it turns

---

[6]If a constellation does not have zero mean, the receiver can always remove the mean and then proceed as if the mean was zero, resulting in a loss of transmitted power. Indeed, the mean of the signal does not carry any information and can always be set to zero saving power at the transmitter.
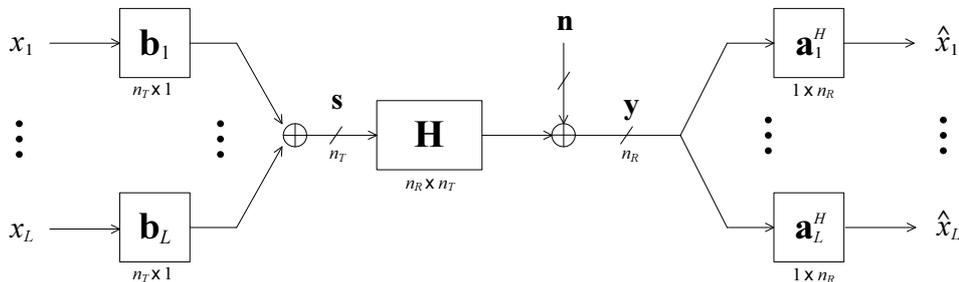
Figure 2.2: Interpretation of a linear MIMO transceiver as a multiple beamforming scheme.

out that complex linear filtering is equivalent to (augmented) real linear filtering if the random vectors involved are proper [46] or circular [47]; otherwise, complex linear filtering is suboptimal and it is necessary to consider either real linear filtering or widely complex linear filtering [47, 48]. Fortunately, many of the commonly employed constellations, such as the family of QAM constellations, are proper [49] and this allows the use of a nice complex notation (although some other constellations, such as BPSK and GMSK, are improper and a complex notation is not adequate anymore).

As a final comment, it is worth noting that multicarrier systems, although they can always be modeled as in (2.23) by properly defining the channel matrix $\mathbf{H}$ as a block-diagonal matrix containing in each diagonal block the channel at each carrier, may be more conveniently modeled as a set of parallel and non-interfering MIMO channels (c.f. §2.4.4).

2.3.2   Measures of Quality

The quality of the ith established substream or link in (2.27) can be conveniently measured, among others, in terms of MSE, SINR, or bit error rate (BER), defined, respectively, as

$$\mathrm{MSE}_i \triangleq \mathbb{E}[\,|\hat{x}_i - x_i|^2\,] = \left|\mathbf{a}_i^H \mathbf{H} \mathbf{b}_i - 1\right|^2 + \mathbf{a}_i^H \mathbf{R}_{n_i} \mathbf{a}_i \tag{2.28}$$

$$\mathrm{SINR}_i \triangleq \frac{\text{desired component}}{\text{undesired component}} = \frac{\left|\mathbf{a}_i^H \mathbf{H} \mathbf{b}_i\right|^2}{\mathbf{a}_i^H \mathbf{R}_{n_i} \mathbf{a}_i} \tag{2.29}$$

$$\mathrm{BER}_i \triangleq \frac{\text{\# bits in error}}{\text{\# transmitted bits}} \approx \tilde{g}_i\left(\mathrm{SINR}_i\right) \tag{2.30}$$

where $\tilde{g}_i$ is a function that relates the BER to the SINR at the ith substream. For most types of modulations, the BER can indeed be analytically expressed as a function of the SINR when the interference-plus-noise term follows a Gaussian distribution [50, 51, 52]. Otherwise, it is an approximation, although when the number of interfering signals is sufficiently large, the central limit theory can be invoked to show that the distribution converges almost surely to a Gaussian distribution (c.f. [53]) (see [54] for

a more detailed discussion). For example, for square M-ary QAM constellations, the BER is [50, 52]

$$\mathrm{BER}\,(\mathrm{SINR}) \approx \frac{4}{\log_2 M}\left(1 - \frac{1}{\sqrt{M}}\right)\mathcal{Q}\left(\sqrt{\frac{3}{M-1}\,\mathrm{SINR}}\right) \qquad (2.31)$$

where $\mathcal{Q}$ is defined as $\mathcal{Q}\,(x) \triangleq \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\lambda^2/2}d\lambda$ [51].[7] It is sometimes convenient to use the Chernoff upper bound of the tail of the Gaussian distribution function $\mathcal{Q}\,(x) \leq \frac{1}{2}e^{-x^2/2}$ [51] to approximate the symbol error probability (which becomes a reasonable approximation for high values of the SINR).

It is worth pointing out that expressing the BER as in (2.30) implicitly assumes that the different links are independently detected after the joint linear processing with the receive matrix $\mathbf{A}$. This reduces the complexity drastically compared to a joint ML detection and is indeed the main advantage of using the receive matrix $\mathbf{A}$.

Any properly designed system should attempt to somehow minimize the MSEs, maximize the SINRs, or minimize the BERs, as is mathematically formulated in §2.4.1.

2.3.3 Optimum Linear Receiver

The optimum linear receiver can be easily designed independently of the particular criterion chosen to design the system (c.f. §2.4.1), provided that the system quality improves with smaller MSEs, higher SINRs, and smaller BERs (for more details, the reader is referred to [15, 17]).

It is notationally convenient to define the MSE matrix as

$$\begin{aligned}\mathbf{E} &\triangleq \mathbb{E}[\,(\hat{\mathbf{x}} - \mathbf{x})\,(\hat{\mathbf{x}} - \mathbf{x})^H\,]\\ &= \left(\mathbf{A}^H\mathbf{H}\mathbf{B} - \mathbf{I}\right)\left(\mathbf{B}^H\mathbf{H}^H\mathbf{A} - \mathbf{I}\right) + \mathbf{A}^H\mathbf{R}_n\mathbf{A}\end{aligned} \qquad (2.32)$$

from which the MSE of the ith link is obtained as the ith diagonal element of $\mathbf{E}$, i.e., $\mathrm{MSE}_i = [\mathbf{E}]_{ii}$.

The receive matrix $\mathbf{A}$ can be easily optimized for a given fixed transmit matrix $\mathbf{B}$, since the minimization of the MSE of a substream with respect to $\mathbf{A}$ does not incur any penalty on the other substreams (see, for example, (2.27) where $\mathbf{a}_i$ only affects $\hat{x}_i$). In other words, there is no tradeoff among the MSEs and the problem decouples. Therefore, it is possible to minimize simultaneously all MSEs and this is precisely how the well-known linear MMSE receiver, also termed Wiener filter, is obtained [55] (see also [15, 16]). If the additional ZF constraint $\mathbf{A}^H\mathbf{H}\mathbf{B} = \mathbf{I}$ is imposed to avoid crosstalk among the substreams (which happens with the MMSE receiver), then the well-known linear ZF receiver is obtained. Interestingly, the MMSE and ZF linear receivers are also optimum (within the class of linear receivers) in the sense that they maximize simultaneously all SINRs and, consequently, minimize simultaneously all BERs (c.f. [15, 17]).

---

[7]The $\mathcal{Q}$-function and the commonly used complementary error function "erfc" are related as $\mathrm{erfc}\,(x) = 2\,\mathcal{Q}(\sqrt{2}x)$ [51].

The MMSE and ZF linear receivers can be compactly written as

$$\mathbf{A} = \mathbf{R}_n^{-1} \mathbf{H} \mathbf{B} \left(\nu \mathbf{I} + \mathbf{B}^H \mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H} \mathbf{B} \right)^{-1} \tag{2.33}$$

where ν is a parameter defined as $\nu \triangleq \begin{cases} 1 & \text{for the MMSE receiver} \\ 0 & \text{for the ZF receiver} \end{cases}$. The MSE matrix (2.32) reduces then to the following concentrated MSE matrix

$$\mathbf{E} = \left(\nu \mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B} \right)^{-1} \tag{2.34}$$

where $\mathbf{R}_H \triangleq \mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H}$ is the squared whitened channel matrix.

Relation among Different Measures of Quality

It is convenient now to relate the different measures of quality, namely, MSE, SINR, and BER, to the concentrated MSE matrix in (2.34). From the definition of MSE matrix, the individual MSEs are given by the diagonal elements:

$$\text{MSE}_i = \left[\left(\nu \mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B} \right)^{-1} \right]_{ii}. \tag{2.35}$$

It turns out that the SINRs and the MSEs are trivially related when using the MMSE or ZF linear receivers as [15, 16, 17]

$$\text{SINR}_i = \frac{1}{\text{MSE}_i} - \nu. \tag{2.36}$$

Finally, the BERs can also be written as a function of the MSEs:

$$\text{BER}_i = g_i \left(\text{MSE}_i \right) \triangleq \tilde{g}_i \left(\text{SINR}_i = \text{MSE}_i^{-1} - \nu \right) \tag{2.37}$$

where $\tilde{g}_i$ was defined in (2.30).

2.4 Beamforming Design for MIMO Channels: A Convex Optimization Approach

The design of transmit-receive beamforming or linear MIMO transceivers has been studied since the 1970s where cable systems were the main application [56, 57]. The traditional results existing in the literature have dealt with the problem from a narrow perspective (due to the complexity of the problem); the basic approach has been to choose a measure of quality of the system sufficiently simple such that the problem can be analytically solved. Some examples include: the minimization of the (weighted) sum of the MSEs of the substreams or, equivalently, the trace of the MSE matrix [56, 57, 58, 6, 59]; the minimization of the determinant of the MSE matrix [60]; and the maximization of the SINR with a ZF constraint [6]. For these criteria, the original complicated design problem is greatly simplified because the channel turns out to be diagonalized by the optimal transmit-receive processing and the transmission is effectively performed on a diagonal or parallel fashion. The diagonal transmission allows a

scalarization of the problem (meaning that all matrix equations are substituted with scalar ones) with the consequent simplification.

Recent results have considered more elaborated and meaningful measures of quality. In [61], the minimization of the BER (and also of the Chernoff upper bound) averaged over the channel substreams was treated in detail when a diagonal structure is imposed. The minimum BER design without the diagonal structure constraint has been independently obtained in [62] and [15], resulting in an optimal nondiagonal structure. This result, however, only holds when the constellations used in all channel substreams are equal. The general case of different constellations was treated and optimally solved in [54] (different constellations are typically obtained when some kind of bit allocation strategy is used such as the gap-approximation method [63, Part II] which chooses the constellations as a function of the channel realization). In [15], a general unifying framework was developed that embraces a wide range of different design criteria; in particular, the optimal design was obtained for the family of Schur-concave and Schur-convex cost functions [44].

Clearly, the problem faced when designing a MIMO system not only lies on the design itself but also on the choice of the appropriate measure of the system quality (which may depend on the application at hand and/or on the type of coding used on top of the uncoded system). In fact, to fully characterize such a problem, a multi-objective optimization approach should be taken to characterize the Pareto-optimal set.[8] Following the results in [16, 17], this section deals first with the optimal design subject to a set of independent QoS constraints for each of the channel substreams. This allows the characterization of the Pareto-optimal set and the feasible region (see the numerical example in Figure 2.7). However, since it is generally more convenient to use a single measure of the system quality to simplify the characterization, the optimal design subject to a global QoS constraint that measures the system quality is also considered based on [15, 17].

2.4.1 Problem Formulation

The problems addressed in this section are the minimization of the transmitted power $\mathrm{Tr}(\mathbf{BB}^H)$ subject to either independent QoS constraints or a global QoS constraint. The problems are originally formulated in terms of the variables $\mathbf{A}$ and $\mathbf{B}$. However, as was obtained in §2.3.3, the optimal receive matrix $\mathbf{A}$ is always given by (2.33) and the problems can then be rewritten as optimization problems with respect to only the transmit matrix $\mathbf{B}$.

Independent QoS Constraints

Independent QoS constraints can always be expressed in terms of MSE constraints $\mathrm{MSE}_i \leq \rho_i$, where ρ_i denotes the maximum MSE value for the ith substream (recall that SINR and BER constraints can always be rewritten as MSE constraints, c.f.

---

[8]A Pareto optimal solution is an optimal solution to a multi-objective optimization problem; it is defined as any solution that cannot be improved with respect to any component without worsening the others [8].

§2.3.3). The problem can then be formulated as

$$\begin{array}{ll} \underset{\mathbf{A},\mathbf{B}}{\text{minimize}} & \text{Tr}\left(\mathbf{BB}^H\right) \\ \text{subject to} & \text{MSE}_i \leq \rho_i \qquad 1 \leq i \leq L. \end{array} \qquad (2.38)$$

This problem is optimally solved in §2.4.2 and is then used in a numerical example in §2.4.5 to obtain the achievable region for a given power budget P_0 (a given set of constraints $\{\rho_i\}$ is achievable if and only if the minimum required power is no greater than P_0).

Global QoS Constraint

In this case, it is assumed that the performance of the system is measured by a global cost function f of the MSEs (as before, cost functions of the SINRs and BERs can always be rewritten in terms of MSEs, c.f. §2.3.3). The problem can be formulated as

$$\begin{array}{ll} \underset{\mathbf{A},\mathbf{B}}{\text{minimize}} & \text{Tr}\left(\mathbf{BB}^H\right) \\ \text{subject to} & f\left(\{\text{MSE}_i\}\right) \leq \alpha_0 \end{array} \qquad (2.39)$$

where α_0 is the required level of the global performance as measured by the cost function f. In principle, any function can be used to measure the system quality as long as it is increasing in each argument (this is a mild and completely reasonable assumption: if the quality of one of the substream improves while the rest remain unchanged, any reasonable function should properly reflect this difference).

It is important to point out that this problem can be similarly formulated as the minimization of the cost function subject to a given power budget P_0. Both formulations are essentially equivalent since they describe the same tradeoff curve of performance vs. power. In fact, numerically speaking, each problem can be solved by iteratively solving the other combined with the bisection method [14, Alg. 4.1].

Illustrative Example

As a motivation to the need of solving problems (2.38) and (2.39) optimally, the following example shows that a simple design imposing a diagonal transmission is not necessarily good.

Consider a system with the following characteristics: a diagonal 2×2 MIMO channel $\mathbf{H} = \text{diag}\left(\{1, \epsilon\}\right)$, a white normalized noise $\mathbf{R}_n = \mathbf{I}$, two established substreams $L = 2$ with an MMSE receiver (see (2.33) with $\nu = 1$), with a power budget P_0, and a design based on maximizing the minimum of the SINRs of the substreams $\text{SINR}_0 = \min\{\text{SINR}_1, \text{SINR}_2\}$.

A naive design imposing a diagonal transmission is suboptimal in this case (c.f. Theorem 2.4.3) and is given by

$$\mathbf{B} = \left[\begin{array}{cc} \sqrt{p_1} & 0 \\ 0 & \sqrt{p_2} \end{array}\right], \qquad \mathbf{A} = \left[\begin{array}{cc} \sqrt{p_1}/\left(1 + p_1\right) & 0 \\ 0 & \epsilon\sqrt{p_2}/\left(1 + \epsilon^2 p_2\right) \end{array}\right] \qquad (2.40)$$

with SINRs given by $\text{SINR}_1 = p_1$ and $\text{SINR}_2 = \epsilon^2 p_2$. The optimal power allocation is given by $p_1 = P_0\epsilon^2/\left(1 + \epsilon^2\right)$ and $p_2 = P_0/\left(1 + \epsilon^2\right)$, and then both substreams have the same SINR given by $\text{SINR}_0 = P_0\epsilon^2/\left(1 + \epsilon^2\right)$.

An optimal design yields a nondiagonal transmission (c.f. Theorem 2.4.3) given by

$$\mathbf{B} = \begin{bmatrix} \sqrt{p_1} & 0 \\ 0 & \sqrt{p_2} \end{bmatrix} \bar{\mathbf{H}}_2, \qquad \mathbf{A} = \begin{bmatrix} \sqrt{p_1}/\left(1+p_1\right) & 0 \\ 0 & \epsilon\sqrt{p_2}/\left(1+\epsilon^2 p_2\right) \end{bmatrix} \bar{\mathbf{H}}_2 \quad (2.41)$$

with equal SINRs given by $\text{SINR}_0 = \text{MSE}_0^{-1} - 1$ (from (2.36)), where $\bar{\mathbf{H}}_2$ is a 2×2 unitary Hadamard matrix (c.f. Theorem 2.4.3) and $\text{MSE}_0 = \frac{1}{2}\left(1/\left(1+p_1\right) + 1/\left(1+\epsilon^2 p_2\right)\right)$ (from (2.61)). The optimal power allocation in this case is $p_1 = (\mu - 1)^+$ and $p_2 = \left(\mu\epsilon^{-1} - \epsilon^{-2}\right)^+$ (from (2.63)), which simplifies to $p_1 = P_0$ and $p_2 = 0$ for sufficiently small ϵ ($\epsilon < 1/\left(1+P_0\right)$) with a final SINR given by $\text{SINR}_0 = P_0/\left(2+P_0\right)$.

Both solutions can be easily compared for small ϵ: the performance of the suboptimal diagonal transmission goes to zero with ϵ whereas the optimal transmission is robust and less sensitive.

2.4.2 Optimal Design with Independent QoS Constraints

Using the optimal receive matrix (2.33), problem (2.38) can be rewritten as

$$\begin{aligned} \underset{\mathbf{B}}{\text{minimize}} \quad & \text{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \\ \text{subject to} \quad & \mathbf{d}\left(\left(\nu\mathbf{I} + \mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}\right) \le \boldsymbol{\rho} \end{aligned} \quad (2.42)$$

where the elements of $\boldsymbol{\rho}$ are assumed in decreasing order without loss of generality (by properly relabeling the elements).

It is now possible to write a simpler and equivalent problem by using a fundamental result of majorization theory [44] that says that a matrix with given eigenvalues $\boldsymbol{\lambda}$ and diagonal elements less than $\mathbf{d}$ can be constructed if and only if $\boldsymbol{\lambda}$ weakly majorizes $\mathbf{d}$, i.e., $\boldsymbol{\lambda} \succ^w \mathbf{d}$ [44, 9.B.1, 9.B.2, & 5.A.9.a].[9] Defining $\tilde{\mathbf{B}} = \mathbf{B}\mathbf{Q}^H$, where $\mathbf{Q}^H$ is a unitary matrix that diagonalizes $\mathbf{B}^H\mathbf{R}_H\mathbf{B}$ (with the resulting diagonal elements in increasing order), the equivalent problem is

$$\begin{aligned} \underset{\tilde{\mathbf{B}}}{\text{minimize}} \quad & \text{Tr}\left(\tilde{\mathbf{B}}\tilde{\mathbf{B}}^H\right) \\ \text{subject to} \quad & \tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}} \quad \text{diagonal (increasing diag. elements)} \\ & \boldsymbol{\lambda}\left(\left(\nu\mathbf{I} + \tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}}\right)^{-1}\right) \succ^w \boldsymbol{\rho}. \end{aligned} \quad (2.43)$$

Conversely, given $\tilde{\mathbf{B}}$, it is not difficult to compute a unitary matrix $\mathbf{Q}$ and form $\mathbf{B} = \tilde{\mathbf{B}}\mathbf{Q}$ such that $\mathbf{d}\left(\left(\nu\mathbf{I} + \mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}\right) \le \boldsymbol{\rho}$ (with equality at an optimal solution) with the practical algorithm given in [64, IV-A].

Since $\tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}}$ is diagonal with diagonal elements in increasing order, it can be shown [15, Lem. 12][17, Lem. 5.11] that $\tilde{\mathbf{B}}$ can be assumed without loss of optimality of the form $\tilde{\mathbf{B}} = \mathbf{U}_{H,1}\boldsymbol{\Sigma}_B$, where $\mathbf{U}_{H,1} \in \mathbb{C}^{n_T \times L}$ is a (semi-)unitary matrix that has as columns the eigenvectors of $\mathbf{R}_H$ corresponding to the L largest eigenvalues $\{\lambda_{H,i}\}$

---

[9] The weakly majorization relation $\mathbf{y} \succ^w \mathbf{x}$ is defined as $\sum_{j=i}^n y_j \le \sum_{j=i}^n x_j$ for $1 \le i \le n$, where the elements of $\mathbf{y}$ and $\mathbf{x}$ are assumed in decreasing order [44].

in increasing order and $\mathbf{\Sigma}_B = \text{diag}\left(\left\{\sqrt{p_i}\right\}\right) \in \mathbb{R}^{L \times L}$ is a diagonal matrix with a power allocation $\{p_i\}$ over the channel eigenmodes (note the need for the additional constraints $p_i \geq 0$).

Using the form of the optimal $\tilde{\mathbf{B}}$ and writing the weakly majorization relation explicitly, the problem finally becomes

$$
\begin{array}{ll}
\underset{\mathbf{p}}{\text{minimize}} & \sum_{i=1}^{L} p_i \\
\text{subject to} & \sum_{j=i}^{L} \frac{1}{\nu + p_j \lambda_{H,j}} \leq \sum_{j=i}^{L} \rho_j \qquad 1 \leq i \leq L \\
& p_i \geq 0
\end{array}
\tag{2.44}
$$

which is a very simple convex problem. In principle, problem (2.44) is a relaxation since it lacks the constraints $p_i \lambda_{H,i} \leq p_{i+1} \lambda_{H,i+1}$ (to guarantee that the diagonal elements of $\tilde{\mathbf{B}}^H \mathbf{R}_H \tilde{\mathbf{B}}$ are in increasing order). However, it is not difficult to see that any optimal point must necessarily satisfy them and, hence, problem (2.44) is indeed equivalent to the original one (suppose $p_i \lambda_{H,i} > p_{i+1} \lambda_{H,i+1}$ for some i, then the terms $p_i \lambda_{H,i}$ and $p_{i+1} \lambda_{H,i+1}$ could be swapped to satisfy the ordering constraint without affecting the problem (2.44), but this cannot be at an optimal point because the objective value could be further reduced by using the optimal increasing ordering of the $\lambda_{H,i}$'s). The following theorem summarizes the whole simplification.

Theorem 2.4.1 *The original complicated nonconvex problem (2.42), with the elements of $\boldsymbol{\rho}$ in decreasing order w.l.o.g., is equivalent to the simple convex problem (2.44), where the $\lambda_{H,i}$'s are the L largest eigenvalues of $\mathbf{R}_H$ in increasing order. The mapping from (2.44) to (2.42) is given by*

$$
\mathbf{B} = \mathbf{U}_{H,1} \text{diag}\left(\left\{\sqrt{p_i}\right\}\right) \mathbf{Q},
\tag{2.45}
$$

where $\mathbf{U}_{H,1}$ diagonalizes the channel matrix, $\{p_i\}$ is the power allocation over the channel eigenmodes, and $\mathbf{Q}$ is a rotation, chosen such that $\mathbf{d}\left(\left(\nu\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}\right) = \boldsymbol{\rho}$ (e.g., with the algorithm in [64, IV-A]), that spreads the transmitted symbols over the channel eigenmodes (see Figure 2.3).

Furthermore, the convex problem (2.44) can be easily solved in practice with a simple multilevel waterfilling algorithm as given in [16, 17].

The global transmit-receive process $\hat{\mathbf{x}} = \mathbf{A}^H \left(\mathbf{HBx} + \mathbf{n}\right)$ using the optimal transmission structure of Theorem 2.4.1 can be written as

$$
\hat{\mathbf{x}} = \mathbf{Q}^H \left(\nu\mathbf{I} + \mathbf{\Sigma}_B^H \mathbf{D}_{H,1} \mathbf{\Sigma}_B\right)^{-1} \mathbf{\Sigma}_B^H \mathbf{D}_{H,1}^{1/2} \left(\mathbf{D}_{H,1}^{1/2} \mathbf{\Sigma}_B \mathbf{Q}\mathbf{x} + \mathbf{w}\right)
\tag{2.46}
$$

or, equivalently, as

$$
\hat{x}_i^Q = \alpha_i \left(\sqrt{p_i \lambda_{H,i}} \, x_i^Q + w_i\right) \qquad 1 \leq i \leq L
\tag{2.47}
$$

where $\mathbf{w}$ is an equivalent normalized white noise ($\mathbb{E}\left[\mathbf{w}\mathbf{w}^H\right] = \mathbf{I}$), $\mathbf{D}_{H,1} = \mathbf{U}_{H,1}^H \mathbf{R}_H \mathbf{U}_{H,1}$, $\mathbf{x}^Q = \mathbf{Q}\mathbf{x}$, and $\alpha_i = \sqrt{p_i \lambda_{H,i}}/\left(\nu + p_i \lambda_{H,i}\right)$ (see Figure 2.3 with $\lambda_i \triangleq \lambda_{H,i}$).
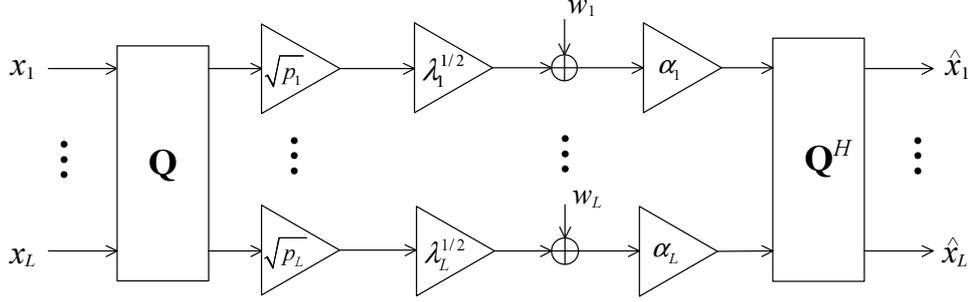
Figure 2.3 Decomposition of the optimal transmission through a MIMO channel.

2.4.3 Optimal Design with a Global QoS Constraint

Using the optimal receive matrix (2.33), problem (2.39) reduces to

$$
\begin{aligned}
\underset{\mathbf{B}}{\text{minimize}} \quad & \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \\
\text{subject to} \quad & f\left(\mathbf{d}\left(\left(\nu\mathbf{I}+\mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}\right)\right) \leq \alpha_0.
\end{aligned}
\tag{2.48}
$$

This problem can be easily simplified by using the previous result in Theorem 2.4.1. First, rewrite (2.48) as

$$
\begin{aligned}
\underset{\mathbf{B},\boldsymbol{\rho}}{\text{minimize}} \quad & \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \\
\text{subject to} \quad & \mathbf{d}\left(\left(\nu\mathbf{I}+\mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}\right) \leq \boldsymbol{\rho} \\
& f\left(\boldsymbol{\rho}\right) \leq \alpha_0
\end{aligned}
\tag{2.49}
$$

which can always be done since f is increasing in each argument. Then, use Theorem 2.4.1 to reformulate the problem as

$$
\begin{aligned}
\underset{\mathbf{p},\boldsymbol{\rho}}{\text{minimize}} \quad & \sum_{i=1}^{L} p_i \\
\text{subject to} \quad & \sum_{j=i}^{L} \frac{1}{\nu+p_j\lambda_{H,j}} \leq \sum_{j=i}^{L}\rho_{[j]} \qquad 1 \leq i \leq L \\
& p_i \geq 0 \\
& f\left(\rho_1,\cdots,\rho_L\right) \leq \alpha_0
\end{aligned}
\tag{2.50}
$$

where the ordering constraints $p_i\lambda_{H,i} \leq p_{i+1}\lambda_{H,i+1}$ are not necessary for the same reasons as in (2.44) and $\rho_{[i]}$ denotes the ρ_i's in decreasing order which can be explicitly written as [14]:

$$
\sum_{j=i}^{L}\rho_{[j]} = \min\left\{\rho_{j_1}+\cdots+\rho_{j_{L-i+1}} \mid 1 \leq j_1 < \cdots < j_{L-i+1} \leq L\right\}.
\tag{2.51}
$$

This is clearly a concave function since it is the pointwise minimum of concave (affine) functions.

To avoid the constraint (2.51), it is convenient to assume that the cost function f is minimum when the arguments are sorted in decreasing order.[10] The problem can then be rewritten as

$$
\begin{aligned}
\underset{\mathbf{p}, \boldsymbol{\rho}}{\text{minimize}} \quad & \textstyle\sum_{i=1}^{L} p_i \\
\text{subject to} \quad & \textstyle\sum_{j=i}^{L} \frac{1}{\nu + p_j \lambda_{H,j}} \leq \sum_{j=i}^{L} \rho_j \qquad 1 \leq i \leq L \\
& p_i \geq 0 \\
& \rho_i \geq \rho_{i+1} \\
& f\left(\rho_1, \cdots, \rho_L\right) \leq \alpha_0.
\end{aligned}
\tag{2.52}
$$

If, in addition, the cost function f is convex, then the constraints $\rho_i \geq \rho_{i+1}$ are not necessary (since any optimal solution cannot have $\rho_i < \rho_{i+1}$ because the problem would have a lower objective value by using instead $\tilde{\rho}_i = \tilde{\rho}_{i+1} = \left(\rho_i + \rho_{i+1}\right)/2$ [54]) and the problem can be finally written in convex form as

$$
\begin{aligned}
\underset{\mathbf{p}, \boldsymbol{\rho}}{\text{minimize}} \quad & \textstyle\sum_{i=1}^{L} p_i \\
\text{subject to} \quad & \textstyle\sum_{j=i}^{L} \frac{1}{\nu + p_j \lambda_{H,j}} \leq \sum_{j=i}^{L} \rho_j \qquad 1 \leq i \leq L \\
& p_i \geq 0 \\
& f\left(\rho_1, \cdots, \rho_L\right) \leq \alpha_0.
\end{aligned}
\tag{2.53}
$$

The following theorem summarizes the simplification.

Theorem 2.4.2 *The original complicated nonconvex problem (2.48), with a cost function f increasing in each variable, is equivalent to the simple problem (2.50), where the $\lambda_{H,i}$'s are the L largest eigenvalues of $\mathbf{R}_H$ in increasing order. In addition, if f is convex and is minimum when its arguments are sorted in decreasing order, the problem further simplifies to the convex problem (2.53). The mapping from (2.53) to (2.48) is given by*

$$
\mathbf{B} = \mathbf{U}_{H,1} \operatorname{diag}\left(\left\{\sqrt{p_i}\right\}\right) \mathbf{Q},
$$

where $\mathbf{U}_{H,1}$ diagonalizes the channel matrix, $\{p_i\}$ is the power allocation over the channel eigenmodes, and $\mathbf{Q}$ is a rotation, chosen such that $\mathbf{d}\left(\left(\nu\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}\right) = \boldsymbol{\rho}$ (e.g., with the algorithm in [64, IV-A]), that spreads the transmitted symbols over the channel eigenmodes (see Figure 2.3).

Interestingly, Theorem 2.4.2 can be further simplified for the family of Schur-concave/convex functions as shown next. First, rewrite the MSE constraints of (2.49) (knowing that they are satisfied with equality at an optimal point) as

$$
\boldsymbol{\rho} = \mathbf{d}\left(\mathbf{Q}^H (\nu\mathbf{I} + \tilde{\mathbf{B}}^H \mathbf{R}_H \tilde{\mathbf{B}})^{-1} \mathbf{Q}\right).
\tag{2.54}
$$

Now it suffices to use the definition of Schur-concavity/convexity to obtain the desired result. In particular, if f is Schur-concave, it follows from the definition of Schur-concavity [44] (the diagonal elements and eigenvalues are assumed in decreasing order)

---

[10]In practice, most cost functions are minimized when the arguments are in a specific ordering (if not, one can always use instead the function $\tilde{f}_0\left(\mathbf{x}\right) = \min_{\mathbf{P} \in \mathcal{P}} f_0\left(\mathbf{P}\mathbf{x}\right)$, where $\mathcal{P}$ is the set of all permutation matrices) and, hence, the decreasing ordering can be taken without loss of generality.

that

$$f\left(\mathbf{d}\left(\mathbf{X}\right)\right) \geq f\left(\boldsymbol{\lambda}\left(\mathbf{X}\right)\right) \qquad (2.55)$$

which means that $f\left(\boldsymbol{\rho}\right)$ is minimum when $\mathbf{Q} = \mathbf{I}$ in (2.54) (since $(\nu\mathbf{I}+\tilde{\mathbf{B}}^{H}\mathbf{R}_{H}\tilde{\mathbf{B}})^{-1}$ is already diagonal with diagonal elements in decreasing order by definition). If f is Schur-convex, the opposite happens:

$$f\left(\mathbf{d}\left(\mathbf{X}\right)\right) \geq f\left(\mathbf{1}\times\operatorname{Tr}\left(\mathbf{X}\right)/L\right) \qquad (2.56)$$

where $\mathbf{1}$ denotes the all-one vector. This means that $f_{0}\left(\boldsymbol{\rho}\right)$ is minimum when $\mathbf{Q}$ is such that $\boldsymbol{\rho}$ has equal elements in (2.54), i.e., when $\mathbf{Q}^{H}(\nu\mathbf{I}+\tilde{\mathbf{B}}^{H}\mathbf{R}_{H}\tilde{\mathbf{B}})^{-1}\mathbf{Q}$ has equal diagonal elements. The following theorem summarizes this results.

Theorem 2.4.3 *The solution to the original problem (2.48) can be further characterized for two particular families of cost functions:*

- *If f is Schur-concave, then an optimal solution is*

$$\mathbf{B} = \mathbf{U}_{H,1}\operatorname{diag}\left(\{\sqrt{p_{i}}\}\right). \qquad (2.57)$$

- *If f is Schur-convex, then an optimal solution is*

$$\mathbf{B} = \mathbf{U}_{H,1}\operatorname{diag}\left(\{\sqrt{p_{i}}\}\right)\mathbf{Q}, \qquad (2.58)$$

where $\mathbf{Q}$ is a unitary matrix such that $\left(\mathbf{I} + \mathbf{B}^{H}\mathbf{R}_{H}\mathbf{B}\right)^{-1}$ has identical diagonal elements. This rotation matrix $\mathbf{Q}$ can be computed with the algorithm in [64, IV-A], as well as with any unitary matrix that satisfies $|[\mathbf{Q}]_{ik}| = |[\mathbf{Q}]_{il}|$, $\forall i, k, l$ such as the unitary Discrete Fourier Transform (DFT) matrix or the unitary Hadamard matrix (when the dimensions are appropriate such as a power of two [51, p. 66]).

Interestingly, Theorem 2.4.3 implies that Schur-concave cost functions lead to parallel transmissions (from the fully channel diagonalization), whereas Schur-convex cost functions result in transmission schemes that spread all the symbols equally through all channel eigenmodes in a CDMA fashion (see Figure 2.4). Hence, the designs obtained with Schur-convex cost functions are inherently more robust to ill-conditioned substreams (due, for example, to fading) and exhibit a better performance (as can be observed from the numerical results in §2.4.5).

It is important to remark that all the results obtained in this section are directly applicable to the opposite problem formulation consisting in the minimization of a cost function subject to a power constraint.

Schur-Concave Cost Functions

For Schur-concave cost functions, the optimal rotation is $\mathbf{Q} = \mathbf{I}$ (from Theorem 2.4.3) and the MSEs are given by

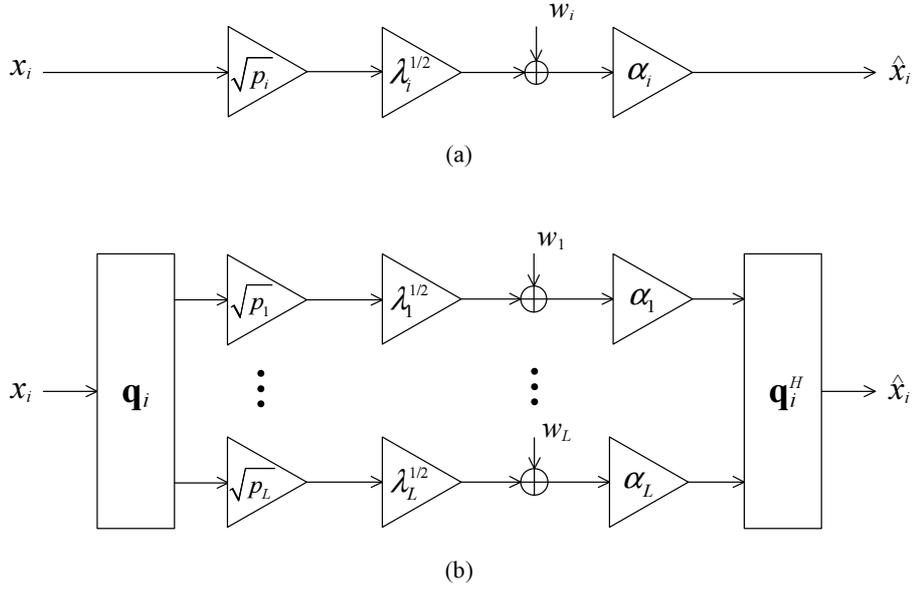$$\mathrm{MSE}_{i} = \frac{1}{\nu + p_{i}\,\lambda_{H,i}} \qquad 1 \leq i \leq L. \qquad (2.59)$$

Figure 2.4: Transmission structure for the ith symbol: (a) Diagonal (parallel) transmission for Schur-concave functions; and (b) Nondiagonal (distributed or spread) transmission for Schur-convex functions.

Note that the SINRs in this case are easily given by $\text{SINR}_i = p_i \lambda_{H,i}$ (from (2.36) and (2.59)) which do not depend on ν.

The original optimization problem (2.48) can be finally written as

$$
\begin{aligned}
\underset{\mathbf{p}}{\text{minimize}} \quad & \textstyle\sum_{j=1}^{L} p_j \\
\text{subject to} \quad & f\left(\left\{\frac{1}{\nu + p_i \lambda_{H,i}}\right\}\right) \leq \alpha_0 \\
& p_i \geq 0 \qquad\qquad 1 \leq i \leq L,
\end{aligned}
\tag{2.60}
$$

whose solution clearly depends on the particular choice of f.

Schur-Convex Cost Functions

For Schur-concave cost functions, the diagonal elements of the MSE matrix $\mathbf{E}$ are equal at the optimal solution (Theorem 2.4.3) and the MSEs are then given by

$$
\text{MSE}_i = \frac{1}{L} \text{Tr}(\mathbf{E}) = \frac{1}{L} \sum_{j=1}^{L} \frac{1}{\nu + p_j \lambda_{H,j}} \qquad 1 \leq i \leq L.
\tag{2.61}
$$

The original optimization problem (2.48) can be finally written as

$$
\begin{aligned}
\underset{\mathbf{P}}{\text{minimize}} \quad & \sum_{j=1}^{L} p_j \\
\text{subject to} \quad & \frac{1}{L} \sum_{j=1}^{L} \frac{1}{\nu + p_j \lambda_{H,j}} \le \rho_0 \\
& p_i \ge 0 \qquad\qquad 1 \le i \le L
\end{aligned}
\qquad (2.62)
$$

where $\rho_0 \triangleq \{\rho \mid f(\mathbf{1} \times \rho) = \alpha_0\}$ is the MSE level required on all substreams to achieve the required global quality.

Surprisingly, this simplified problem for Schur-convex functions does not explicitly depend on the cost function f; in other words, once the required MSE level ρ_0 has been calculated, problem (2.62) is independent of f. The reason is that, since all the MSEs are equal, the cost function simply defines a one-to-one mapping between α_0 and ρ_0. In addition, problem (2.62) is solved by the following waterfilling solution obtained from the KKT optimality conditions:

$$
p_i = \left(\mu \lambda_{H,i}^{-1/2} - \nu \lambda_{H,i}^{-1} \right)^{+} \qquad 1 \le i \le L
\qquad (2.63)
$$

where μ is the waterlevel chosen such that $\frac{1}{L} \sum_{j=1}^{L} \frac{1}{\nu + p_j \lambda_{H,j}} = \rho_0$ (see [16, 17] for a practical numerical evaluation of the waterfilling expression). Note that for the ZF receiver ($\nu = 0$), the waterfilling solution (2.63) simplifies to $p_i = \lambda_{H,i}^{-1/2} \sum_{j=1}^{L} \lambda_{H,j}^{-1/2} / (L\rho_0)$.

List of Schur-Concave and Schur-Convex Cost Functions

The following list of Schur-concave and Schur-convex functions, along with the corresponding closed-form solutions obtained from the KKT conditions, illustrates how powerful is the unifying framework developed in Theorem 2.4.3 (see [15, 17] for a detailed treatment of each case).

Examples of Schur-concave functions (when expressed as functions of the MSEs) for which the diagonal transmission is optimal:

- Minimization of the sum of the MSEs or, equivalently, of $\mathrm{Tr}(\mathbf{E})$ [58, 6] with solution $p_i = \left(\mu \lambda_{H,i}^{-1/2} - \nu \lambda_{H,i}^{-1} \right)^{+}$.

- Minimization of the weighted sum of the MSEs or, equivalently, of $\mathrm{Tr}(\mathbf{WE})$, where $\mathbf{W} = \mathrm{diag}(\{w_i\})$ is a diagonal weighting matrix, [59] with solution $p_i = \left(\mu w_i^{1/2} \lambda_{H,i}^{-1/2} - \nu \lambda_{H,i}^{-1} \right)^{+}$.

- Minimization of the (exponentially weighted) product of the MSEs with solution $p_i = \left(\mu w_i - \nu \lambda_{H,i}^{-1} \right)^{+}$.

- Minimization of $\det(\mathbf{E})$ [60] with solution $p_i = \left(\mu - \nu \lambda_{H,i}^{-1} \right)^{+}$.

- Maximization of the mutual information, e.g. [10], with solution $p_i = \left(\mu - \lambda_{H,i}^{-1} \right)^{+}$.

- Maximization of the (weighted) sum of the SINRs with solution given by allocating the power on the eigenmode with maximum weighted gain $w_i \lambda_{H,i}$.

- Maximization of the (exponentially weighted) product of the SINRs with solution $p_i \propto w_i / \sum_j w_j$ (for the unweighted case, it results in a uniform power allocation).

Examples of Schur-convex functions for which the optimal transmission is nondiagonal with solution given by $p_i = \left(\mu \lambda_{H,i}^{-1/2} - \nu \lambda_{H,i}^{-1} \right)^+$ plus the rotation $\mathbf{Q}$:

- Minimization of the maximum of the MSEs.

- Maximization of the minimum of the SINRs.

- Maximization of the harmonic mean of the SINRs.[11]

- Minimization of the average BER (with equal constellations).

- Minimization of the maximum of the BERs.

A Practical Example: Minimum BER Design

The average (uncoded) BER is a good measure of the uncoded part of a system. Hence, guaranteeing a minimum average BER may be regarded as an excellent criterion:

$$\frac{1}{L} \sum_{i=1}^{L} g_i \left(\mathrm{MSE}_i \right) \leq \mathrm{BER}_0 \qquad (2.64)$$

where the functions g_i were defined in (2.37) and happen to be convex increasing in the MSE for sufficiently small values of the argument (see Figure 2.5) [15, 17]. As a rule-of-thumb, the BER is convex in the MSE for a BER less than 2×10^{-2} (this is a mild assumption, since practical systems have in general a smaller uncoded BER[12]); interestingly, for BPSK and QPSK constellations the BER function is always convex [15, 17].

The optimal receive matrix is given by (2.33) and the problem is

$$\begin{aligned} \underset{\mathbf{B}}{\mathrm{minimize}} \quad & \mathrm{Tr} \left(\mathbf{B} \mathbf{B}^H \right) \\ \mathrm{subject\ to} \quad & \frac{1}{L} \sum_i g_i \left(\left[\left(\nu \mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B} \right)^{-1} \right]_{ii} \right) \leq \mathrm{BER}_0 \end{aligned} \qquad (2.65)$$

where it has been implicitly assumed that the constellations used have been previously chosen with some bit allocation strategy such as the gap-approximation method [63, Part II] or simply equal fixed constellations. Theorem 2.4.2 can now be invoked (provided that the constellations are chosen with increasing cardinality) to simplify

---

[11] For the ZF receiver, the maximization of the harmonic mean of the SINRs is equivalent to the minimization of the unweighted sum of the MSEs, which can be classified as both Schur-concave and Schur-convex (since it is invariant to rotations).

[12] Given an uncoded bit error probability of at most 10^{-2} and using a proper coding scheme, coded bit error probabilities with acceptable low values such as 10^{-6} can be obtained.
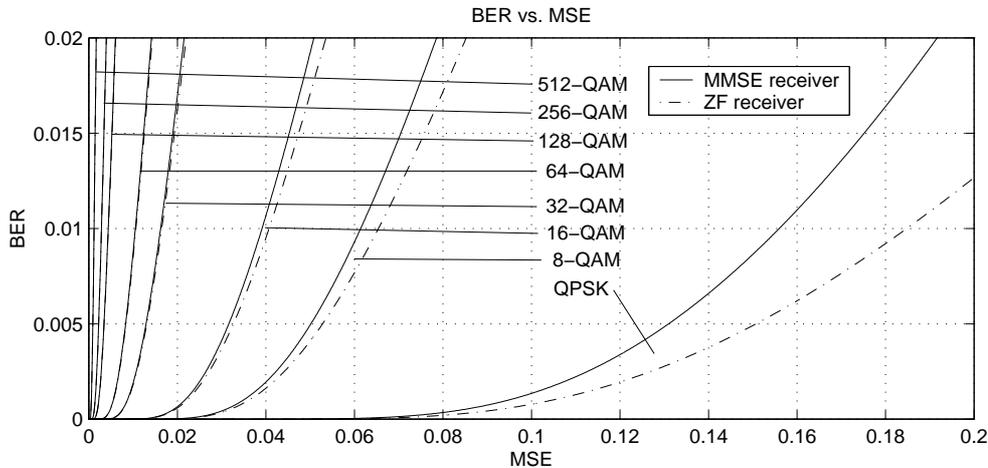
Figure 2.5 BER as a function of the MSE for different QAM constellations.

the problem and reformulate it in convex form as in (2.53). This problem was extensively treated in [54] for the multicarrier case via a primal decomposition approach, which allowed the resolution of the problem with extremely simple algorithms (rather than using general purpose iterative algorithms such as interior-point methods).

In the particular case when the constellations used in the L substreams are equal, the average BER cost function turns out to be Schur-convex since it is the sum of identical convex functions [44, 3.H.2]. Hence, the final problem to be solved is (2.62) with $\rho_0 = g^{-1}(\text{BER}_0)$ and the solution is given by (2.63) (recall that the rotation matrix $\mathbf{Q}$ is needed as indicated in Theorem 2.4.3).

2.4.4 Extension to Multicarrier Systems

As mentioned in §2.3, multicarrier systems may be more conveniently modeled as a communication through a set of parallel MIMO channels

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{s}_k + \mathbf{n}_k \qquad 1 \le k \le N, \tag{2.66}$$

where N is the number of carriers and k is the carrier index, rather than as a single MIMO channel as in (2.23) with $\mathbf{H} = \text{diag}(\{\mathbf{H}_k\})$. The parallel modeling in (2.66) is useful when the signal processing operates independently at each MIMO channel (implying block diagonal matrices $\mathbf{B} = \text{diag}(\{\mathbf{B}_k\})$ and $\mathbf{A} = \text{diag}(\{\mathbf{A}_k\})$), whereas the modeling with a single MIMO channel is more convenient when the signal processing operates jointly over all carriers (meaning full matrices $\mathbf{B}$ and $\mathbf{A}$).

The optimal linear receiver and MSE matrix for multiple MIMO channels as in (2.66) still have the same form as (2.33) and (2.34), respectively, for each MIMO channel (L_k denotes the number of established substreams at the kth MIMO channel).

The extension of the design with independent QoS constraint of §2.4.2 to the case of a set of N parallel MIMO channels is straightforward. In fact, the minimization of the total power is equivalent to the independent minimization of the power used at each carrier; hence, the original problem decouples into N subproblems like (2.42).

The design with a global QoS constraint of §2.4.3 can also be extended to N parallel MIMO channels, under the mild assumption that the quality of each carrier k is measured by an increasing function f_k and the global quality of the system, as measured by f, depends only on the f_k's:

$$
\begin{aligned}
&\underset{\{\mathbf{A}_k,\mathbf{B}_k,\alpha_k\}}{\text{minimize}} && \sum_{k=1}^{N} \text{Tr}\left(\mathbf{B}_k\mathbf{B}_k^H\right) \\
&\text{subject to} && f_k\left(\{\text{MSE}_{k,i}\}_{i=1}^{L_k}\right) \leq \alpha_k && 1 \leq k \leq N \\
& && f\left(\alpha_1,\cdots,\alpha_N\right) \leq \alpha_0
\end{aligned}
\tag{2.67}
$$

where the optimization is also over the α_k's that measure the quality of each carrier. If the α_k's are held fixed, problem (2.67) decouples into a set of N parallel optimization subproblems like (2.48), for which all the results of §2.4.3 apply. However, problem (2.67) has the additional difficulty that the α_k's have to be optimized as well. Such a problem can sometimes be directly solved as was done in [17] for some particular cases, obtaining more complicated solutions than the simple waterfillings expressions previously listed for Schur-concave/convex functions. Alternatively, the problem can also be easily tackled with a decomposition approach [13], by which (2.67) is conveniently decomposed into a set of parallel subproblems controlled by a master problem, c.f. [54].

Single Beamforming

A significant simplification occurs when a single substream is established per MIMO channel, i.e., when $L_k = 1$ $\forall k$. In such a case, the structure of each transmit matrix $\mathbf{B}_k$ is trivially given by a vector (or beamvector, hence the name single beamforming) $\mathbf{b}_k$ parallel to the eigenvector corresponding to the maximum eigenvalue at the kth MIMO channel and with squared norm p_k (from Theorems 2.4.1-2.4.3). Problem (2.67) reduces then to (see [65] for a comparison of several criteria)

$$
\begin{aligned}
&\underset{\{p_k,\alpha_k\}}{\text{minimize}} && \sum_{k=1}^{N} p_k \\
&\text{subject to} && f_k\left(\{\text{MSE}_k\}\right) \leq \alpha_k && 1 \leq k \leq N \\
& && f\left(\alpha_1,\cdots,\alpha_N\right) \leq \alpha_0.
\end{aligned}
\tag{2.68}
$$

2.4.5 Numerical Results

This section illustrates with numerical results the power of the tools developed for the design of linear MIMO transceivers in §2.4.2 and §2.4.3. Once the design criterion has been chosen, the transceiver is optimally designed using the general framework of Theorems 2.4.1-2.4.3.

A simple model has been used to randomly generate different realizations of the MIMO channel. In particular, the channel matrix $\mathbf{H}$ has been generated from a Gaussian distribution with i.i.d. zero-mean unit-variance elements, and the noise has been
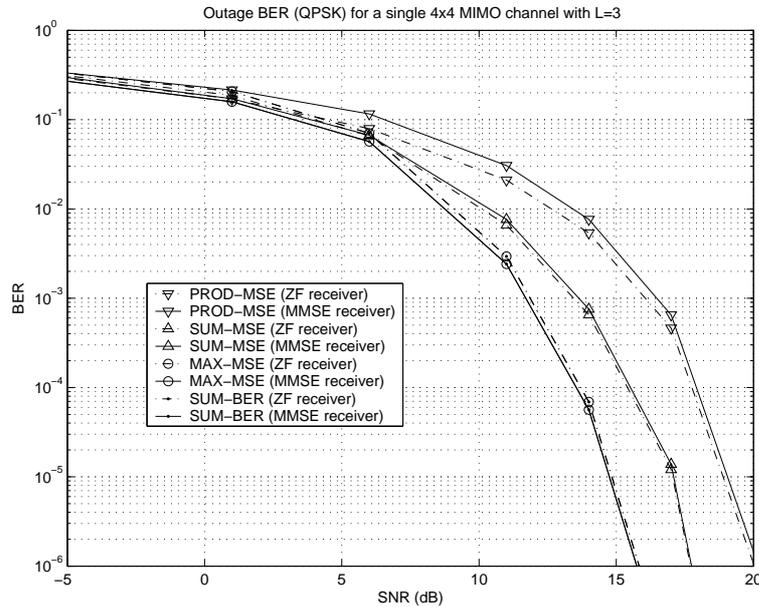
Figure 2.6: BER (at an outage probability of 5%) vs. SNR when using QPSK in a 4×4 MIMO channel with $L = 3$ (with MMSE and ZF receivers) for the methods: PROD-MSE, SUM-MSE, MAX-MSE, and SUM-BER.

modeled as white $\mathbf{R}_n = \sigma_n^2 \mathbf{I}$, where σ_n^2 is the noise power. (For simulations with more realistic wireless multi-antenna channel models including spatial and frequency correlation, the reader is referred to [15, 16, 17].) The SNR is defined as $\mathrm{SNR} = P_T/\sigma_n^2$, which is essentially a measure of the transmitted power normalized with respect to the noise.

In the first example, four different methods have been simulated by minimizing a cost function subject to a power constraint (recall that this is equivalent to minimizing the power subject to a global constraint as in §2.4.3): the classical minimization of the sum of the MSEs (SUM-MSE), the minimization of the product of the MSEs (PROD-MSE), the minimization of the maximum of the MSEs (MAX-MSE), and the minimization of the average/sum of the BERs (SUM-BER). The methods are evaluated in terms of BER averaged over the substreams; to be more precise, the outage BER[13] (over different realizations of $\mathbf{H}$) is considered since it is a more realistic measure than the average BER (which only makes sense when the system does not have delay constraints and the duration of the transmission is sufficiently long such that the fading statistics of the channel can be averaged out).

In Figure 2.6, the BER (for a QPSK constellation) is plotted as a function of

---

[13]The outage BER is the BER that is attained with some given probability (when it is not satisfied, an outage event is declared).
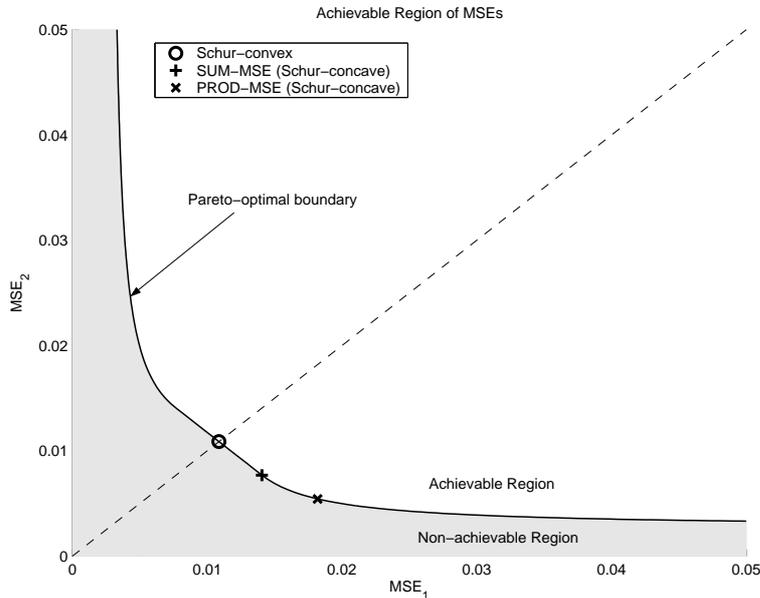
Figure 2.7: Achievable region of the MSEs for a given channel realization of a 4×4 MIMO channel with $L = 2$, along with the location of the design with the methods: PROD-MSE, SUM-MSE, and a Schur-convex (MAX-MSE and SUM-BER).

the SNR for a single 4×4 MIMO channel with $L = 3$ for the cases of ZF and MMSE receivers. It can be observed that the ZF receiver performs almost the same as the MMSE receiver thanks to the joint optimization of the transmitter and receiver (as opposed to the typically worse performance of the ZF receiver in the classical equalization setup where only the receiver is optimized). The methods MAX-MSE and SUM-BER correspond to Schur-convex functions and, as expected, are exactly the same (c.f. §2.4.3). The superiority of Schur-convex designs (MAX-MSE and SUM-BER) with respect to Schur-concave designs (SUM-MSE and PROD-MSE) is very clear from Figure 2.6, as was argued in §2.4.3, due to the increase robustness against fading of the channel eigenmodes.

In Figure 2.7, the achievable region in terms of MSEs is plotted for a given realization of a single 4×4 MIMO channel with $L = 2$ (MMSE receiver) and with an SNR of 15 dB. The achievable region has been computed with the method developed in §2.4.2 that allows to specify independent QoS constraints on each substream. The boundary between the achievable and non-achievable regions corresponds to the Pareto-optimal designs, characterized by not being outperformed by any other solution simultaneously in all substreams. The solutions corresponding to the previous methods, SUM-MSE, PROD-MSE, Schur-convex method (which includes MAX-MSE and SUM-BER), are also indicated. They clearly lie on the Pareto-optimal frontier,

although in different points of it. In fact, since Schur-convex methods have equal MSEs on all substreams, they all correspond to the intersection of the Pareto-optimal boundary with the line $\rho_1 = \rho_2 = \cdots = \rho_L$, which corresponds to a complete fairness among substreams.

2.5 An Application to Robust Transmitter Design in MIMO Channels

2.5.1 Introduction and State of the Art

In §2.4, a general framework has been presented to jointly design optimum linear transmitters and receivers according to a set of optimization criteria based on convex optimization theory. In those cases, it has been assumed that a perfect channel estimate or CSI is available during the design stage. In a realistic scenario, however, the channel knowledge is generally imperfect. In such a situation, the design should take into account explicitly the errors in the channel estimate, leading to the so called *robust designs*, which are less sensitive to these errors. It is interesting to note that the first applications of robust designs were not for wireless communications, but for control theory (see [66, 67] and references therein). Indeed, the concepts of signal state space and MIMO were originally used in that area. Afterwards, all these techniques and concepts were extended to other fields due to their potential benefits.

In some works such as [65, 68, 69, 70], the performance degradation of several non-robust designs for multi-antenna systems was analyzed, in which the errors in the CSI were considered negligible. The main conclusion is that this degradation increases rapidly with the error level.

In a communication system, the receiver usually acquires the channel estimate using a training sequence (pilot symbols). At the transmitter, the CSI can be obtained through a feedback channel or from previous received signals, exploiting the channel reciprocity principle in a time division duplexing (TDD) system (see [36] for an overview of different channel estimation strategies).

Different sources of errors in the CSI can be identified. In case of exploiting the channel reciprocity, the Gaussian noise from the estimation process and the outdated estimate due to the channel variability have to be considered. If a feedback channel is used, additional effects arise, such as the quantization of the channel estimate and the errors in the communication through the feedback channel.

According to the way the error in the channel estimate is modeled, the robust techniques can be classified into two families: the Bayesian (or stochastic) and the maximin (or worst-case) approaches [14, 17]. In the Bayesian philosophy, the statistics of the error are assumed to be known and a stochastic measure of the system performance is optimized, such as the mean value. On the other hand, the maximin approach considers that the error belongs to a predefined uncertainty region and the final objective is the optimization of the worst system performance for any error in this region.

The Bayesian philosophy has been considered in works such as [71], where a multi-antenna transmitter was designed to maximize the mean SNR and the mutual in-

formation assuming two sources of errors: the Gaussian noise and the quantization errors. The minimization of the BER was considered in [72]. Transmit FIR filters in a multi-antenna frequency-selective channel were designed in [73] to maximize the mean SNR and minimize the MSE assuming Gaussian errors. The more general case of MIMO flat fading channels was considered in [74], where the transmitter was composed of an orthogonal space-time block code (OSTBC) and a matrix performing a linear transformation. This matrix was designed to minimize an upper bound of the BER assuming Gaussian errors. A similar scheme consisting on the combination of an OSTBC a set of beamformers was considered in [75] to minimize the error probability. The same objective was taken in [76] and [77] for a MIMO frequency-selective channel using a multicarrier modulation: in [76], the transmitter and the receiver were based on matrices performing a linear transformation, whereas in [77], the transmitter was composed of an Alamouti's code [11] combined with two beamformers.

Regarding the maximin approach, [78] and [79] provide a general insight using a game theoretic formulation and describing several applications in signal processing. See also [18] for a reference on the theory of saddle-functions and maximin. This approach has been recently used in the classical problem of designing a receive beamformer under mismatches in the presumed model, as in [33], where the errors were assumed to be in the estimated steering vector and to belong to a spherical uncertainty region. This was afterwards generalized in [35] to embrace uncertainties both in the array response and the covariance matrix. The classical Capon's beamformer [31] was extended to its robust version in [32], [34], and [80], taking generic uncertainty regions and different formulations. In some of these examples, the robustness was obtained by minimizing the output power of the beamformer while guaranteeing a minimum gain for any direction modeled by the uncertainty region (see §2.2.5 for a more detailed description). Finally, several applications of this robust approach to multiuser systems with multi-antenna base stations can also be found in [36], [81], and [82].

This subsection starts with the presentation of the generic formulation of the Bayesian and the maximin approaches in §2.5.2. Afterwards, a MIMO system is considered, where the robust transmitter is designed under the maximin philosophy and the receiver is based on an optimum ML detector assuming a perfect channel knowledge. This follows and extends the results obtained in [83].

2.5.2   A Generic Formulation of Robust Approaches

A generic formulation can be stated for both the Bayesian and the maximin approaches (see [17] for more details). In all the cases, the imperfect CSI can be represented as

$$\mathbf{H} = \widehat{\mathbf{H}} + \boldsymbol{\Delta} \tag{2.69}$$

where $\mathbf{H}$ is the actual MIMO channel matrix (as defined in §2.3), $\widehat{\mathbf{H}}$ is the channel estimate, and $\boldsymbol{\Delta}$ is the error. The system performance is usually measured by a cost function f, whose minimization is the objective of the design (usual cost functions are based on the BER, the MSE, or the SINR, among others, c.f. §2.3.2).

In the Bayesian approach, the error and the actual channel are modeled statistically with the probability density functions (pdf's) $p_{\boldsymbol{\Delta}}(\boldsymbol{\Delta})$ and $p_{\mathbf{H}}(\mathbf{H})$, respectively,

which are assumed to be known. Note that knowing these pdf's is equivalent to knowing the pdf of the actual channel conditioned to the channel estimate $p_{\mathbf{H}|\widehat{\mathbf{H}}}\big(\mathbf{H}|\widehat{\mathbf{H}}\big)$, which is equal to $p_{\boldsymbol{\Delta}}\big(\mathbf{H} - \widehat{\mathbf{H}}\big)p_{\mathbf{H}}(\mathbf{H})/p_{\widehat{\mathbf{H}}}\big(\widehat{\mathbf{H}}\big)$ by the Bayes rule. A possible design strategy consists in the minimization of the mean value of the cost function f (note, however, that other stochastic measures of the performance could have been used, such as the outage performance). If this criterion is adopted and the average transmitted power is upper bounded by P_0, the following optimization problem has to be solved:

$$\begin{aligned} \underset{\mathbf{A},\mathbf{B}}{\text{minimize}} \quad & \mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\big[f\big(\mathbf{H},\mathbf{A},\mathbf{B}\big)\big] \\ \text{subject to} \quad & \text{Tr}\big(\mathbf{B}\mathbf{B}^H\big) \leq P_0 \end{aligned} \tag{2.70}$$

where the optimization variables are the transmit $\mathbf{B}$ and the receive $\mathbf{A}$ matrices. Note that, in this case, although the average performance is optimized, no guarantee can be given in terms of the instantaneous performance.

In the maximin approach, instead of modeling the error statistically, it is assumed that it belongs to a predefined uncertainty region $\mathcal{R}$, i.e., $\boldsymbol{\Delta} \in \mathcal{R}$. Then, the worst performance for any error in $\mathcal{R}$ is expressed as $\sup_{\boldsymbol{\Delta}\in\mathcal{R}} f\big(\mathbf{H},\mathbf{A},\mathbf{B}\big)$. The robust design problem consists in the optimization of the worst performance, which can be formulated as

$$\begin{aligned} \underset{\mathbf{A},\mathbf{B}}{\text{minimize}} \quad & \sup_{\boldsymbol{\Delta}\in\mathcal{R}} f\big(\mathbf{H},\mathbf{A},\mathbf{B}\big) \\ \text{subject to} \quad & \text{Tr}\big(\mathbf{B}\mathbf{B}^H\big) \leq P_0. \end{aligned} \tag{2.71}$$

In this case, a full statistical characterization is not necessary. Besides, this approach guarantees a minimum instantaneous performance for any error modeled by the uncertainty region, i.e., when the actual error behaves as expected (in a real situation, this will be satisfied with a high probability, declaring an outage otherwise). Note that this guarantee cannot be provided by the Bayesian approach optimizing the average performance.

2.5.3 Problem Formulation

As in §2.3 and §2.4, in the following, the transmission through a MIMO channel $\mathbf{H}$ is considered corresponding to the signal model $\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}$ (see (2.23)). In particular, a multi-antenna flat fading wireless channel is assumed, where the number of transmit and receive antennas is n_T and n_R, respectively. The interference-plus-noise covariance matrix is $\mathbf{R}_n = \sigma_n^2\mathbf{I}$, corresponding to a scenario with only additive white Gaussian noise (AWGN). Note, however, that the derived design strategy can be applied to other kinds of MIMO channels as well.

The objective is to obtain a maximin robust design of the system, as in (2.71), according to an imperfect channel estimate $\widehat{\mathbf{H}}$ at the transmitter as modeled in (2.69). The CSI at the receiver will be assumed to be perfect.

The Transmitter Architecture

Consider that one symbol is to be transmitted at one time instant. From §2.4 it is concluded that, in the case of having a perfect CSI, the optimum solution is based on

single beamforming ($L = 1$), consisting in the transmission through the eigenvector of $\mathbf{H}^H\mathbf{H}$ associated to the maximum eigenvalue (note that the squared whitened channel matrix is $\mathbf{R}_H = \frac{1}{\sigma_n^2}\mathbf{H}^H\mathbf{H}$). In case that the channel knowledge is imperfect, transmitting through the maximum eigenmode of $\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}$ constitutes the naive or non-robust solution, which may be quite sensitive to the errors in the CSI. Therefore, a robust design is expected to use more eigenmodes than the maximum one.

The design of the robust transmitter will be based on a linear processing, as in §2.3 ($\mathbf{s} = \mathbf{Bx}$), whereas at the receiver, a ML detector will be used assuming a perfect CSI. The direct design of a robust transmit matrix $\mathbf{B}$ seems very complicated and, therefore, a structure will be imposed on it to simplify the problem. The proposed robust solution consists in the transmission of the symbols through all the eigenmodes of $\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}$ using an adequate power distribution among them, as will be shown later and as opposed to the non-robust design, where only the maximum eigenmode is used.

Consider the simultaneous transmission of R independent complex symbols over N_t periods of time (leading to a transmission rate equal to R/N_t), corresponding to the following linear signal model, similarly to linear dispersion codes [84] and OSTBC [85, 86, 87]:

$$\mathbf{S} = \sum_{l=1}^{R}\left(\mathbf{B}_l^{(r)}x_l^{(r)} + j\mathbf{B}_l^{(i)}x_l^{(i)}\right) \in \mathbb{C}^{n_T \times N_t} \qquad (2.72)$$

where x_l is the lth transmitted complex symbol with a normalized energy ($\mathbb{E}\left[|x_l|^2\right] = 1$), $x_l^{(r)}$ and $x_l^{(i)}$ are its real and imaginary parts, respectively, $\mathbf{B}_l^{(r)}, \mathbf{B}_l^{(i)} \in \mathbb{C}^{n_T \times N_t}$ are the associated complex transmit matrices for $x_l^{(r)}$ and $x_l^{(i)}$, and the rows of $\mathbf{S}$ are the N_t signal samples transmitted through each antenna. The structure imposed on the transmit matrices is

$$\mathbf{B}_l^{(r)} = \widehat{\mathbf{U}}\operatorname{diag}\left(\{\sqrt{p_i}\}\right)\mathbf{T}_l^{(r)}, \quad \mathbf{B}_l^{(i)} = \widehat{\mathbf{U}}\operatorname{diag}\left(\{\sqrt{p_i}\}\right)\mathbf{T}_l^{(i)} \qquad (2.73)$$

where $\widehat{\mathbf{U}} = \left[\widehat{\mathbf{u}}_1 \cdots \widehat{\mathbf{u}}_{n_T}\right] \in \mathbb{C}^{n_T \times n_T}$ is the unitary matrix containing the n_T eigenvectors of $\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}$ with eigenvalues $\{\widehat{\lambda}_i\}$ sorted in decreasing order, p_i is the power allocated to the ith estimated eigenmode, and $\mathbf{T}_l^{(r)}, \mathbf{T}_l^{(i)} \in \mathbb{C}^{n_T \times N_t}$ are matrices modeling a temporal spreading of the symbols and fulfilling $\mathbf{T}_l^{(r)}\mathbf{T}_l^{(r)H} = \mathbf{I}$ and $\mathbf{T}_l^{(i)}\mathbf{T}_l^{(i)H} = \mathbf{I}$. These matrices $\mathbf{T}_l^{(r)}$ and $\mathbf{T}_l^{(i)}$ are based on the Hurwitz-Radon family of matrices (see [86] and [85]), so that the ML detector reduces to a bank of linear filters, and the detection scheme is the same as that used for OSTBC.

This apparently complicated signaling scheme (2.72) can be represented as shown in Figure 2.8, where it is seen that the symbols are encoded by an OSTBC, and each output of the OSTBC is transmitted through a different estimated eigenmode. Finally, the parameters $\{p_i\}$ distribute the available power among the eigenmodes. This transmission scheme is similar to those presented in [74, 75, 88], among others, in which the transmitter architecture also consisted in the combination of an OSTBC and a beamforming stage, although the design of the power allocation was different to the one proposed in the following.
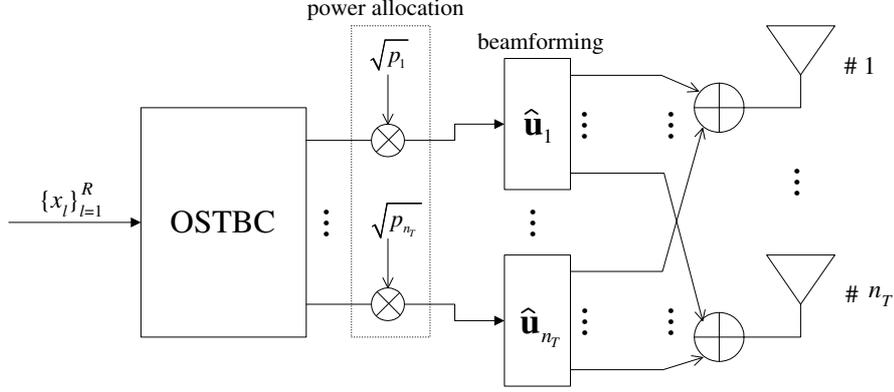
Figure 2.8: General architecture for the transmitter based on the combination of an OSTBC block, a power allocation, and a beamforming stage.

Note that, in the signal model (2.72), not only the spatial, but also the temporal dimensions are exploited. This signal model can be rewritten using the generic matrix-vector notation presented in §2.3 as follows. Let $\mathbf{s}_n$ be the nth column of $\mathbf{S}$ representing the transmitted vector during the nth period of time. This vector can be expressed as

$$\mathbf{s}_n = \overline{\mathbf{B}}_n^{(r)}\mathbf{x}_r + j\overline{\mathbf{B}}_n^{(i)}\mathbf{x}_i \qquad (2.74)$$

where $\overline{\mathbf{B}}_n^{(r)} = \widehat{\mathbf{U}}\operatorname{diag}\left(\{\sqrt{p_i}\}\right)\overline{\mathbf{T}}_n^{(r)}$ and $\overline{\mathbf{B}}_n^{(i)} = \widehat{\mathbf{U}}\operatorname{diag}\left(\{\sqrt{p_i}\}\right)\overline{\mathbf{T}}_n^{(i)}$, the matrices $\overline{\mathbf{T}}_n^{(r)}$ and $\overline{\mathbf{T}}_n^{(i)}$ contain the nth columns of $\mathbf{T}_l^{(r)}$ and $\mathbf{T}_l^{(i)}$ for $l = 1, \ldots, R$, and $\mathbf{x}_r = \left[x_1^{(r)} \cdots x_R^{(r)}\right]^T$, $\mathbf{x}_i = \left[x_1^{(i)} \cdots x_R^{(i)}\right]^T$.

The design objective is to calculate the optimum power allocation strategy $\{p_i\}$ subject to a transmit power constraint and adopting an adequate performance criterion. If the transmit power budget is P_0, the power constraint can be expressed in terms of the factors $\{p_i\}$ as

$$\frac{1}{2}\left(\|\mathbf{B}_l^{(r)}\|_F^2 + \|\mathbf{B}_l^{(i)}\|_F^2\right) = \sum_{i=1}^{n_T} p_i \leq P_0, \quad p_i \geq 0. \qquad (2.75)$$

Note that the set of feasible power distributions is convex, since all the constraints given above are linear.

For the considered system ((2.72) and (2.73)) using an OSTBC with ML detection, the performance can be measured by the SNR expressed as (see [86] and [85])

$$\mathrm{SNR} = \frac{1}{\sigma_n^2}\operatorname{Tr}\left(\widehat{\mathbf{U}}^H\mathbf{H}^H\mathbf{H}\widehat{\mathbf{U}}\operatorname{diag}(\mathbf{p})\right) \qquad (2.76)$$

where $\mathbf{p} = \begin{bmatrix} p_1 \cdots p_{n_T} \end{bmatrix}^T$. Based on this, the performance function f is defined as

$$f(\mathbf{p}, \mathbf{\Delta}) = \mathrm{Tr}\left(\widehat{\mathbf{U}}^H \mathbf{H}^H \mathbf{H} \widehat{\mathbf{U}} \, \mathrm{diag}(\mathbf{p})\right) \tag{2.77}$$

$$= \mathrm{Tr}\left(\widehat{\mathbf{U}}^H \big(\widehat{\mathbf{H}} + \mathbf{\Delta}\big)^H \big(\widehat{\mathbf{H}} + \mathbf{\Delta}\big) \widehat{\mathbf{U}} \, \mathrm{diag}(\mathbf{p})\right), \tag{2.78}$$

whose maximization is the objective of the design and where the error model (2.69) has been used. This function is linear in $\mathbf{p}$ and, therefore, concave; and convex-quadratic in the error $\mathbf{\Delta}$. Note that, in this case, an opposite strategy is taken from that presented in §2.4, in which the transmitted power was minimized subject to QoS constraints, although both kinds of problems are essentially equivalent, as explained in §2.4.1.

The Maximin Problem Formulation

As stated in §2.5.2, the maximin approach can be used to include robustness in the design. Accordingly, an uncertainty region $\mathcal{R}$ for $\mathbf{\Delta}$ has to be defined, which, in the following, will be assumed to be convex.[14] The robust power distribution $\mathbf{p}^\star$, which optimizes the worst performance for any error in the uncertainty region, can be found as the solution to the following optimization problem:

$$\begin{array}{ll} \underset{\mathbf{p}}{\mathrm{maximize}} & \underset{\mathbf{\Delta} \in \mathcal{R}}{\inf} f(\mathbf{p}, \mathbf{\Delta}) \\ \mathrm{subject\ to} & \mathbf{1}^T \mathbf{p} \leq P_0 \\ & p_i \geq 0 \end{array} \tag{2.79}$$

where $\mathbf{1} = \begin{bmatrix} 1 \cdots 1 \end{bmatrix}^T \in \mathbb{R}^{n_T \times 1}$ is the all-one vector.

The direct way to solve the maximin problem is to obtain the minimization of f with respect to $\mathbf{\Delta}$ in an analytical way and then solve the outer maximization. Such an approach, however, is difficult because it is not clear what is the minimizing $\mathbf{\Delta}$ for a given $\mathbf{p}$ in closed form.

Of course, one can also consider solving the problem numerically, i.e., solving the inner minimization

$$\widetilde{f}(\mathbf{p}) = \underset{\mathbf{\Delta} \in \mathcal{R}}{\inf} f(\mathbf{p}, \mathbf{\Delta}) \tag{2.80}$$

numerically for a given $\mathbf{p}$, and then solving the outer maximization $\max_{\mathbf{p}} \widetilde{f}(\mathbf{p})$ also numerically. Note that the inner minimization is a convex problem, since f is convex in $\mathbf{\Delta}$ and the constraint set $\mathcal{R}$ is also convex. The outer maximization is also a convex problem since the constraint set for $\mathbf{p}$ is convex and the function $\widetilde{f}$ is concave. Consequently, the numerical algorithms referenced in §2.2.4 could be used. The proof that $\widetilde{f}$ is concave is given below, where $\mathbf{p}_1$ and $\mathbf{p}_2$ are any feasible power distributions,

---

[14]The set $\mathcal{R}$ is convex if $\mathbf{\Delta} = \theta \mathbf{\Delta}_1 + (1 - \theta)\mathbf{\Delta}_2 \in \mathcal{R}$ for any $\mathbf{\Delta}_1, \mathbf{\Delta}_2 \in \mathcal{R}$ and $\theta \in [0, 1]$.

and $\theta \in [0,1]$:

$$
\begin{aligned}
\widetilde{f}(\theta \mathbf{p}_1 + (1-\theta)\mathbf{p}_2) &= \inf_{\boldsymbol{\Delta} \in \mathcal{R}} f(\theta \mathbf{p}_1 + (1-\theta)\mathbf{p}_2, \boldsymbol{\Delta}) \\
&= \inf_{\boldsymbol{\Delta} \in \mathcal{R}} [\theta f(\mathbf{p}_1, \boldsymbol{\Delta}) + (1-\theta)f(\mathbf{p}_2, \boldsymbol{\Delta})] \\
&\geq \theta \inf_{\boldsymbol{\Delta} \in \mathcal{R}} f(\mathbf{p}_1, \boldsymbol{\Delta}) + (1-\theta) \inf_{\boldsymbol{\Delta} \in \mathcal{R}} f(\mathbf{p}_2, \boldsymbol{\Delta}) \\
&= \theta \widetilde{f}(\mathbf{p}_1) + (1-\theta)\widetilde{f}(\mathbf{p}_2)
\end{aligned}
\tag{2.81}
$$

where the linearity of f in $\mathbf{p}$ has been used in the second equality. This numerical procedure guarantees that the desired solution $\mathbf{p}^\star$ is found, however, it is computationally costly because each iteration of the method for the outer maximization requires an evaluation of $\widetilde{f}(\mathbf{p})$ (and possibly its gradient as well), which in turn requires solving the inner minimization numerically with as many iterations as needed to converge.

Other kinds of numerical iterative methods could be used, such as the algorithm proposed in [89] to find saddle-points of maximin problems based on the method of steepest descent. In [90], an alternative algorithm for the same problem is derived based on the interior point approach.

2.5.4   Reformulating the Problem in a Simplified Convex Form

The objective is to rewrite the maximin problem (2.79), which is already convex but not amenable for efficient resolution, into an equivalent simplified convex problem so that it can be solved requiring less computational effort (see §2.2).

The function $f(\mathbf{p}, \boldsymbol{\Delta})$, which is concave-convex, and the optimization sets for the power distribution and the error satisfy the conditions given in Corollary 37.6.2 in [18]. From it, it can be concluded that there exists a saddle-point of the maximin problem, i.e., there exist $\mathbf{p}^\star$ and $\boldsymbol{\Delta}^\star$ fulfilling the constraints and satisfying

$$
f(\mathbf{p}, \boldsymbol{\Delta}^\star) \leq f(\mathbf{p}^\star, \boldsymbol{\Delta}^\star) \leq f(\mathbf{p}^\star, \boldsymbol{\Delta})
\tag{2.82}
$$

for any feasible $\mathbf{p}$ and $\boldsymbol{\Delta}$. The solution to the original maximin problem (2.79) can be shown to be $\mathbf{p}^\star$, and the saddle-value $f^\star \triangleq f(\mathbf{p}^\star, \boldsymbol{\Delta}^\star)$ is the minimum value of f given $\mathbf{p}^\star$, i.e., $\widetilde{f}(\mathbf{p}^\star)$ (see Lemma 36.2 in [18]). The existence of the saddle-point allows to rewrite the original maximin problem (2.79) using a minimax formulation, i.e., the inner and the outer optimizations can be interchanged:

$$
\begin{aligned}
&\underset{\boldsymbol{\Delta}}{\text{minimize}} && \sup_{\mathbf{1}^T \mathbf{p} \leq P_0, p_i \geq 0} f(\mathbf{p}, \boldsymbol{\Delta}) \\
&\text{subject to} && \boldsymbol{\Delta} \in \mathcal{R},
\end{aligned}
\tag{2.83}
$$

with the advantage that the inner maximization is now very simple. In particular, $\sup_{\mathbf{p}} f(\mathbf{p}, \boldsymbol{\Delta})$ gives as a result the maximum element of the diagonal of the matrix $\widehat{\mathbf{U}}^H(\widehat{\mathbf{H}} + \boldsymbol{\Delta})^H(\widehat{\mathbf{H}} + \boldsymbol{\Delta})\widehat{\mathbf{U}}$ multiplied by the power budget P_0, i.e.:

$$
\sup_{\mathbf{1}^T \mathbf{p} \leq P_0, p_i \geq 0} f(\mathbf{p}, \boldsymbol{\Delta}) = P_0 \max_i \left[\widehat{\mathbf{U}}^H(\widehat{\mathbf{H}} + \boldsymbol{\Delta})^H(\widehat{\mathbf{H}} + \boldsymbol{\Delta})\widehat{\mathbf{U}} \right]_{ii}
\tag{2.84}
$$

$$
= P_0 \max_i \left\{ \widehat{\mathbf{u}}_i^H(\widehat{\mathbf{H}} + \boldsymbol{\Delta})^H(\widehat{\mathbf{H}} + \boldsymbol{\Delta})\widehat{\mathbf{u}}_i \right\}.
\tag{2.85}
$$

Note that the power allocation $\mathbf{p}$ achieving this optimum value is not unique if the maximum value is attained by more than one element of the diagonal of the matrix $\widehat{\mathbf{U}}^H\big(\widehat{\mathbf{H}} + \boldsymbol{\Delta}\big)^H\big(\widehat{\mathbf{H}} + \boldsymbol{\Delta}\big)\widehat{\mathbf{U}}$.

As a consequence from the previous result, the original problem can now be written as the following simple convex minimization problem:

$$
\begin{aligned}
\underset{t,\boldsymbol{\Delta}}{\text{minimize}} \quad & t \\
\text{subject to} \quad & t \geq P_0\widehat{\mathbf{u}}_i^H\big(\widehat{\mathbf{H}} + \boldsymbol{\Delta}\big)^H\big(\widehat{\mathbf{H}} + \boldsymbol{\Delta}\big)\widehat{\mathbf{u}}_i \qquad \forall i \\
& \boldsymbol{\Delta} \in \mathcal{R}.
\end{aligned}
\qquad (2.86)
$$

Solving the previous problem gives the saddle-value $f^\star = f(\mathbf{p}^\star, \boldsymbol{\Delta}^\star) = t^\star$ and the worst-case error $\boldsymbol{\Delta}^\star$ of the saddle-point of the problem (see [18]); however, the optimal robust power distribution is still unknown. It turns out that the optimum Lagrange multipliers $\gamma_i^\star$ associated with the constraints $t \geq P_0\widehat{\mathbf{u}}_i^H\big(\widehat{\mathbf{H}} + \boldsymbol{\Delta}\big)^H\big(\widehat{\mathbf{H}} + \boldsymbol{\Delta}\big)\widehat{\mathbf{u}}_i$ in problem (2.86) provide the optimum normalized power distribution $p_i^\star$, i.e., $p_i^\star = P_0\gamma_i^\star$, as proved below.

The problem (2.86) can be solved by formulating the KKT conditions, which are satisfied by the worst-case error $\boldsymbol{\Delta}^\star$ along with the optimum dual variables (see §2.2.3). On the other hand, it is clear that the worst-case error $\boldsymbol{\Delta}^\star$ is also the solution to the problem $\inf_{\boldsymbol{\Delta}}(\mathbf{p}^\star, \boldsymbol{\Delta})$ (from (2.82)), where $\mathbf{p}^\star = \arg\max_{\mathbf{p}} f(\mathbf{p}, \boldsymbol{\Delta}^\star)$ is the robust power distribution, and, therefore, the worst-case error $\boldsymbol{\Delta}^\star$ must satisfy the KKT conditions for the problem $\inf_{\boldsymbol{\Delta}} f(\mathbf{p}^\star, \boldsymbol{\Delta})$. By a simple comparison of the KKT conditions for both problems, it is straightforward to see that for $p_i^\star = P_0\gamma_i^\star$, the worst-case error $\boldsymbol{\Delta}^\star$ satisfies both sets of KKT conditions and, hence, that is an optimal robust power allocation.

The Lagrangian of the optimization problem (2.86) (characterizing for convenience and w.l.o.g. the uncertainty convex region $\mathcal{R}$ as the intersection of a set of convex constraints of the form $f_i(\boldsymbol{\Delta}) \leq 0$) is

$$
\begin{aligned}
L_1 &= t + \sum_{i=1}^{n_T} \gamma_i \left(P_0\widehat{\mathbf{u}}_i^H\big(\widehat{\mathbf{H}} + \boldsymbol{\Delta}\big)^H\big(\widehat{\mathbf{H}} + \boldsymbol{\Delta}\big)\widehat{\mathbf{u}}_i - t \right) + \sum_i \mu_i f_i(\boldsymbol{\Delta}) \\
&= t \left(1 - \sum_{i=1}^{n_T} \gamma_i \right) + P_0 \operatorname{Tr}\left(\big(\widehat{\mathbf{H}} + \boldsymbol{\Delta}\big)^H\big(\widehat{\mathbf{H}} + \boldsymbol{\Delta}\big)\widehat{\mathbf{U}} \operatorname{diag}(\{\gamma_i\})\widehat{\mathbf{U}}^H \right) + \sum_i \mu_i f_i(\boldsymbol{\Delta}),
\end{aligned}
\qquad (2.87)
$$

where the relation $\sum_{i=1}^{n_T} \gamma_i\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i^H = \widehat{\mathbf{U}} \operatorname{diag}(\{\gamma_i\})\widehat{\mathbf{U}}^H$ has been used. The KKT conditions for this problem are

$$
f_i(\boldsymbol{\Delta}^\star) \leq 0, \quad t^\star \geq P_0\widehat{\mathbf{u}}_i^H\big(\widehat{\mathbf{H}} + \boldsymbol{\Delta}^\star\big)^H\big(\widehat{\mathbf{H}} + \boldsymbol{\Delta}^\star\big)\widehat{\mathbf{u}}_i, \qquad (2.88)
$$

$$
\mu_i^\star \geq 0, \quad \gamma_i^\star \geq 0, \qquad (2.89)
$$

$$
\sum_{i=1}^{n_T} \gamma_i^\star = 1, \quad P_0\big(\widehat{\mathbf{H}} + \boldsymbol{\Delta}^\star\big)\widehat{\mathbf{U}} \operatorname{diag}(\{\gamma_i^\star\})\widehat{\mathbf{U}}^H + \sum_i \mu_i^\star \nabla f_i(\boldsymbol{\Delta}^\star) = \mathbf{0}, \qquad (2.90)
$$

$$
\mu_i^\star f_i(\boldsymbol{\Delta}^\star) = 0, \quad \gamma_i^\star \left(P_0\widehat{\mathbf{u}}_i^H\big(\widehat{\mathbf{H}} + \boldsymbol{\Delta}^\star\big)^H\big(\widehat{\mathbf{H}} + \boldsymbol{\Delta}^\star\big)\widehat{\mathbf{u}}_i - t^\star \right) = 0. \qquad (2.91)
$$

On the other hand, the Lagrangian for the problem $\inf_{\boldsymbol{\Delta}} f(\mathbf{p}^\star, \boldsymbol{\Delta})$ is

$$L_2 = \mathrm{Tr}\left(\widehat{\mathbf{U}}^H(\widehat{\mathbf{H}} + \boldsymbol{\Delta})^H(\widehat{\mathbf{H}} + \boldsymbol{\Delta})\widehat{\mathbf{U}}\,\mathrm{diag}(\mathbf{p}^\star)\right) + \sum_i \alpha_i f_i(\boldsymbol{\Delta}) \qquad (2.92)$$

and the KKT conditions for the optimal error and multipliers are:

$$f_i(\boldsymbol{\Delta}^\star) \leq 0, \qquad\qquad (2.93)$$

$$\alpha_i^\star \geq 0, \qquad\qquad (2.94)$$

$$(\widehat{\mathbf{H}} + \boldsymbol{\Delta}^\star)\widehat{\mathbf{U}}\,\mathrm{diag}(\mathbf{p}^\star)\widehat{\mathbf{U}}^H + \sum_i \alpha_i^\star \nabla f_i(\boldsymbol{\Delta}^\star) = \mathbf{0}, \qquad\qquad (2.95)$$

$$\alpha_i^\star f_i(\boldsymbol{\Delta}^\star) = 0. \qquad\qquad (2.96)$$

From the comparison of both sets of KKT conditions (2.88)-(2.91) and (2.93)-(2.96), it is clear that they are satisfied by the same worst-case error $\boldsymbol{\Delta}^\star$ if $\alpha_i^\star = \mu_i^\star$ and $p_i^\star = P_0 \gamma_i^\star$. Note that the transmit power constraints are automatically fulfilled, since the optimum dual variables $\{\gamma_i^\star\}$ are required to satisfy $\gamma_i^\star \geq 0$ (see (2.89)) and $\sum_{i=1}^{n_T} \gamma_i^\star = 1$ (see (2.90)).

It is important to remark that the saddle-value $t^\star$ can be attained by more than one element of the diagonal of the matrix $P_0 \widehat{\mathbf{U}}^H(\widehat{\mathbf{H}} + \boldsymbol{\Delta}^\star)^H(\widehat{\mathbf{H}} + \boldsymbol{\Delta}^\star)\widehat{\mathbf{U}}$. Taking this into account, and using the condition $\gamma_i^\star\left(P_0\widehat{\mathbf{u}}_i^H(\widehat{\mathbf{H}} + \boldsymbol{\Delta}^\star)^H(\widehat{\mathbf{H}} + \boldsymbol{\Delta}^\star)\widehat{\mathbf{u}}_i - t^\star\right) = 0$ in (2.91), it is concluded that there can exist more than one Lagrange multipler $\gamma_i^\star$ that can be different from zero; in other words, the robust power distribution can use more than one estimated eigenmode for transmission, as expected. This will be proved with some examples in the simulations section.

Summarizing, the original maximin robust power allocation problem (2.79) can be solved by means of the simplified convex problem (2.86). The values of the optimum Lagrange multipliers of this simplified problem provide the normalized power distribution to be applied among the estimated eigenmodes. Currently, most of the existing software packages provide simultaneously the optimum values of both the primal and the dual variables and, therefore, they could be applied to find the optimum solution to this robustness problem (see §2.2.4 and references therein).

2.5.5   Convex Uncertainty Regions

The definition of the uncertainty region $\mathcal{R}$ may impact importantly on the system performance. The size and the shape of this region should take into account the quality of the channel estimate and the imperfections that generate the error (see §2.5.1 for some examples of sources of errors).

In the following, two sources of errors are identified and three different uncertainty regions, jointly with their sizes, are described. In all the cases, the proposed uncertainty regions are convex, as required to solve the optimization problem.

Estimation Gaussian Noise

A common situation when estimating the channel corresponds to the presence of AWGN, especially in TDD systems, where the transmitter can estimate the channel
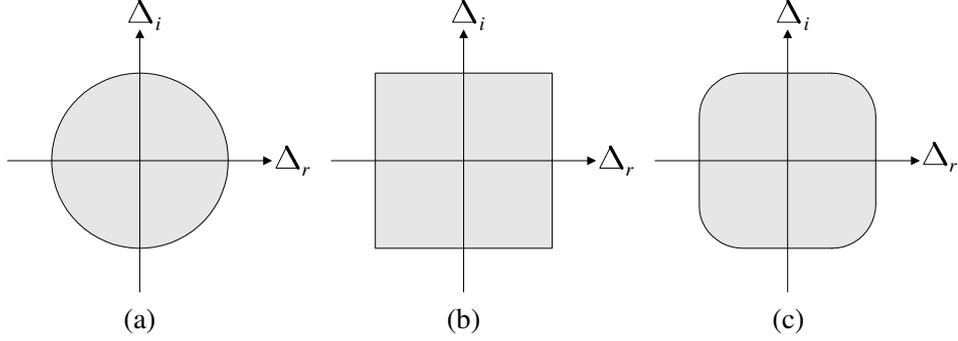
Figure 2.9: Different uncertainty regions for the case of a scalar error Δ, where $\Delta_r = \mathrm{Re}\left\{\Delta\right\}$ and $\Delta_i = \mathrm{Im}\left\{\Delta\right\}$. (a) estimation Gaussian noise, (b) quantization errors, and (c) combined estimation and quantization errors.

while receiving in the reverse link, and use it as an estimate in the forward link, due to the channel reciprocity principle.

The following assumptions are considered: all the components of the actual channel $\mathbf{H}$ are i.i.d. and follow a zero-mean circularly symmetric Gaussian distribution with a variance equal to σ_h^2, whereas the estimation AWGN is also zero-mean and circularly symmetric with variance σ_e^2. Based on this, the estimation SNR, i.e., the received SNR during the transmission of the training sequence, is defined as $\mathrm{SNR}_{\mathrm{est}} = \sigma_h^2/\sigma_e^2$.

Under these assumptions, the actual channel $\mathbf{H}$ conditioned to the channel estimate follows a circulary symmetric complex Gaussian distribution with a mean value equal to the MMSE Bayesian channel estimate and a white covariance matrix with diagonal elements equal to $\frac{\sigma_h^2 \sigma_e^2}{\sigma_h^2 + \sigma_e^2} = \frac{\sigma_h^2}{1+\mathrm{SNR}_{\mathrm{est}}}$ (see [55]), i.e., the actual channel can be assumed to be in a region near the MMSE Bayesian channel estimate, where the distance to it is measured indirectly by $\frac{\sigma_h^2}{1+\mathrm{SNR}_{\mathrm{est}}}$. Therefore, in this case, it is natural to identify $\widehat{\mathbf{H}}$ as the MMSE Bayesian estimate of the MIMO channel and to define the uncertainty region $\mathcal{R}$ as a sphere centered at $\widehat{\mathbf{H}}$ with a radius equal to $\sqrt{\epsilon}$ as follows:

$$\mathcal{R} = \left\{ \mathbf{\Delta} : \ \|\mathbf{\Delta}\|_F^2 \leq \epsilon \right\}. \tag{2.97}$$

Since the error $\mathbf{\Delta}$ is Gaussian distributed, it will be inside the region $\mathcal{R}$ with a certain probability P_{in} lower than 1. This probability will be equal to the probability of providing the required QoS to the user. The mathematical relationship between the size of the uncertainty region, measured by ϵ, and P_{in} is given be $\epsilon = \varphi^{-1}\left(P_{\mathrm{in}}\right)$, where the function φ is the cumulative density function of the chi-square distribution corresponding to $\|\mathbf{\Delta}\|_F^2$ with $2n_R n_T$ degrees of freedom and normalized variance $\frac{\sigma_h^2}{2+2\mathrm{SNR}_{\mathrm{est}}}$.

Quantization Errors

In frequency division duplexing (FDD) systems, the estimate of the channel at the transmitter is typically obtained through a feedback channel from the receiver to the transmitter. Since this feedback is expected to be discrete, the channel estimate has to be quantized introducing an error in the CSI available at the transmitter. Assuming that the receiver has a perfect knowledge of the channel response $\mathbf{H}$, let us consider, for illustrative purposes, that a suboptimum uniform quantization is applied to the real and imaginary parts of all the components of $\mathbf{H}$ using a quantization step equal to Δ_q, obtaining $\widehat{\mathbf{H}}$. In this situation, the quantization SNR is given by $\mathrm{SNR}_q = 6\sigma_h^2/\Delta_q^2$. Consequently, the uncertainty region can be defined as a cube centered at $\widehat{\mathbf{H}}$ as

$$\mathcal{R} = \left\{ \boldsymbol{\Delta} : \left|\mathrm{Re}\big\{[\boldsymbol{\Delta}]_{ij}\big\}\right| \leq \frac{\Delta_q}{2}, \left|\mathrm{Im}\big\{[\boldsymbol{\Delta}]_{ij}\big\}\right| \leq \frac{\Delta_q}{2} \right\}. \tag{2.98}$$

As the capacity of the feedback channel increases, more bits can be used in the quantization and, therefore, the size of the uncertainty region can be reduced.

Combined Estimation and Quantization Errors

In a realistic scenario with feedback, the two effects considered previously, i.e., the presence of AWGN and the quantization errors, are expected to be combined. This can be modeled mathematically by defining an appropriate uncertainty region as

$$\mathcal{R} = \left\{ \begin{array}{l} \boldsymbol{\Delta} = \boldsymbol{\Delta}_1 + \boldsymbol{\Delta}_2 : \|\boldsymbol{\Delta}_1\|_F^2 \leq \epsilon, \\ \left|\mathrm{Re}\big\{[\boldsymbol{\Delta}_2]_{ij}\big\}\right| \leq \frac{\Delta_q}{2}, \left|\mathrm{Im}\big\{[\boldsymbol{\Delta}_2]_{ij}\big\}\right| \leq \frac{\Delta_q}{2} \end{array} \right\}, \tag{2.99}$$

which is convex because it is described by linear and norm constraints [14]. With this region, the optimization problem (2.86) can be rewritten as the following quadratic convex minimization problem:

$$\begin{array}{ll} \underset{t,\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2}{\mathrm{minimize}} & t \\ \mathrm{subject\ to} & t \geq P_0\widehat{\mathbf{u}}_i^H\big(\widehat{\mathbf{H}} + \boldsymbol{\Delta}_1 + \boldsymbol{\Delta}_2\big)^H\big(\widehat{\mathbf{H}} + \boldsymbol{\Delta}_1 + \boldsymbol{\Delta}_2\big)\widehat{\mathbf{u}}_i \qquad \forall i \\ & \mathrm{Tr}\big(\boldsymbol{\Delta}_1^H\boldsymbol{\Delta}_1\big) \leq \epsilon \\ & \left|\mathrm{Re}\big\{[\boldsymbol{\Delta}_2]_{ij}\big\}\right| \leq \frac{\Delta_q}{2}, \left|\mathrm{Im}\big\{[\boldsymbol{\Delta}_2]_{ij}\big\}\right| \leq \frac{\Delta_q}{2}, \end{array} \tag{2.100}$$

which comprises the previous uncertainty regions and the corresponding optimization problems as particular cases.

Figure 2.9 illustrates the shape of the three considered uncertainty regions for the concrete case of a scalar error Δ, where $\Delta_r = \mathrm{Re}\{\Delta\}$ and $\Delta_i = \mathrm{Im}\{\Delta\}$.

2.5.6 Numerical Results

This section provides some numerical results to illustrate the performance of the proposed robust design when compared to other known transmission techniques.

The robust design takes into account the uncertainty in the channel knowledge to find an optimum power distribution among the estimated eigenmodes. When the
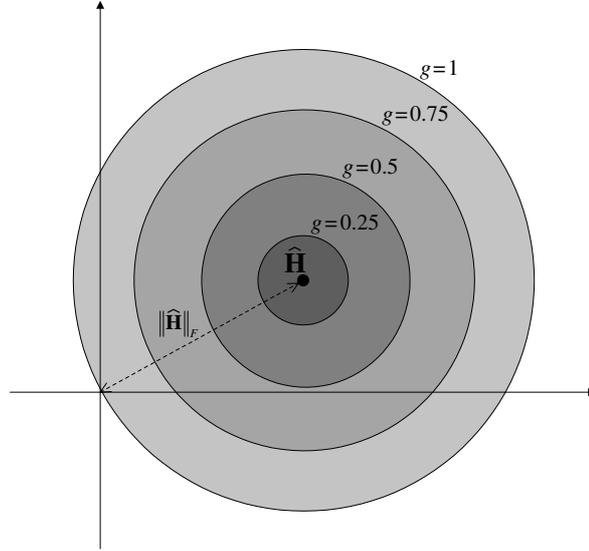
Figure 2.10 Spherical uncertainty regions for different values of the parameter g.

channel estimate is perfect, the robust solution must lead to the same power distribution as for the non-robust beamforming. As stated in §2.5.3, the non-robust design consists in transmitting all the symbols only through the maximum estimated eigenmode, which can be formulated in terms of the power distribution as $p_1 = P_0$, $p_i = 0$, $i = 2, \ldots, n_T$, i.e., all the power is allocated to the maximum eigenmode. When the channel uncertainty increases, the robust design tends to distribute the power in a more uniform way, increasing the power for the weaker eigenmodes.

In the first example, a system with 4 transmit and 6 receive antennas is studied. Spherical uncertainty regions with a radius equal to $\sqrt{\epsilon} = g\|\widehat{\mathbf{H}}\|_F$, $0 < g \leq 1$ are considered. Note that for these uncertainty regions, $\mathbf{H} = \widehat{\mathbf{H}} + \boldsymbol{\Delta} \neq \mathbf{0}$, $\forall \boldsymbol{\Delta} \in \mathcal{R}$. This condition is imposed since, otherwise, the saddle-value would be equal to 0. In Figure 2.10, the spherical uncertainty regions for different sizes are represented. Since $n_T = 4$, the total transmitted power has to be distributed among the 4 estimated eigenmodes. Figure 2.11 shows the mean value of the normalized robust power distribution given by $\{\gamma_i^\star\}$, that is, assuming $P_0 = 1$. As it is seen, for $g = 0$ the power distribution corresponds to the non-robust approach, as expected. As g increases, the errors in the channel estimate are higher and the power allocation profile changes so that the power is distributed in a more uniform way.

Figures 2.12 and 2.13 show some simulations results in order to compare three different techniques: the proposed maximin robust approach, the non-robust technique, and a pure OSTBC in which no CSI is available at the transmitter. The techniques are compared in terms of the minimum required transmitted power so that the resulting SNR is higher than a target $\text{SNR}_0 = 10$ dB for any error in the uncertainty
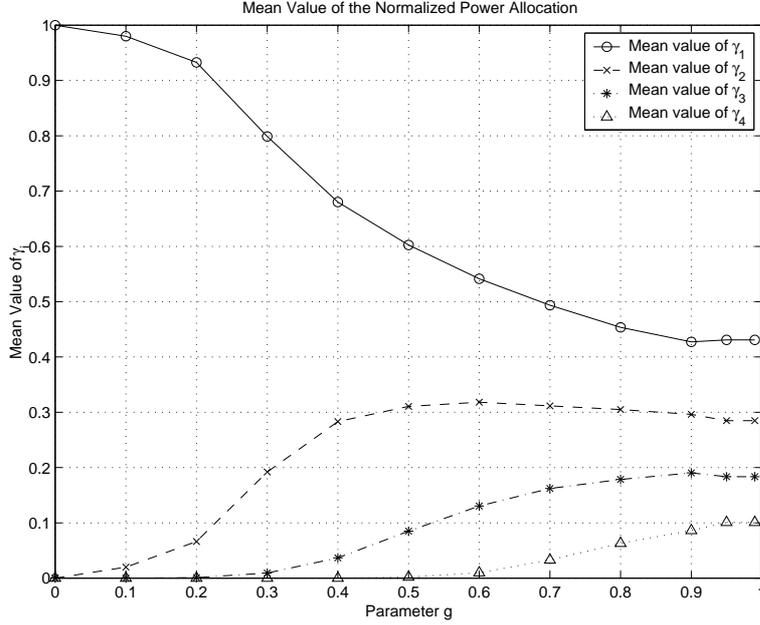
Figure 2.11: Mean value of the normalized robust power distribution for different sizes of the uncertainty regions.

region. These results are given as a function of the estimation and the quantization SNR.

In Figure 2.12, a TDD system has been considered with Gaussian estimation noise and assuming spherical uncertainty regions. The number of antennas is $n_T = 2$, $n_R = 2$ and, therefore, the transmission rate is 1. Two different probabilities of providing the required QoS have been considered: $P_{in} = 0.6$ and $P_{in} = 0.85$. As P_{in} increases, more power is necessary since the uncertainty region grows in order to guarantee the required performance with a higher probability. Besides, as the estimation SNR increases, less power is necessary, since the quality of the CSI improves. An interesting conclusion obtained from the figure is that, if the estimation SNR is high enough, the OSTBC technique needs more power than the non-robust and the robust solutions, since it does not exploit the channel knowledge. On the contrary, if the estimation SNR is low enough, the non-robust solution may need more power than OSTBC, concluding that in case that the CSI has a very low quality, it is not convenient to exploit that knowledge, unless using the robust solution. Note that, for all the estimation SNR range, the robust technique is the one requiring less power.

Very similar conclusions can be obtained from Figure 2.13 in terms of the quantization SNR. Once again, the robust technique needs less transmitted power than
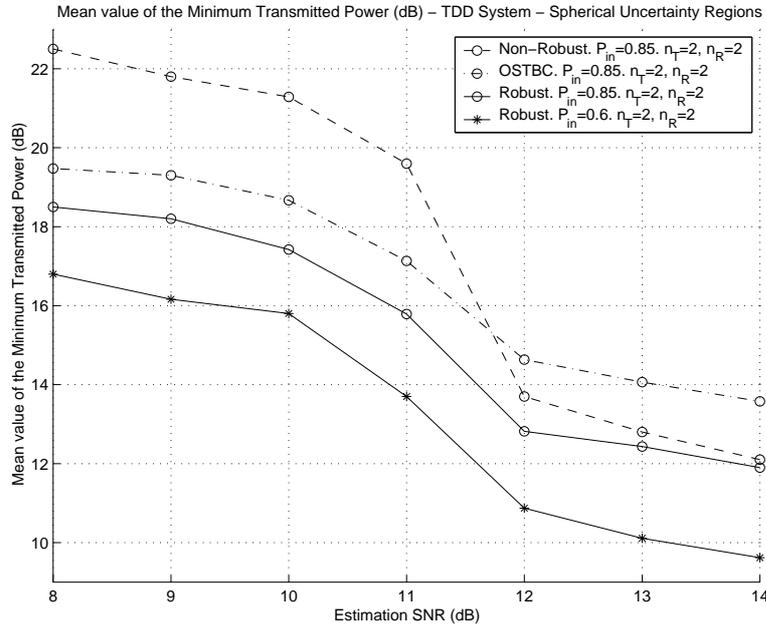
Figure 2.12: Minimum transmitted power in a TDD system assuming spherical uncertainty regions.

the other ones to fulfill the performance requirements represented by $\text{SNR}_0 = 10$ dB. In the same figure, the performance is also compared for two different antennas configurations: $n_T = 4$, $n_R = 4$ (rate 3/4) and $n_T = 6$, $n_R = 6$ (rate 1/2). From the simulations it is concluded that, as expected, increasing the number of antennas implies a reduction in the required transmitted power.

2.6   Summary

This chapter has given an overview of convex optimization theory with emphasis on the art of unveiling the hidden convexity of engineering problems and then has considered the design of linear MIMO transceivers or beamforming under the powerful framework of convex optimization theory.

The design of linear MIMO transceivers is a complicate nonconvex problem with matrix-valued variables. After several manipulations, the problem has been reformulated as a simple convex problem with scalar-valued variables. Then, the theory of convex optimization has been used to derive simple and efficient algorithms to compute the achievable region and, in particular, closed-form solutions have been obtained for the family of Schur-concave/convex functions such as the minimization
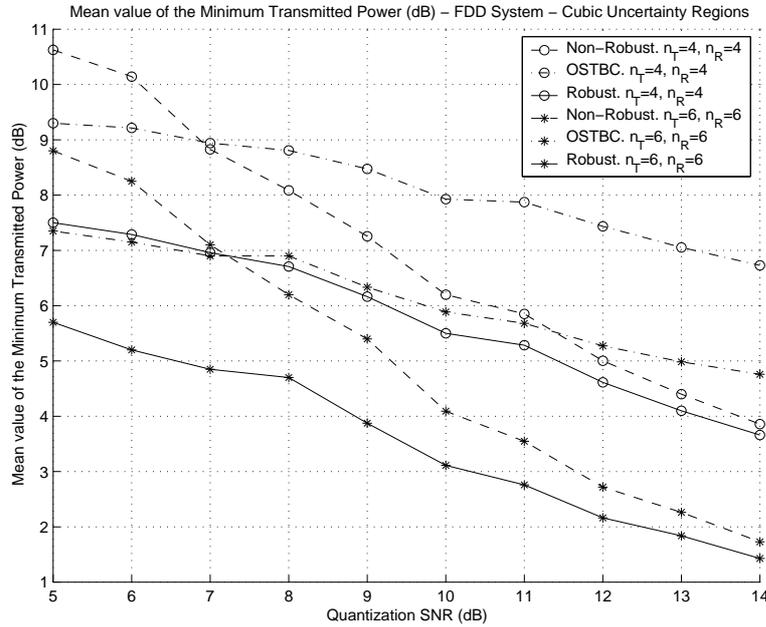
Figure 2.13: Minimum transmitted power in a FDD system assuming cubic uncertainty regions.

of the average BER of the system.

Finally, a robust transmission scheme for MIMO channels has been proposed based on the combination of an OSTBC, a power allocation, and a beamforming stage. The design of the power allocation is performed according to an imperfect channel estimate under the maximin philosophy, by which the worst performance is optimized for any error in the channel estimate described by an uncertainty region. The original maximin optimization problem corresponding to the design of the robust power allocation, which consists of two stages (inner and outer optimizations), is first formulated and then is transformed into a simple single-stage convex optimization problem, whose optimal Lagrange multipliers provide the optimum power allocation. For several examples of uncertainty regions, this optimization problem is shown to reduce to a convex quadratic problem.

References

[1] 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, Autumn 1996.

[2] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–366, March 1998.

[3] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov.-Dec. 1999.

[4] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.

[5] M. L. Honig, K. Steiglitz, and B. Gopinath, "Multichannel signal processing for data communications in the presence of crosstalk," *IEEE Trans. Commun.*, vol. 38, no. 4, pp. 551–558, April 1990.

[6] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers Part I: Unification and optimal designs," *IEEE Trans. Signal Processing*, vol. 47, no. 7, pp. 1988–2006, July 1999.

[7] H. Bölcskei and A. J. Paulraj, "Multiple-input multiple-output (MIMO) wireless systems," in *The Communications Handbook*, 2nd ed., pp. 90.1–90.14, J. Gibson, Ed. CRC Press, 2002.

[8] K. Miettinen, *Multi-Objective Optimization.* Kluwer Academic Publishers, 1999.

[9] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July-Oct. 1948.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[11] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[12] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 744–765, March 1998.

[13] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, Massachusetts: Athena Scientific, 1999.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004.

[15] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Processing*, vol. 51, no. 9, pp. 2381–2401, Sept. 2003.

[16] D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints," *IEEE Trans. Signal Processing*, vol. 52, no. 5, pp. 1179–1197, May 2004.

[17] D. P. Palomar, "A unified framework for communications through MIMO channels," Ph.D. dissertation, Technical University of Catalonia (UPC), Barcelona, Spain, May 2003.

[18] R. T. Rockafellar, *Convex Analysis*, 2nd ed. Princeton, NJ: Princeton Univ. Press, 1970.

[19] D. G. Luenberger, *Optimization by Vector Space Methods.* New York: Wiley, 1969.

[20] R. T. Rockafellar, "Lagrange multipliers and optimality," *SIAM Review*, vol. 35, no. 2, pp. 183–238, 1993.

[21] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra and Applications*, vol. 284, no. 1-3, pp. 193–228, July 1998.

[22] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, vol. 38, no. 1, pp. 49–95, March 1996.

[23] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*.   Society for Industrial and Applied Mathematics, 2001.

[24] Y. Nesterov and A. Nemirovski, "Interior-point polynomial methods in convex programming," *SIAM Studies in Applied Mathematics*, vol. 13, 1994.

[25] J. F. Sturm, "Using sedumi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11-12, pp. 625–653, 1999.

[26] J.-L. Goffin and J.-P. Vial, "Convex nondifferentiable optimization: A survey focussed on the analytic center cutting plane method," Department of Management Studies, University of Geneva, Switzerland, Tech. Rep., Feb. 1999.

[27] Z.-Q. Luo, "Applications of convex optimization in signal processing and digital communication," *Mathematical Programming, Series B*, vol. 97, no. 1-2, pp. 177–207, July 2003.

[28] H. Lebret and S. Boyd, "Antenna array pattern synthesis via convex optimization," *IEEE Trans. Signal Processing*, vol. 45, no. 3, pp. 526–532, March 1997.

[29] S.-P. Wu, S. Boyd, and L. Vandenberghe, "FIR filter design via spectral factorization and convex optimization," in *Applied and Computational Control, Signals and Circuits*, vol. 1, ch. 5, pp. 215–245, B. Datta, Ed.   Boston, MA: Birkhauser, 1999.

[30] T. N. Davidson, Z.-Q. Luo, and J. F. Sturm, "Linear matrix inequality formulation of spectral mask constraints with applications to FIR filter design," *IEEE Trans. Signal Processing*, vol. 50, no. 11, pp. 2702–2715, Nov. 2002.

[31] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*.   New York: Wiley, 1980.

[32] R. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Processing*, to appear 2004.

[33] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Processing*, vol. 51, no. 2, pp. 313–324, Feb. 2003.

[34] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Trans. Signal Processing*, vol. 51, no. 7, pp. 1702–1715, July 2003.

[35] S. Shahbazpanahi, A. B. Gershman, Z.-Q. Luo, and K. M. Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Trans. Signal Processing*, vol. 51, no. 9, pp. 2257–2269, Sept. 2003.

[36] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," in *Handbook of Antennas in Wireless Communications*, L. C. Godara, Ed.   Boca Raton, FL: CRC Press, 2001.

[37] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, pp. 142–163, March 1959.

[38] M. Chiang and S. Boyd, "Geometric programming duals of channel capacity and rate distortion," *IEEE Trans. Inform. Theory*, vol. 50, no. 2, pp. 245–258, Feb. 2004.

[39] L. Xiao, M. Johansson, and S. Boyd, "Simultaneous routing and resource allocation via dual decomposition," *IEEE Trans. Commun.*, vol. 52, no. 7, pp. 1136–1144, July 2004.

[40] S. H. Low, L. L. Perterson, and L. Wang, "Understanding vegas: A duality model," *Journal of the ACM*, vol. 49, no. 2, pp. 207–235, March 2002.

[41] M. Chiang, "To layer or not to layer: balancing transport and physical layers in wireless multihop networks," in *Proc. IEEE INFOCOM*, Hong Kong, China, March 2004.

[42] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.

[43] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods.* Prentice Hall, 1989.

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

[45] Z.-Q. Luo, T. N. Davidson, G. B. Giannakis, and K. M. Wong, "Tranceiver optimization for block-based multiple access through ISI channels," *IEEE Trans. Signal Processing*, vol. 52, no. 4, pp. 1037–1052, April 2004.

[46] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inform. Theory*, vol. 39, no. 4, pp. 1293–1302, July 1993.

[47] B. Picinbono, "On circularity," *IEEE Trans. Signal Processing*, vol. 42, no. 12, pp. 3473–3482, Dec. 1994.

[48] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Processing*, vol. 43, no. 8, pp. 2030–2033, Aug. 1995.

[49] P. J. Schreier and L. L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Trans. Signal Processing*, vol. 51, no. 3, pp. 714–725, March 2003.

[50] S. Benedetto and E. Biglieri, *Principles of Digital Transmission: With Wireless Applications.* New York: Kluwer Academic, 1999.

[51] S. Verdú, *Multiuser Detection.* New York: Cambridge University Press, 1998.

[52] K. Cho and D. Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1074–1080, July 2002.

[53] H. V. Poor and S. Verdú, "Probability of error in MMSE multiuser detection," *IEEE Trans. Inform. Theory*, vol. 43, no. 3, pp. 858–871, May 1997.

[54] D. P. Palomar, M. Bengtsson, and B. Ottersten, "Minimum BER linear transceivers for MIMO channels via primal decomposition," *accepted in IEEE Trans. Signal Processing*, to appear 2005.

[55] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Englewood Cliffs, NJ: Prentice Hall, 1993.

[56] K. H. Lee and D. P. Petersen, "Optimal linear coding for vector channels," *IEEE Trans. Commun.*, vol. COM-24, no. 12, pp. 1283–1290, Dec. 1976.

[57] J. Salz, "Digital transmission over cross-coupled linear channels," *At&T Technical Journal*, vol. 64, no. 6, pp. 1147–1159, July-Aug. 1985.

[58] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3221–3231, Dec. 1994.

[59] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.

[60] J. Yang and S. Roy, "Joint transmitter-receiver optimization for multi-input multi-output systems with decision feedback," *IEEE Trans. Inform. Theory*, vol. 40, no. 5, pp. 1334–1347, Sept. 1994.

[61] E. N. Onggosanusi, A. M. Sayeed, and B. D. V. Veen, "Efficient signaling schemes for wideban space-time wireless channels using channel state information," *IEEE Trans. Veh. Technol.*, vol. 52, no. 1, pp. 1–13, Jan. 2003.

[62] Y. Ding, T. N. Davidson, Z.-Q. Luo, and K. M. Wong, "Minimum BER block precoders for zero-forcing equalization," *IEEE Trans. Signal Processing*, vol. 51, no. 9, pp. 2410–2423, Sept. 2003.

[63] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, "MMSE decision-feedback equalizers and coding - Parts I-II: Equalization results and coding results," *IEEE Trans. Commun.*, vol. 43, no. 10, pp. 2582–2604, Oct. 1995.

[64] P. Viswanath and V. Anantharam, "Optimal sequences and sum capacity of synchronous CDMA systems," *IEEE Trans. Inform. Theory*, vol. 45, no. 6, pp. 1984–1991, Sept. 1999.

[65] A. Pascual-Iserte, A. I. Pérez-Neira, and M. A. Lagunas, "On power allocation strategies for maximum signal to noise and interference ratio in an OFDM-MIMO system," *IEEE Trans. Wireless Commun.*, vol. 3, no. 3, pp. 808–820, May 2004.

[66] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control.* Upper Saddle River, NJ: Prentice-Hall, 1996.

[67] B. Hassibi, A. H. Sayed, and T. Kailath, *Indefinite Quadratic Estimation and Control: A Unified Approach to $\mathcal{H}_2$ and $\mathcal{H}_\infty$ Theories.* Philadelphia, PA: SIAM, 1999.

[68] S. Zhou and G. B. Giannakis, "How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over Rayleigh MIMO channels?" *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1285–1294, July 2004.

[69] J. Choi, "Performance analysis for transmit antenna diversity with/without channel information," *IEEE Trans. Veh. Technol.*, vol. 51, no. 1, pp. 101–113, January 2002.

[70] ——, "Performance limitation of closed-loop transmit antenna diversity over fast rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 51, no. 4, pp. 771–775, July 2002.

[71] A. Narula, M. J. Lopez, M. D. Trot, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1423–1436, October 1998.

[72] A. Wittneben, "Optimal predictive TX combining diversity in correlated fading for microcellular mobile radio applications," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM 1995)*, pp. 13–17, Nov. 1995.

[73] A. Pascual-Iserte, A. I. Pérez-Neira, and M. A. Lagunas, "Exploiting transmission spatial diversity in frequency selective systems with feedback channel," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003)*, vol. 4, pp. 85–88, April 2003.

[74] G. Jöngren, M. Skoglund, and B. Ottersen, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 611–627, March 2002.

[75] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel mean feedback," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2599–2613, October 2002.

[76] F. Rey, M. Lamarca, and G. Vázquez, "Transmit filter optimization based on partial CSI knowledge for wireless applications," in *Proc. IEEE 2003 International Conference on Communications (ICC 2003)*, vol. 4, pp. 2567 –2571, Anchorage, AK, May 11-15, 2003.

[77] P. Xia, S. Zhou, and G. B. Giannakis, "Adaptive MIMO-OFDM based on partial channel state information," *IEEE Trans. Signal Processing*, vol. 52, no. 1, pp. 202–213, January 2004.

[78] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proceedings of the IEEE*, vol. 73, no. 3, pp. 433–481, March 1985.

[79] S. Verdú and V. Poor, "On minimax robustness: A general approach and applications," *IEEE Trans. Inform. Theory*, vol. 30, no. 2, pp. 328–340, March 1984.

[80] P. Stoica, Z. Wang, and J. Li, "Robust Capon beamforming," *IEEE Signal Processing Lett.*, vol. 10, no. 6, pp. 172–175, June 2003.

[81] M. Bengtsson and B. Ottersten, "Optimal downlink beamforming using semidefinite optimization," in *Proc. 37th Annual Allerton Conference on Communications, Control, and Computing*, pp. 987–996, Sept. 1999.

[82] M. Biguesh, S. Shahbazpanahi, and A. B. Gershman, "Robust power adjustment for transmit beamforming in cellular communications systems," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003)*, vol. 5, pp. 105–108, April 2003.

[83] A. Pascual-Iserte, A. I. Pérez-Neira, and M. A. Lagunas, "A maximin approach for robust MIMO design: Combining OSTBC and beamforming with minimum transmission power requirements," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004)*, vol. 2, pp. 1–4, May 2004.

[84] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1804–1824, July 2002.

[85] G. Ganesan and P. Stoica, "Space-time block codes: A maximum SNR approach," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1650–1656, May 2001.

[86] V. Tarokh, H. Jafharkani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[87] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge University Press, 2003.

[88] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel correlations," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1673–1690, July 2003.

[89] R. T. Rockafellar, "Saddle-points and convex analysis," in *Differential Games and Related Topics*, pp. 109–127, H. W. Kuhn and G. P. Szego, Eds. North-Holland Publ. Co., 1971.

[90] S. Žaković and C. Pantelides, "An interior point method algorithm for computing saddle points of constrained continuous minimax," *Annals of Operations Research*, vol. 99, pp. 59–77, December 2000.