Part 1 Communication fundamentals

1

Introduction

1.1 Introduction

Digital communication systems have been studied for many decades, and they have become an integral part of the technological world we live in. Many excellent books in recent years have told the story of this communication revolution, and have explained in considerable depth the theory and applications. Since the late 1990s particularly, there have been a number of significant contributions to digital communications from the signal processing community. This book presents a number of these recent developments, with emphasis on the use of filter bank precoders and equalizers. Optimization of these systems will be one of the main themes in this book. Both multiple-input multiple-output (MIMO) systems and single-input single-output (SISO) systems will be considered. It is assumed that the reader has had some exposure to digital communications and signal processing at the introductory level. Many text books cover this prerequisite, and some are mentioned at the beginning of Sec. 1.5.

Before we describe the contents of the book we first give an introductory description of analog and digital communication systems in the next few sections. The scope and outline of the book will be described in Sec. 1.5.

1.2 Communications systems

Figure 1.1(a) shows the schematic of a simple analog communication system. Here we have a message signal $s(t)$ which is transmitted over a channel to produce the signal $y(t)$ at the receiver end. In many practical systems the channel can be modeled as a linear time invariant, or **LTI**, system followed by an additive noise source $q(t)$. This is shown in Fig. 1.1(b), where the channel impulse response is indicated as $h(t)$.

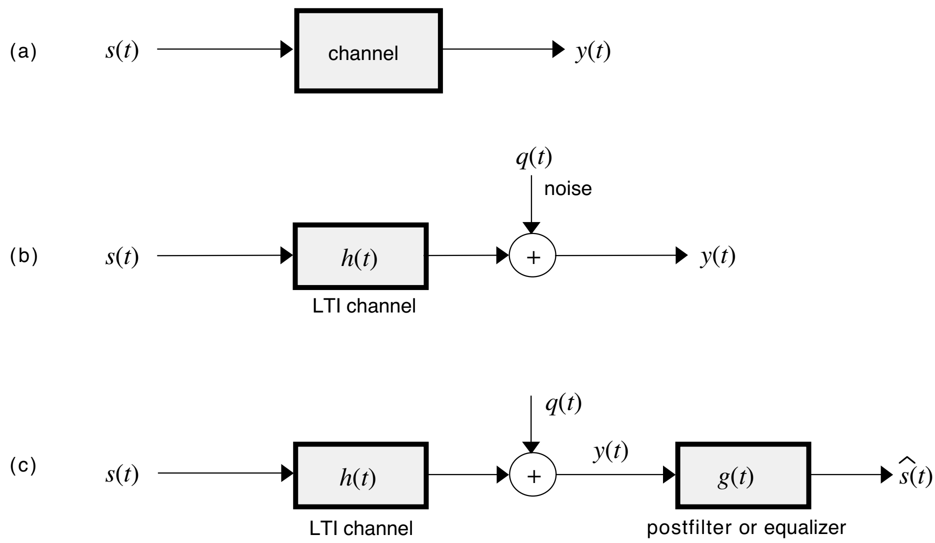

Figure 1.1. An analog communication system. (a) Channel with input $s(t)$ and output $y(t)$. (b) The channel modeled as a linear time invariant system followed by an additive noise source. (c) The channel followed by a postfilter or equalizer at the receiver.

The received signal $y(t)$ in Fig. 1.1(b) can be expressed in the form

$$y(t) = \int_{-\infty}^{\infty} h(\tau)s(t-\tau)d\tau \quad + q(t). \tag{1.1}$$

The first term above represents a convolution integral. In practice the channel is *causal* so that $h(t)$ is zero for $t < 0$. In this case the lower limit of the integral can be taken as 0 rather than $-\infty$.

Next consider Fig. 1.1(c), where the received signal $y(t)$ is processed using an LTI system called the **equalizer** or the **postfilter.** The purpose of an equalizer is to compensate for the distortion caused by the convolution with the channel $h(t)$, and to reduce the effect of the channel noise. The equalizer should be designed by taking into account the knowledge of $h(t)$ and whatever knowledge might be available about the statistics of the noise $q(t)$. The reconstructed signal $\widehat{s}(t)$ then serves as an approximation of $s(t)$. The reconstruction error is given by

$$e(t) = \widehat{s}(t) - s(t). \tag{1.2}$$

Figure 1.2 shows a further enhancement at the transmitter end. The message signal $s(t)$ is first passed through an LTI system called the **prefilter** or **precoder**. This system has impulse response $f(t)$. The prefilter "shapes" the message $s(t)$ appropriately.

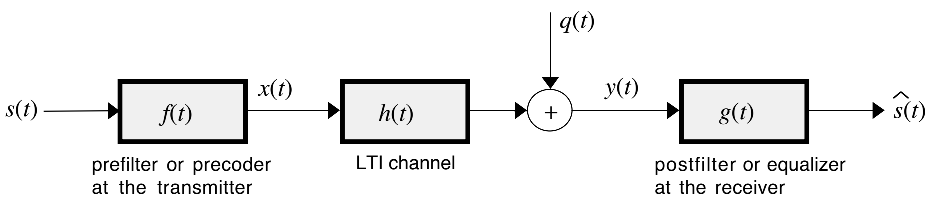

Figure 1.2. An analog communication system with a linear precoder at the transmitter and a linear equalizer at the receiver.

Given our knowledge of the channel and the noise statistics, it is possible to choose the prefilter (jointly with the equalizer) such that $\widehat{s}(t)$ approximates $s(t)$ as best as possible. For this we have to specify some error criterion such as, for example, the mean square error. Also, appropriate constraints on the transmitted power have to be specified. The precoder at the transmitter and the equalizer at the receiver are together referred to as a **transceiver**. Thus, we often talk of optimal design of a transceiver $\{f(t), g(t)\}$ for a given channel.[1] In later chapters we will consider many recent variations of this classical problem, especially in the context of *digital* communications. The multi-input multi-output (MIMO) version of this problem is especially important as we shall see. It is often convenient to use a frequency domain representation for the communication system. Thus, let $H(j\omega)$ denote the Fourier transform of $h(t)$, that is,

$$H(j\omega) = \int_{-\infty}^{\infty} h(t)e^{-j\omega t}dt. \tag{1.3}$$

This is called the **frequency response** of the channel. Similarly, let $F(j\omega)$ and $G(j\omega)$ represent the frequency responses of the precoder $f(t)$ and equalizer $g(t)$.

Figure 1.3 shows a redrawing of Fig. 1.2 in terms of these Fourier transform quantities. In terms of this notation the effective channel from $s(t)$ to $\widehat{s}(t)$ is

$$H_{eff}(j\omega) = G(j\omega)H(j\omega)F(j\omega). \tag{1.4}$$

With $(s * f)(t)$ denoting the **convolution** of $s(t)$ with $f(t)$, the effective channel impulse response can be written in the form

$$h_{eff}(t) = (g * h * f)(t). \tag{1.5}$$

There are many variations of the above channel model. In some situations, like mobile communications, the channel is modeled as a slowly time varying system rather than an LTI system. In some scenarios the channel impulse response $h(t)$ is regarded as a random variable drawn from a known distribution. The noise source $q(t)$ is often modeled as a Gaussian random process with known power spectrum. We shall come to the details later.

[1]Sometimes the entire system in the figure is loosely referred to as the transceiver.

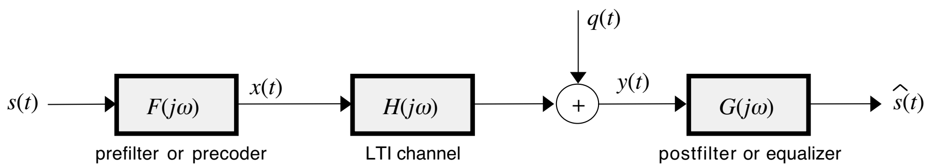

Figure 1.3. The analog communication system represented in terms of frequency responses.

1.3 Digitial communication systems

In the preceding section, the signal $s(t)$ was regarded as a continuous time signal with continuous (unquantized) amplitude. In a digital communication system, the messages are *quantized* amplitudes, transmitted in *discrete* time. Figure 1.4 shows the schematic of a digital communication system. Here we have a *discrete-time* message or signal $s(n)$ which we wish to transmit over a *continuous-time* channel. The amplitudes of $s(n)$ are "digitized," that is they come from a finite set of **symbols**. This collection of symbols is called a **constellation**. We shall come to details of digitization later.[2] Since $s(n)$ is a discrete-time signal and the channel is continuous-time, the signal is first converted into a continuous-time signal $x(t)$ as indicated in the figure.

The conversion from $s(n)$ to $x(t)$ can be described schematically in two steps. The building block indicated as D/C is a *discrete-to-continuous-time converter*, and it converts $s(n)$ to a signal $s_c(t)$ given by

$$s_c(t) = \sum_{n=-\infty}^{\infty} s(n)\delta_c(t-nT). \tag{1.6}$$

Here $\delta_c(t)$ is the *impulse* or *Dirac delta* function [Oppenheim and Willsky, 1997]. Thus, the sample $s(n)$ is converted into an impulse positioned at time nT. The sample spacing T determines the speed with which the message samples are conveyed. Since we have $1/T$ symbols per second, the **symbol rate** is given by

$$f_s = \frac{1}{T} \text{ Hz}. \tag{1.7}$$

The prefilter $F(j\omega)$ at the transmitter performs a convolution to produce the output

$$x(t) = \sum_{n=-\infty}^{\infty} s(n)f(t-nT), \tag{1.8}$$

[2]Examples of constellations include PAM and QAM systems to be described in Sec. 2.2.

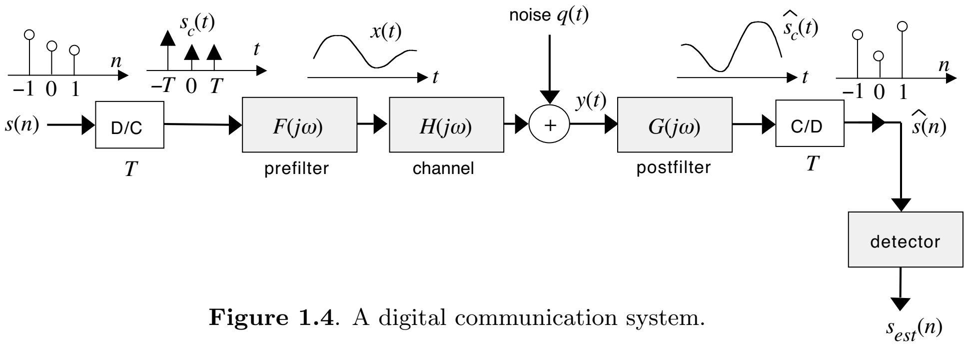

Figure 1.4. A digital communication system.

where $f(t)$ is the impulse response of $F(j\omega)$. Typically $f(t)$ is a smooth, finite duration function, as demonstrated in Fig. 1.5(a). In practice, $f(t)$ is **causal**, that is, it is zero for negative time. In the figure it is shown to be noncausal for generality. Note that $x(t)$ is a weighted sum of uniformly shifted versions of the impulse response $f(t)$. The weight on the nth shifted version $f(t-nT)$ is the nth message sample $s(n)$. This construction of $x(t)$ from $s(n)$ is demonstrated in Fig. 1.5(b).

The signal $x(t)$ is then transmitted over the continuous-time channel $H(j\omega)$, which also adds noise $q(t)$:

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau \quad + q(t). \tag{1.9}$$

This is the signal that will be observed at the receiver. The goal at the receiver is to reconstruct the original discrete signal $s(n)$ from this noisy and distorted continuous-time signal $y(t)$. First, the postfilter $G(j\omega)$ at the receiver processes $y(t)$ to produce $\widehat{s}_c(t)$, which is then **sampled at the rate** $f_s = 1/T$ to obtain a reconstructed version of $s(n)$:

$$\widehat{s}(n) = \widehat{s}_c(nT). \tag{1.10}$$

The box labeled C/D is the *continuous-to-discrete-time converter*, and performs the sampling operation (1.10). The reconstruction error is given by

$$e(n) = \widehat{s}(n) - s(n). \tag{1.11}$$

Given the knowledge of the channel $H(j\omega)$ and the noise statistics, it is possible to design the filters $F(j\omega)$ and $G(j\omega)$ to minimize an appropriate measure of reconstruction error. A simple example of an "appropriate measure" is the mean square error (i.e., the average value of $|e(n)|^2$).

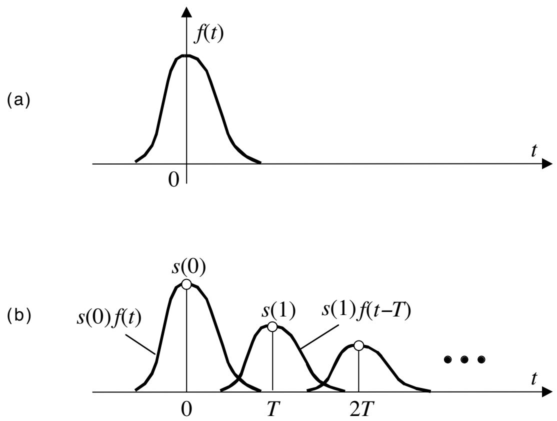

Figure 1.5. (a) Impulse response $f(t)$ of the prefilter $F(j\omega)$, and (b) the signal $x(t)$ generated from the samples $s(n)$ by interpolation with the function $f(t)$.

Note that, while $s(n)$ belongs to the signal constellation, the quantity $\widehat{s}(n)$ does not. In practice there is a device called the **detector** at the receiver (Fig. 1.4), which obtains an *estimated constellation symbol* $s_{est}(n)$ from the quantity $\widehat{s}(n)$. The *probability of symbol error* is defined to be the probability that $s_{est}(n)$ differs from $s(n)$. The minimization of this probability is another important optimization problem. Such optimizations and several generalizations will be discussed in appropriate sections of the book.

Bandlimiting. In practice the bandwidth allowed for a transceiver is limited. This bandlimiting is enforced by using lowpass filters at the transmitter and receiver. These filters can be incorporated as parts of $F(j\omega)$ and $G(j\omega)$. The transmitted signal $x(t)$ therefore occupies a fairly narrow bandwidth of the form

$$-\sigma < \omega < \sigma, \tag{1.12}$$

and is called the **baseband** signal. The bandwidth can be a few kHz to MHz, depending on application. The signal $x(t)$ is actually used to modulate a high-frequency carrier, and the modulated signal is transmitted either wirelessly using antennas or on wirelines. A discussion of carrier modulation is included in Sec. 2.4. The channel model discussed above is called the *baseband model*, as it does not show the carrier explicitly. Similarly the continuous-time system described in Sec. 1.2 also represents a baseband model. $\triangledown \triangledown \triangledown$

1.3.1 Discrete-time equivalent

We will see in later sections that the problem of designing the digital communication system of Fig. 1.4 can be reformulated entirely in terms of discrete-time transfer functions as in Fig. 1.6.

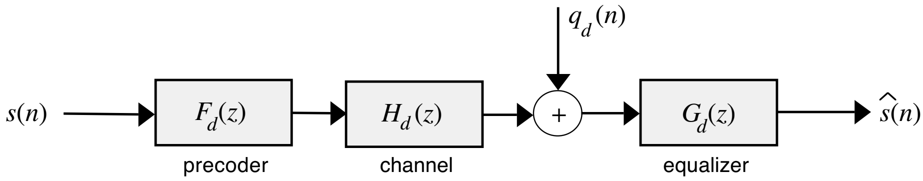

Figure 1.6. An all-discrete equivalent of the digital communication system.

Here $H_d(z)$ is the transfer function of an equivalent *discrete-time* channel. It is the z-transform of an equivalent digital channel impulse response $h_d(n)$, that is

$$H_d(z) = \sum_{n=-\infty}^{\infty} h_d(n)z^{-n}. \tag{1.13}$$

Similarly, $F_d(z)$ and $G_d(z)$ are the transfer functions of the discrete-time precoder and equalizer. The subscript d (for "discrete"), which is just for clarity, is usually dropped. In practice $H_d(z)$ is causal and can be approximated by a finite impulse response, or **FIR**, system so that

$$H_d(z) = \sum_{n=0}^{L} h_d(n)z^{-n}. \tag{1.14}$$

The problem of optimizing the precoder $F_d(z)$ and equalizer $G_d(z)$ for fixed channel $H_d(z)$ and fixed noise statistics will be addressed in later chapters.

1.4 MIMO channels

The transceivers described so far have one input signal $s(n)$ and a corresponding output $\widehat{s}(n)$. These are called *single-input single-output*, or **SISO**, transceivers. An important communication system that comes up frequently in this book is the *multi-input multi-output*, or **MIMO**, channel. Figure 1.7 shows a MIMO channel assumed to be linear and time-invariant with a transfer function matrix $\mathbf{H}(z)$, usually an FIR system:

$$\mathbf{H}(z) = \sum_{n=0}^{L} \mathbf{h}(n)z^{-n}. \tag{1.15}$$

The sequence $\mathbf{h}(n)$, called the MIMO impulse response, is a sequence of matrices. If the channel has P inputs and J outputs then $\mathbf{H}(z)$ has size $J \times P$, and so does each of the matrices $\mathbf{h}(n)$. The MIMO communication channel is used to transmit a vector signal $\mathbf{s}(n)$ with M components:

$$\mathbf{s}(n) = [\, s_0(n) \quad s_1(n) \quad \ldots \quad s_{M-1}(n)\,]^T. \tag{1.16}$$

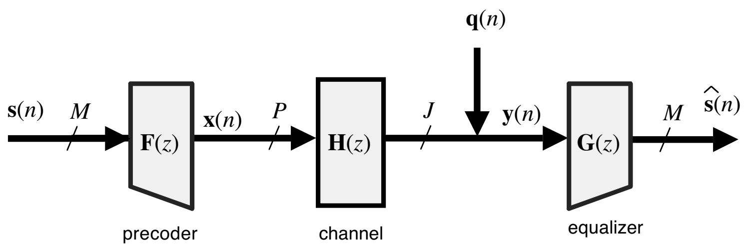

Figure 1.7. A digital communication system.

The precoder $\mathbf{F}(z)$ transforms this sequence $\mathbf{s}(n)$ into another sequence $\mathbf{x}(n)$. We will see that the choice of $\mathbf{F}(z)$ plays an important role in the performance of the communication system. The channel produces the inevitable distortion represented by the transfer function $\mathbf{H}(z)$ and the noise vector $\mathbf{q}(n)$. Thus the signal obtained at the receiver is

$$\mathbf{y}(n) = \sum_{k=0}^{L} \mathbf{h}(k)\mathbf{x}(n-k) \; + \; \mathbf{q}(n). \tag{1.17}$$

The equalizer $\mathbf{G}(z)$ seeks to reconstruct $\mathbf{s}(n)$ from this distorted version:

$$\widehat{\mathbf{s}}(n) = \sum_{k} \mathbf{g}(k)\mathbf{y}(n-k). \tag{1.18}$$

The joint design of the transceiver $\{\mathbf{F}(z), \mathbf{G}(z)\}$ is an important problem in modern digital communications. The MIMO transceiver shown in the figure can be used to transmit messages

$$s_k(n), \; 0 \leq k \leq M-1, \tag{1.19}$$

from M separate users. It can also be used to transmit information from one user by representing the message $s(n)$ from the user in the form of a vector $\mathbf{s}(n)$; such systems are called block-based transceivers for SISO channels. They have many advantages as we shall see. MIMO channels also arise from the use of multiple antennas for single users; a detailed discussion of how MIMO channels arise will be given in Sec. 4.5.

A special case of the MIMO system arises when the channel is **memoryless**, that is, the transfer function $\mathbf{H}(z)$ is just a constant $\mathbf{H}$. This corresponds to the situation where $L = 0$ in Eq. (1.15). Optimization of transceivers for memoryless MIMO channels will be the focus of some of the chapters in this book.

1.5 Scope and outline

The reader is assumed to have some familiarity with introductory topics in communications and signal processing. References for such background material include Proakis [1995], Oppenheim and Willsky [1997], Oppenheim and Schafer [1999], Lathi [1998], Haykin [2001], Mitra [2001], and Antoniou [2006], among other excellent texts. Advanced material related to the topics in this book can also be found in Ding and Li [2001], and Giannakis, *et al.* [2001]. Books on the very important related areas of wireless and multiuser communications include Rappaport [1996], Verdu [1998], Goldsmith [2005], Haykin and Moher [2005], and Tse and Viswanath [2005].

The book is divided into four parts. Part 1 contains introductory material on digital communication systems and signal processing aspects. In part 2 we discuss the optimization of transceivers, with emphasis on MIMO channels. Part 3 provides mathematical background material for optimization of transceivers. This part can be used as a reference, and will be very useful for readers wishing to pursue more detailed literature on optimization. Part 4 contains eight appendices on commonly used material such as matrix theory, Wiener filtering, and so forth.

The history of digital communication theory is fascinating. It is a humbling experience to look back and reflect on the tremendous insights and accomplishments of the communications and signal processing society in the last six decades. Needless to say, much of the recent research is built upon six to seven decades of this solid foundation. A detour into history will be provided in Chap. 9, where we present a historical perspective of transceiver design, equalization, and optimization, all of which originated in the early 1960s, and have continued to this day to be research topics. All references to literature will be given in the specific chapters as appropriate. An extensive reference list is given at the end of the book. In what follows we briefly describe the four parts of the book.

Part 1: Communication fundamentals

Part 1 consists of Chapters 1 to 8. In Chap. 2 we review basic topics in digital communication systems, such as signal constellations, carrier modulation, and so forth. Formulas for probabilities of error in symbol detection are derived. Matched filtering, which is used in some receiver systems, is discussed in some detail. In Chap. 3 we describe digital communication systems using the language of multirate filter banks. Such a representation is very useful for transceivers with or without redundancy, and has many applications as we shall see throughout the book.

In Chap. 4 we describe digital communication systems using discrete-time language. This chapter also introduces symbol spaced equalizers (SSE) and fractionally spaced equalizers (FSE). The minimum mean square error (MMSE) equalizer is also introduced in this chapter. Chapter 5 discusses a number of fundamental techniques that are commonly used in digital communications. First a detailed discussion of the matched filter is provided. Then we discuss optimal sequence estimators, such as the maximum likelihood (ML) detector and the Viterbi alogrithm. Nonlinear methods, such as the decision feedback equalizer and nonlinear precoders, are introduced. Chapter 6 is a brief discussion of channel capacity with emphasis on MIMO channels.

Chapter 7 introduces *redundant precoders*, including *zero-padded* and *cyclic-*

prefixed precoders. The redundant precoder is an integral part of many of the transceiver designs today. For example, cyclic prefix systems are employed in *orthogonal frequency division multiplexing* (OFDM) systems and discrete multitone (DMT) systems, used in *digital subscriber loop* (DSL) technology. The introduction of redundancy allows us to compensate or equalize the effects of a linear channel very efficiently – for example, an FIR channel can be equalized without the use of IIR equalizers. In Chap. 8 we discuss zero-padded systems in greater detail and introduce zero-forcing FIR equalizers, which can perfectly equalize FIR channels by exploiting the redundancy in the transmitted symbol stream. A number of properties of such equalizers are studied.

Part 2: Transceiver optimization

Part 2 consists of Chapters 9 to 19. Chapter 9 gives a brief historical introduction to transceiver optimization, and provides a detailed outline for Chapters 10 to 19. Briefly, Chap. 10 discusses the optimization of transceivers for scalar channels, and Chap. 11 discusses the optimization of transceivers for MIMO diagonal channels.

Chapters 12 and 13 discuss the minimization of mean square error in transceivers (MMSE transceivers) for general (nondiagonal) channels with and without the so-called *zero-forcing* constraint. Chapter 14 discusses the minimization of transmitted power for fixed performance criteria (such as error probability). This chapter also shows how one can perform *bit allocation* among the symbol streams optimally.

Chapter 15 discusses transceiver optimization for the special case where the precoder at the transmitter is constrained to be orthogonal. In Chap. 16 we consider the minimization of symbol error rates or bit error rates (BER), which are more directly related to practical performance than mean square errors. There is a close connection between MMSE transceivers and minimum-BER transceivers as we shall see in that chapter. The results of transceiver optimization are applied in Chaps. 17 and 18 to the case of cyclic-prefix systems and zero-padded systems, respectively. These are single-input single-output (SISO) channels turned into multi-input multi-output (MIMO) channels by introducing redundancy as described in Chap. 7. Chapter 19 discusses the decision feedback equalizer for MIMO channels. The joint optimization of transceiver matrices with decision feedback is discussed in detail.

Part 3: Mathematical background

Part 3 consists of Chapters 20 to 22. Some of the mathematical background needed for the optimization chapters is given in these chapters. This includes matrix calculus, Schur convex functions, and nonlinear optimization tools. Matrix calculus is a less commonly reviewed topic, so Chap. 20 offers a detailed review. Schur convex functions have played a major role in transceiver optimization in recent years, and the review in Chap. 21 will be useful to readers wishing to pursue the literature in depth. Chapter 22 is a review of constrained optimization theory, which is useful in some of the chapters on transceiver opti-