TECHNISCHE UNIVERSITÄT MÜNCHEN
Fachgebiet Methoden der Signalverarbeitung

Communications in Multi-Antenna Interference Networks

David A. Schmidt

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. sc. techn. Gerhard Kramer
Prüfer der Dissertation:

1. Univ.-Prof. Dr.-Ing. Wolfgang Utschick
2. Prof. Michael L. Honig, Ph. D.
   Northwestern University, Evanston, USA

Die Dissertation wurde am 28. Juni 2011 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 6. Dezember 2011 angenommen.

To my sister Christine

Acknowledgments

It would not have been possible for me to complete this work without the guidance, support, and small and big contributions of everyone at TU München's Associate Institute for Signal Processing. I am deeply grateful to Professor Wolfgang Utschick for supervising my research, and for his patience and trust in me. Also, many of the results in this work originated from discussions with Professor Mike Honig, whom I would like to thank for the fruitful cooperation and for being a great host during my visit to Northwestern University.

Of all my colleagues at the Associate Institute I am especially grateful to Michael Joham, for his excellent advice in every situation, and to Raphael Hunger, for being a great office partner and friend for many years. I would also like to thank Peter Breun for his help with all things statistical, Max Riemensberger for sharing his expert knowledge of optimization, and Christoph Hellings and Andreas Gründinger, whose excellent work as students laid some of the foundations for this thesis.

It was a great pleasure to work with the whole group: Johannes Brehmer, Andreas Dotzler, Lennart Gerdes, Christian Guthy, and Alexander Krebs each have their own share in the success of my thesis, both by helping with specific mathematical and technical problems and by contributing to the friendly and creative working atmosphere at the Associate Institute. Finally, I would like to thank Hans Brunner for the good times administrating and Leo Baltar for bringing a little bit of Rio to Munich.

Munich, December 2011

Contents

Abbreviations

AWGN	additive white Gaussian noise
bpcu	bits per channel use
DoF	degrees of freedom
EVD	eigenvalue decomposition
FDMA	frequency-division multiple access
GDoF	generalized degrees of freedom
i. i. d.	independent and identically distributed
INR	interference-to-noise ratio
KKT	Karush-Kuhn-Tucker
MAC	multiple access channel
ME	maximum eigenvalue/eigenvector
MF	matched filter
MIMO	multiple-input multiple-output
MISO	multiple-input single-output
MMSE	minimum mean squared error
MSE	mean squared error
NBS	Nash bargaining solution
PDF	probability density function
SINR	signal-to-interference-plus-noise ratio
SISO	single-input single-output
SNR	signal-to-noise ratio
SOCP	second-order cone problem
SVD	singular value decomposition
TDMA	time-division multiple access
ZF	zero-forcing

Notation

Throughout this work, we use bold upper case letters to denote matrices and bold lower case letters to denote vectors. We furthermore use the following notational conventions:

- $\log(\cdot)$ is the natural logarithm,
- $\lfloor \cdot \rfloor$ is the floor operator,
- $\mathrm{E}[\cdot]$ is the expected value of a random variable,
- $|\cdot|$ is the absolute value of a complex scalar,
- $\mathrm{Re}(\cdot)$ is the real part and $(\cdot)^*$ is the complex conjugate of a complex scalar, vector, or matrix,
- $(\cdot)^{\mathrm{T}}$ is the transpose and $(\cdot)^{\mathrm{H}}$ is the conjugate transpose of a vector or matrix,
- $(\cdot)^+$ is the pseudo-inverse of a matrix,
- $\mathrm{tr}(\cdot)$ is the trace and $\det(\cdot)$ is the determinant of a square matrix,
- $\mathbf{I}$ is the identity matrix, where the dimensions follow from the context,
- $\boldsymbol{e}_k$ is the kth column of the identity matrix $\mathbf{I}$, where the dimensions follow from the context,
- $\mathbf{1}$ is the vector of ones, where the dimensions follow from the context,
- and $\mathbf{0}$ is the vector or matrix of zeros, where the dimensions follow from the context.

We use the notation $(\,\cdot\,,\,\cdot\,)$ for an open interval and $[\,\cdot\,,\,\cdot\,]$ for a closed interval. When the upper limit of a summation is clear from the context, we use abbreviations of the form

$$\sum_{k} \quad \text{for} \quad \sum_{k=1}^{K} \qquad \text{and} \qquad \sum_{j \neq k} \quad \text{for} \quad \sum_{\substack{j=1 \\ j \neq k}}^{K}.$$

Landau Notation

The Landau symbol $\mathrm{O}(f(x))$ denotes the class of functions $g(x)$ for which

$$\limsup_{x \to \infty} \left| \frac{g(x)}{f(x)} \right| < \infty$$

or, equivalently, the class of functions $g(x)$ for which there exist real positive numbers M and x_0, such that

$$|g(x)| \leq M \cdot |f(x)| \qquad \text{for all} \quad x \geq x_0.$$

Similarly, the symbol $\mathrm{o}(f(x))$ denotes the class of functions $g(x)$ for which

$$\lim_{x \to \infty} \left| \frac{g(x)}{f(x)} \right| = 0.$$

The Landau symbols can be used for the set of functions or to denote a member of the set of functions; if $g(x) = 2x^3 + 3x^2 + 1$, we can, e. g., write $g(x) \in \mathrm{O}(x^3)$ as $x \to \infty$ or $g(x) = 2x^3 + \mathrm{O}(x^2)$ as $x \to \infty$. For a more detailed discussion of the Landau notation, cf. [1].

1. Introduction

Due to the increasing demand for spectral efficiency in modern cellular communications systems, it is no longer sufficient to view the cells as isolated entities. Mobile terminals close to the cell edge may receive a considerable amount of interference from neighboring base stations transmitting in the same frequency band; similarly, base stations may receive interference from mobile terminals in neighboring cells. The interference in many cases is the limiting factor when striving for high data rates so that optimal service can only be provided to mobile terminals close to the cell center.

At the same time, the advent of base stations and mobile terminals with multiple antennas opens up new possibilities for canceling interference, both on the receiver side, where the signals received by the different antennas can be combined in a way that separates the desired signal from interference and noise, and on the transmitter side, where beamforming techniques can direct the transmitted signal towards the desired receiver and away from unintended receivers. It is not yet very well understood how these additional degrees of freedom are best put to use in a practical system, such that the overall performance is optimized at a reasonable complexity. It can be assumed, though, that it is generally beneficial for the base stations to communicate among each other using the network infrastructure in order to coordinate their interference avoidance efforts.

There are two fundamentally different approaches for base station cooperation over the network infrastructure: the premise of the first approach is that the base stations cooperatively precode and decode the data; effectively, several base stations are connected to form a single multi-antenna node with antenna groups in different physical locations. The second approach, on the other hand, excludes synchronization on a data symbol level and only allows the base stations to cooperate in designing their *strategies.* The term strategy in this context encompasses the beamforming coefficients, the coding schemes, the transmit power, etc.; the payload data to be transmitted to mobile terminals in other cells, however, is not known and the signals received by other base stations cannot be used to help decode the own received signal. Clearly, the former approach leads to far stricter requirements concerning the bandwidth of the infrastructure, processing power, latency, and synchronicity; at the same time, the potential performance gains are very high. In this work we investigate the latter approach: while the performance limits are significantly lower, implementation is, arguably, more straightforward and the requirements on the connecting links between the base stations are far less stringent.

In order to understand some of the fundamental effects and tradeoffs inherent in the problem of cooperative transmit strategy design, we investigate an idealized elementary subproblem: we assume that the wireless propagation channels are frequency flat and quasi-static, and that they can be estimated without error. Also, in our scenario only one mobile terminal is associated with each base station. These abstractions lead to the *K-user interference channel:* the system model consists of K transmitter-receiver pairs, where each receiver is only interested in the data from its associated transmitter, but receives a superposition of the signals from all K transmitters as well as addititional background noise.

Exploring the capacity limits of such interference channels has been a challenge in information theory for nearly half a century. Even for the simplest such system with two single-antenna users, not all aspects of the optimal transmit strategies have been fully understood. In recent years, sig-

nificant progress has been made, however, especially in the analysis of systems with more than two users and/or multiple-antenna nodes. We give an overview of the current state of the information theoretic study of interference channels in Chapter 2.

The best known transmit strategies—in an information theoretic sense—for interference channels with more than two users turn out to be rather complex and somewhat impractical. For the remainder of the work we instead focus on one specific class of strategies that is information theoretically suboptimal except for some special cases, but far more suitable for practical implementation: the premise of these *linear* strategies is that the users cooperate only in choosing beamformer coefficients and transmit power, but select the channel coding scheme separately, and that the receivers treat interference as noise and thus do not require knowledge of the interfering users' channel codes.

In Chapters 3, 4, and 5, we investigate the optimal linear strategies for increasingly complex scenarios: first, we assume a single antenna at each node, which allows us to gain familiarity with the relevant concepts in a comparatively simple setting; then, in Chapter 4, we allow multiple antennas on the transmitter side, thus generalizing the problem from power control to beamformer design; and finally, in Chapter 5, we examine the general case of multiple antennas on the transmitter and receiver side. For each of these three cases, we give an overview of the known analytical results, discuss different proposed algorithms for transmit strategy design, and numerically compare the performance of these algorithms for random channels. For the multi-antenna setting in Chapter 5, we restrict our attention to the special case of one data stream per user. The more general multi-stream case introduces many additional difficulties and is therefore outside of the scope of this work.

In addition to the systematic overview of the topic, our main contributions are as follows:

- We discuss in detail the update procedure of the multi-antenna interference pricing algorithm. The concept of a distributed sum utility optimization algorithm based on exchanging interference prices was developed for single-antenna interference networks in [2] and a simplified extension to multi-antenna systems, which does not support power control and is not guaranteed to converge, was proposed in [3]. The generalization of the pricing algorithm to multi-antenna systems presented here resolves both of these issues and is shown experimentally to be a very useful technique in a broad range of scenarios.
- We provide a thorough qualitative evaluation and experimental comparison of many different known algorithms for multi-antenna interference networks. While there has been extensive effort recently to propose methods for determining good linear transmit strategies, significantly fewer results are available in which the performance of these different techniques is compared and such comparisons are usually limited to selected scenarios. Furthermore, issues such as the necessary amount of information exchange or the number of iterations required for convergence are often neglected. The comparison in this work covers many different settings and shows that, depending on the scenario, some algorithms are significantly more useful than others.
- We present a probabilistic analysis of the asymptotic *rate offset* in the multi-antenna case. Due to the complexity of the problem, analysis in the multi-antenna case is largely restricted to high-SNR asymptotics; while recent research has focused mainly on the achievable number of interference-free data streams, our rate offset analysis shows that there are considerable performance differences between solutions that achieve the same number of interference-free data streams; our analysis allows us to approximate the high-SNR performance of choosing the best out of many different interference-free solutions and even permits an estimation of the globally optimal high-SNR performance in some smaller systems.

2. The Interference Channel in Information Theory

One of the main objectives of information theory is to characterize the data rates that can be transmitted reliably over different types of communication channels. For many basic set-ups, such as a single noisy communication link, or a receiver trying to decode messages received simultaneously from many different transmitters, the achievable rates are well understood. Scenarios with interfering pairs of transmitters and receivers, however, have been the subject of such information theoretic analysis for nearly half a century and the characterization of the achievable data rates, the *capacity region,* is in general still unknown.

In this chapter we present a brief survey of the information theoretic research on interference scenarios similar to those we examine in the subsequent chapters. Our goal is to give a motivation for the utility functions and the *linear* strategies employed later on, and at the same time place them into a broader context. By no means do we attempt to provide an exhaustive overview over the literature.

2.1 Capacity of the AWGN Channel

Consider the discrete time communications system characterized by the equation

$$y[t] = h \cdot x[t] + n[t] \tag{2.1}$$

where $t \in \mathbb{Z}$ is the time index. The complex channel gain $h \in \mathbb{C}$ is constant over time, the noise $n[t]$ is a complex-valued, stationary random sequence with a jointly Gaussian distribution of the real and imaginary parts and mean $\mathrm{E}\,[n[t]] = 0$ for all $t \in \mathbb{Z}$. We furthermore assume $n[t]$ to be a "proper" random sequence in the sense of [4], i. e., all "pseudo-covariances" of the form $\mathrm{E}\,[n[t]n[t+j]]$ with $j \in \mathbb{Z}$ are zero, a property often referred to as "circular symmetry" of the complex Gaussian distribution. The variance of $n[t]$ is $\mathrm{E}\,[|n[t]|^2] = \sigma^2$, the temporal correlation is $\mathrm{E}\,[n[t]n^*[t+j]] = 0$ when $j \neq 0$. Since the noise is uncorrelated in time, and thus white in frequency, we call this communications system the *additive white Gaussian noise* (AWGN) channel.

In order to transmit information, we apply an input signal $x[t]$ to the channel during the time indices $t = 1, \ldots, N$, which must obey the unit transmit power constraint

$$\frac{1}{N} \sum_{t=1}^{N} |x[t]|^2 \leq 1. \tag{2.2}$$

The input signal $(x[1], \ldots, x[N])$ is the *codeword* and is selected from a *codebook* that contains M different codewords, each fulfilling the power constraint (2.2). The choice of the codeword conveys $\log_2 M$ *bits* of information, or $\log M$ *nats.* In the following we will mostly use the unit nats in analytical expressions for ease of notation and the unit bits for numerical results; the conversion from nats to bits can be accomplished by a division by $\log 2$. The information conveyed by the codeword normalized with the length of the codeword $R = \frac{1}{N} \log M$ is called *code rate* and is measured in *nats per channel use,* or *bits per channel use* (bpcu) when $\log_2$ is used.

The output signal $y[t]$ of the channel is used by the receiver in the following way to *decode* the information: based on the whole received block $(y[1], \ldots, y[N])$ and taking into account the channel gain h and the distribution of $n[t]$, it decides which of the M codewords in the codebook is most likely to have been the transmitted signal $(x[1], \ldots, x[N])$, e. g., according to the maximum likelihood criterion. Note that this can be a very complex task, possibly requiring an exhaustive search over a very large number M of codewords. At this point the issue of computational complexity is not considered, however, as we are interested in theoretic throughput limits.

Depending on the codebook and the specific decoding function, which maps any received block back to one of the M codewords from the codebook, an average *probability of error* can be defined, where an error is the event that the estimate returned by the decoding function is not the same as the codeword that actually was transmitted. A code rate R is called *achievable* if, given an arbitrarily low $\varepsilon > 0$, there exists a codebook and a decoding function with rate R, for which the probability of error is no higher than ε. Note that the block length N is not fixed and is allowed to be as large as necessary when determining whether R is achievable. The *channel capacity* C is defined as the supremum of all achievable rates R. Consequently, all non-negative $R < C$ are achievable and all $R > C$ are not achievable.

The *channel coding theorem* from information theory (see [5] for a thorough discussion) provides us with the means for determining the channel capacity C without specifically optimizing the codebooks and decoding functions; instead, the system (2.1) is examined at one given time instant and the input variable x is considered to be a random variable with the probability density function (PDF) $p(x)$. The output variable (or received symbol)

$$y = h \cdot x + n \tag{2.3}$$

is therefore also a random variable, and we can describe our system by means of the joint PDF $p(x, y) = p(y|x) \cdot p(x)$.[1] Due to our assumptions concerning the properties of the noise, the conditional PDF of y given x, $p(y|x)$, is complex Gaussian with variance σ^2 and mean $h \cdot x$; the PDF of the channel input $p(x)$ depends on the transmit strategy. Next, we define the *mutual information* $I(x; y)$ between the two random variables x and y as

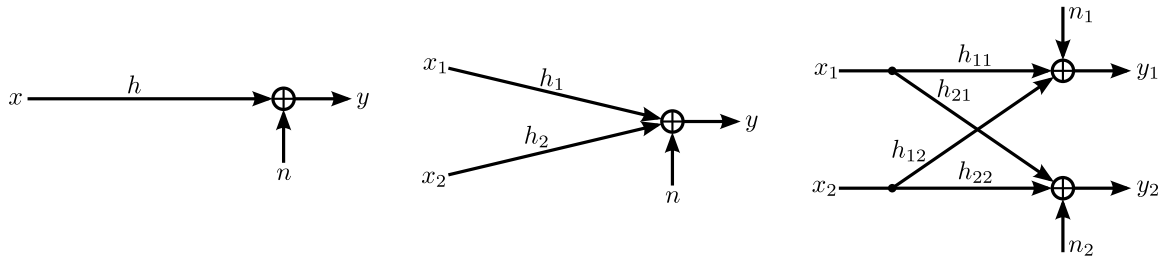
$$I(x; y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \, \mathrm{d}y \, \mathrm{d}x = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(y|x)p(x) \log \frac{p(y|x)}{p(y)} \, \mathrm{d}y \, \mathrm{d}x \tag{2.4}$$

where $p(y)$ is the marginal PDF

$$p(y) = \int_{-\infty}^{\infty} p(y|x)p(x) \, \mathrm{d}x. \tag{2.5}$$

The mutual information can be interpreted as a measure for how much information about the realization of x can be inferred from the observation of y, or vice versa. For the special case that y and x are statistically independent, the mutual information is zero. Again, for a detailed discussion of the concept of mutual information, the reader is referred to [5].

[1] Our notation of random variables is somewhat imprecise here, as we do not distinguish between the random variable and its realization. Also, we denote all probability density functions as $p(\cdot)$ and use the argument to distinguish between the functions. These "inaccuracies" allow for a more compact notation, while ambiguities should not arise as long as the expressions are not taken out of context.



(a) AWGN channel (b) Multiple access channel (c) Interference channel

Figure 2.1: Information theoretic scenarios discussed in this chapter

The channel coding theorem states that the capacity of a channel with an input variable x, an output variable y, and a transmit power constraint as in (2.2) is

$$C = \max_{p(x)} I(x;y) \qquad \text{s.t.:} \quad \mathrm{E}\left[|x|^2\right] \leq 1. \tag{2.6}$$

In words, the channel capacity can be computed by determining the PDF of the input random variable that maximizes the mutual information between the input and output random variable of the channel. The PDF $p(x)$ cannot be chosen freely, however, but must be such that the second absolute moment of x is no larger than one, corresponding to the power constraint (2.2).

The proof of the channel coding theorem consists of two parts. First, it is shown that when the codewords are constructed by drawing random sequences of x, where x has the PDF $p(x)$, and the code rate is below the mutual information $I(x;y)$, for $N \to \infty$ the error probability with a "joint-typicality" decoder, which is fairly simple to analyze, tends towards zero. Second, it is proven that whenever the error probability approaches zero with increasing blocklength N, the number of codewords M in the codebook cannot be higher than e^{NC}, which establishes that rates above C are not achievable.

The PDF $p(x)$ thus has a direct significance for codebooks achieving rates close to capacity: the empirical distribution of $x[t]$ (regardless of the codeword) must be close to $p(x)$. Furthermore, the block length N must be very large in order to be able to achieve a low probability of error.

With (2.6) the capacity of the AWGN channel (2.1) can be shown to be

$$C = \log\left(1 + \frac{|h|^2}{\sigma^2}\right). \tag{2.7}$$

The PDF $p(x)$ that maximizes the mutual information is a complex Gaussian distribution with mean zero and variance one; it is furthermore proper (i. e., circular symmetric). From here on, when referring to a random variable as complex Gaussian, we will imply that it is also proper.

2.2 The Gaussian Multiple Access Channel

Next, we examine a scenario with two transmitters that cannot cooperatively encode their messages and have separate power constraints, and one receiver which would like to reliably decode the messages from both transmitters. Again, we analyze the channel by treating the inputs (i. e., the transmit symbols) as random variables. The Gaussian multiple access channel (MAC) is characterized by the relationship

$$y = h_1 x_1 + h_2 x_2 + n \tag{2.8}$$

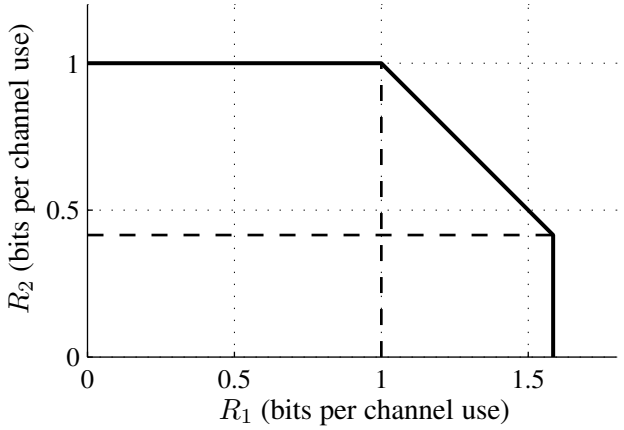


Figure 2.2: The capacity region of the Gaussian MAC with $|h_1|^2 = 2$, $|h_2|^2 = 1$, and $\sigma^2 = 1$. The dashed lines mark the rates $R_1 = \log_2\left(1 + \frac{|h_1|^2}{|h_2|^2+\sigma^2}\right)$ and $R_2 = \log_2\left(1 + \frac{|h_2|^2}{|h_1|^2+\sigma^2}\right)$, which are achievable by the first user in a successive decoding scheme.

where h_1 and h_2 are the complex channel gains from the first and second transmitter to the receiver, n is the complex Gaussian noise variable with variance σ^2, x_1 and x_2 are the input random variables representing the transmitted symbols, and y is the channel output, representing the received symbol. The power constraints on the PDFs of x_1 and x_2 are

$$\mathrm{E}\left[|x_1|^2\right] \leq 1 \qquad \text{and} \qquad \mathrm{E}\left[|x_2|^2\right] \leq 1. \tag{2.9}$$

Furthermore, the random variables x_1 and x_2 are statistically independent, corresponding to the fact that the two transmitters cannot cooperate in choosing the codewords.

For the MAC, the concept of a single achievable rate is not sufficient anymore. Instead, we are interested in the achievable rate pairs (R_1, R_2), where both messages can be recovered from y with arbitrarily low error probability. The *capacity region* is defined as the closure of all achievable rate pairs. Thus, all rate pairs in the interior of the capacity region are achievable, and all rate pairs outside of the capacity region are not achievable.

In [5, Section 14.3], it is shown that a pair of non-negative rates (R_1, R_2) is achievable, if and only if the following inequalities are fulfilled:

$$R_1 < \log\left(1 + \frac{|h_1|^2}{\sigma^2}\right) \tag{2.10}$$

$$R_2 < \log\left(1 + \frac{|h_2|^2}{\sigma^2}\right) \tag{2.11}$$

$$R_1 + R_2 < \log\left(1 + \frac{|h_1|^2 + |h_2|^2}{\sigma^2}\right). \tag{2.12}$$

The capacity region of the Gaussian MAC thus has a pentagonal shape, cf. Figure 2.2.

The justification for the bounds (2.10)–(2.12) is again carried out in two steps: first, it is shown that all rate pairs satisfying the inequalities are achievable; second, it is proven that a rate pair outside of the region cannot be achievable.

For the achievability proof, we assume Gaussian codebooks with unit variance (i. e., equality in (2.9)), which will turn out to be optimal. First, let us consider a decoding scheme where the

receiver decodes the message from the first transmitter while treating $h_2x_2 + n$ as noise. Since x_2 is Gaussian, has variance one, and is independent of n, achievability of R_1 is equivalent to being below the capacity of an AWGN channel with channel gain h_1 and noise power $|h_2|^2 + \sigma^2$. Therefore, $R_1 < \log\left(1 + \frac{|h_1|^2}{|h_2|^2+\sigma^2}\right)$ for this scheme. When the first message is decoded correctly, the receiver can subtract the scaled codeword h_1x_1 from y (for every received symbol of the block) and achievability of R_2 is equivalent to being below the capacity of an AWGN channel with channel gain h_2 and noise power σ^2, i. e., $R_2 < \log\left(1 + \frac{|h_2|^2}{\sigma^2}\right)$. Together, these two bounds on R_1 and R_2 give us a rectangular section of the above capacity region.

In the same way, we can obtain another rectangular section of the capacity region with a scheme in which the second transmitter's message is decoded first and then subtracted from the received block. The remaining triangular section can finally be shown to be achievable with a similar random codebook generation argument as for the AWGN channel. For this part of the region, the decoder must perform a joint estimation of the first and second user's message, however, which is potentially more complex to implement than a successive scheme [5, Section 14.3].

The outer bound argument is fairly straightforward: clearly, setting $h_2 = 0$ cannot reduce the highest achievable rate R_1, which in this case is bounded by (2.10) since the simplification leads to an AWGN channel. In the same way, it is obvious that (2.11) describes an upper bound on any achievable rate R_2. To show that (2.12) is an upper bound on the sum rate, we define $h = \sqrt{|h_1|^2 + |h_2|^2}$ and $x = (h_1x_1 + h_2x_2)/h$. Then $y = hx + n$ and $\mathrm{E}\left[|x|^2\right] \leq 1$ (since x_1 and x_2 are independent). Now, treating x as the transmit symbol of a single transmitter cannot reduce the capacity region, as it is equivalent to allowing some cooperation between the two transmitters when choosing their transmit symbols. Therefore, the combined rate between inputs and output cannot be higher than in this equivalent AWGN channel, which is exactly the sum rate bound (2.12).

The three-user MAC is characterized by the equation

$$y = h_1x_1 + h_2x_2 + h_3x_3 + n \tag{2.13}$$

and the power constraints

$$\mathrm{E}\left[|x_1|^2\right] \leq 1, \qquad \mathrm{E}\left[|x_2|^2\right] \leq 1 \qquad \text{and} \qquad \mathrm{E}\left[|x_3|^2\right] \leq 1. \tag{2.14}$$

With the same arguments as for the two-user MAC, a rate triple (R_1, R_2, R_3) is achievable, if and only if

$$R_1 < \log\left(1 + \frac{|h_1|^2}{\sigma^2}\right) \tag{2.15}$$

$$R_2 < \log\left(1 + \frac{|h_2|^2}{\sigma^2}\right) \tag{2.16}$$

$$R_3 < \log\left(1 + \frac{|h_3|^2}{\sigma^2}\right) \tag{2.17}$$

$$R_1 + R_2 < \log\left(1 + \frac{|h_1|^2 + |h_2|^2}{\sigma^2}\right) \tag{2.18}$$

$$R_1 + R_3 < \log\left(1 + \frac{|h_1|^2 + |h_3|^2}{\sigma^2}\right) \tag{2.19}$$

$$R_2 + R_3 < \log\left(1 + \frac{|h_2|^2 + |h_3|^2}{\sigma^2}\right) \tag{2.20}$$

$$R_1 + R_2 + R_3 < \log\left(1 + \frac{|h_1|^2 + |h_2|^2 + |h_3|^2}{\sigma^2}\right). \tag{2.21}$$

Again, the use of Gaussian codebooks is optimal.

2.3 The Two-User Gaussian Interference Channel

The two-user interference channel in its most general form was first discussed by Shannon [6] and Ahlswede [7]. The defining characteristics of a general two-user interference channel are:

- It has two inputs x_1 and x_2, and two outputs y_1 and y_2.
- It is fully described by the PDFs $p(y_1|x_1, x_2)$ and $p(y_2|x_1, x_2)$ (and possibly power constraints on the PDFs of x_1 and x_2).
- The transmitters cannot cooperate, i. e., x_1 and x_2 are statistically independent.
- The receivers cannot cooperate, i. e., the first receiver's decoder estimates the transmitted codeword using only y_1, the second receiver's decoder only knows y_2.
- The first receiver must only be able to decode the message conveyed by x_1 with vanishing error probability. The second transmitter's codebook is known, however, and this information can be used when decoding. Similarly, the second receiver is only interested in estimating x_2, but can make use of the knowledge of the codebook used for x_1.

The two-user *Gaussian* interference channel (cf. Figure 2.1) is characterized by the two equations

$$y_1 = h_{11}x_1 + h_{12}x_2 + n_1 \tag{2.22}$$

$$y_2 = h_{22}x_2 + h_{21}x_1 + n_2 \tag{2.23}$$

where n_1 and n_2 are complex Gaussian random variables with variance σ^2, and the power constraints

$$\mathrm{E}\left[|x_1|^2\right] \leq 1 \qquad \text{and} \qquad \mathrm{E}\left[|x_2|^2\right] \leq 1. \tag{2.24}$$

A rate pair (R_1, R_2) is achievable, if the message conveyed by x_1 can be recovered from y_1 and the message conveyed by x_2 can be recovered from y_2 with arbitrarily low error probability. Again, the capacity region is the closure of all achievable rate pairs.

The two-user Gaussian interference channel has been the subject of study since Carleial [8, 9]. Up until today, almost half a century after the publication of [6], the capacity region is not known, for either the general or the Gaussian interference channel, with the exception of a few special cases. Nonetheless, significant progress has been made in approximating and understanding the capacity region in recent years.

The methodology for analyzing the interference channel again is to construct achievable schemes and outer bounds on the capacity region. For the MAC, the achievable region and the outer bound were rather straightforward and coincided, giving us a simple description of the capacity region as well as the optimal strategies. For the interference channel, this is not the case. A number of different achievable strategies and outer bounds have been proposed, but in general a gap between the best achievable region and the tightest outer bound remains. In the following we will give a short overview of the relevant results.

2.3.1 Achievable Schemes

Unless stated otherwise, we assume the use of Gaussian codebooks for the following schemes. This is, in contrast to the AWGN and Gaussian MAC, however, not proven to be optimal.

2.3.1.1 Treat Interference as Noise

Perhaps the simplest approach is for the receivers to treat the interference as additional Gaussian noise, reducing the analysis to that of two separate AWGN channels. When the transmitters make use of their full power (equality in (2.24)), the achievable region is rectangular and the achievable rate pairs fulfill

$$R_1 < \log\left(1 + \frac{|h_{11}|^2}{|h_{12}|^2 + \sigma^2}\right) \tag{2.25}$$

$$R_2 < \log\left(1 + \frac{|h_{22}|^2}{|h_{21}|^2 + \sigma^2}\right). \tag{2.26}$$

Intuitively, this is a good strategy when the magnitude of the cross channels is small compared to the noise power, i. e., in a regime of "weak" interference.

The achievable region can be enlarged by allowing the transmitters to decrease their transmit power, by setting $\mathrm{E}\left[|x_k|^2\right] = p_k \leq 1$ for $k \in \{1, 2\}$. For a given pair (p_1, p_2), the achievable rate pairs fulfill

$$R_1 < \log\left(1 + \frac{|h_{11}|^2 p_1}{|h_{12}|^2 p_2 + \sigma^2}\right) \tag{2.27}$$

$$R_2 < \log\left(1 + \frac{|h_{22}|^2 p_2}{|h_{21}|^2 p_1 + \sigma^2}\right). \tag{2.28}$$

By varying p_1 and p_2 over the interval $[0, 1]$ and taking the union of all such achievable regions, we obtain the complete region achievable by treating the interference as noise.

We note that this achievable strategy is somewhat of a special case compared to all following strategies, as the codebook of the interferer must not be known at the receivers. Consequently, it can be of advantage to back off from equality in the power constraints, which is rather unusual in information theoretic analysis.

We also note that (2.27)–(2.28) describe the region achievable by the linear strategies discussed in the following chapters of this work when applied to the two-user single-antenna case with rate utility.

2.3.1.2 Time- (or Frequency-) Division Multiple Access

For the time-division multiple access (TDMA) or frequency-division multiple access (FDMA) region, we partition the transmit blocks into two time slots, or, equivalently, partition the spectrum in the frequency domain into two bands. In each time slot (or frequency band), only one of the transmitters is active. The first time slot, in which only the first transmitter transmits, occupies a fraction of $0 \leq \lambda \leq 1$ of the total time, the second slot occupies a fraction of $1 - \lambda$ of the total time.

Since the power averaged over the whole block consisting of both time slots is constrained, the instantaneous power used during the active time slot can be increased accordingly. Therefore, we have an AWGN channel with a power constraint of $1/\lambda$ for the first, and $1/(1 - \lambda)$ for the second time slot. The instantaneous rates must be normalized accordingly, in order to reflect the information throughput normalized by the length of the whole block.

For a fixed time-sharing parameter $\lambda \in (0,1)$, the region is rectangular, and is characterized by

$$R_1 < \lambda \log\left(1 + \frac{|h_{11}|^2}{\lambda\sigma^2}\right) \tag{2.29}$$

$$R_2 < (1-\lambda) \log\left(1 + \frac{|h_{22}|^2}{(1-\lambda)\sigma^2}\right). \tag{2.30}$$

The union over all $\lambda \in (0,1)$ is the achievable TDMA/FDMA region.

2.3.1.3 Fully Decode Interference

In contrast to treating the interference as noise, the receivers could decide to completely decode the interference along with the desired signal. Then, from the viewpoint of each of the two receivers, the channel is a Gaussian MAC. A rate pair is achievable with this scheme, if it is achievable in both of the receivers' MACs. Therefore, the achievable region is the intersection of the capacity regions of the two MACs (cf. (2.10)–(2.12)), leading to the description

$$R_1 < \min\left\{\log\left(1 + \frac{|h_{11}|^2}{\sigma^2}\right), \log\left(1 + \frac{|h_{21}|^2}{\sigma^2}\right)\right\} \tag{2.31}$$

$$R_2 < \min\left\{\log\left(1 + \frac{|h_{22}|^2}{\sigma^2}\right), \log\left(1 + \frac{|h_{12}|^2}{\sigma^2}\right)\right\} \tag{2.32}$$

$$R_1 + R_2 < \min\left\{\log\left(1 + \frac{|h_{11}|^2 + |h_{12}|^2}{\sigma^2}\right), \log\left(1 + \frac{|h_{21}|^2 + |h_{22}|^2}{\sigma^2}\right)\right\}. \tag{2.33}$$

This region is in general a pentagon, but can (in contrast to the MAC) also be rectangular, if the right-hand side of (2.33) is larger than or equal to the sum of the right-hand sides of (2.31) and (2.32).

Intuitively, this is a good strategy if the magnitude of the cross channels is high compared to the magnitude of the direct channels, i. e., when interference is strong. Otherwise, ensuring that the signal is fully decodable at the unintended receiver is an unnecessarily strict requirement.

In a highly asymmetric setting, a variation of this strategy might be advantageous, in which one receiver decodes the interference and the other receiver treats interference as noise. When the first receiver decodes, we obtain

$$R_1 < \log\left(1 + \frac{|h_{11}|^2}{\sigma^2}\right) \tag{2.34}$$

$$R_2 < \min\left\{\log\left(1 + \frac{|h_{22}|^2}{|h_{21}|^2 + \sigma^2}\right), \log\left(1 + \frac{|h_{12}|^2}{\sigma^2}\right)\right\} \tag{2.35}$$

$$R_1 + R_2 < \log\left(1 + \frac{|h_{11}|^2 + |h_{12}|^2}{\sigma^2}\right). \tag{2.36}$$

2.3.1.4 Han-Kobayashi Scheme

In 1981, Han and Kobayashi [10] proposed a very general strategy, which encompasses all of the previously discussed schemes as extreme cases. First of all, a discrete time-sharing random variable q is introduced, which switches between different modes of transmission, similar to the two time slots in the TDMA scheme. Furthermore, the information conveyed by the channel input

x_k is split into two parts: a *private* and a *common* message, represented by the random variables $x_{\mathrm{P},k}$ and $x_{\mathrm{C},k}$, the PDF of which depends on the transmission mode (or time slot) q. The private message is meant to be treated as noise at the unintended receiver, the common message is meant to be fully decoded at the unintended receiver. The channel input is generated by a mapping function $x_k = f_k(x_{\mathrm{P},k}, x_{\mathrm{C},k}, q)$ from the private and common messages. The mapping function also depends on the transmission mode q.

For a given set of input distributions and mappings, and a fixed time-sharing parameter q, the achievable region can be computed with some effort: from the viewpoint of each of the two receivers, it is a three-user MAC with the own private and both common messages as inputs. However, the complete Han-Kobayashi region is the union over

- all possible PDFs of $x_{\mathrm{P},k}$ and $x_{\mathrm{C},k}$,
- all possible time-sharing schemes (i. e., probability mass functions of q),
- and all possible mapping functions $f_k(x_{\mathrm{P},k}, x_{\mathrm{C},k}, q)$ with $k \in \{1, 2\}$ ensuring that the power constraints are fulfilled,

and is therefore clearly not easily expressed.

A simplified subregion was also proposed in [10], in which the following additional constraints are imposed:

- The mapping function is the summation of the common and private message, i. e., $x_k = x_{\mathrm{P},k} + x_{\mathrm{C},k}$,
- $x_{\mathrm{P},k}$ and $x_{\mathrm{C},k}$ (and consequently also x_k) have a complex Gaussian PDF,
- $\mathrm{E}\left[|x_{\mathrm{P},k}|^2\right] = \alpha_k$ and $\mathrm{E}\left[|x_{\mathrm{C},k}|^2\right] = 1 - \alpha_k$,
- and q is deterministic, i. e., no time-sharing between different transmission modes is employed.

For fixed α_1 and α_2 from the interval $[0, 1]$, we can now express the boundaries of the achievable region. Viewing the whole region as the intersection of the two Gaussian three-user MAC regions seen by the first and second receiver, we obtain seven conditions on $R_{\mathrm{C},1}$, $R_{\mathrm{P},1}$, and $R_{\mathrm{C},2}$, and another seven conditions on $R_{\mathrm{C},1}$, $R_{\mathrm{C},2}$, and $R_{\mathrm{P},2}$ by applying (2.15)–(2.21). By means of *Fourier-Motzkin eliminitation* (e. g., [11, Section 12.2]), these conditions can be restated as a set of conditions on $R_1 = R_{\mathrm{C},1} + R_{\mathrm{P},1}$ and $R_2 = R_{\mathrm{C},2} + R_{\mathrm{P},2}$. The resulting region is a polyhedron confined by conditions on R_1, R_2, $R_1 + R_2$, $2R_1 + R_2$, and $R_1 + 2R_2$, defined as follows:

$$R_1 < C_3 \tag{2.37}$$
$$R_1 < C_1 + C_7 \tag{2.38}$$
$$R_2 < C_8 \tag{2.39}$$
$$R_2 < C_2 + C_6 \tag{2.40}$$
$$R_1 + R_2 < C_1 + C_{10} \tag{2.41}$$
$$R_1 + R_2 < C_4 + C_9 \tag{2.42}$$
$$R_1 + R_2 < C_5 + C_6 \tag{2.43}$$
$$2R_1 + R_2 < C_1 + C_5 + C_9 \tag{2.44}$$
$$R_1 + 2R_2 < C_4 + C_6 + C_{10} \tag{2.45}$$

where the following abbreviations were used:

$$C_1 = \log\left(1 + \frac{|h_{11}|^2\alpha_1}{|h_{12}|^2\alpha_2 + \sigma^2}\right) \tag{2.46}$$

$$C_2 = \log\left(1 + \frac{|h_{12}|^2(1-\alpha_2)}{|h_{12}|^2\alpha_2 + \sigma^2}\right) \tag{2.47}$$

$$C_3 = \log\left(1 + \frac{|h_{11}|^2}{|h_{12}|^2\alpha_2 + \sigma^2}\right) \tag{2.48}$$

$$C_4 = \log\left(1 + \frac{|h_{11}|^2\alpha_1 + |h_{12}|^2(1-\alpha_2)}{|h_{12}|^2\alpha_2 + \sigma^2}\right) \tag{2.49}$$

$$C_5 = \log\left(1 + \frac{|h_{11}|^2 + |h_{12}|^2(1-\alpha_2)}{|h_{12}|^2\alpha_2 + \sigma^2}\right) \tag{2.50}$$

$$C_6 = \log\left(1 + \frac{|h_{22}|^2\alpha_2}{|h_{21}|^2\alpha_1 + \sigma^2}\right) \tag{2.51}$$

$$C_7 = \log\left(1 + \frac{|h_{21}|^2(1-\alpha_1)}{|h_{21}|^2\alpha_1 + \sigma^2}\right) \tag{2.52}$$

$$C_8 = \log\left(1 + \frac{|h_{22}|^2}{|h_{21}|^2\alpha_1 + \sigma^2}\right) \tag{2.53}$$

$$C_9 = \log\left(1 + \frac{|h_{22}|^2\alpha_2 + |h_{21}|^2(1-\alpha_1)}{|h_{21}|^2\alpha_1 + \sigma^2}\right) \tag{2.54}$$

$$C_{10} = \log\left(1 + \frac{|h_{22}|^2 + |h_{21}|^2(1-\alpha_1)}{|h_{21}|^2\alpha_1 + \sigma^2}\right). \tag{2.55}$$

The union over these polyhedra for all possible values of α_1 and α_2, finally, is the *simplified Han-Kobayashi region.*

The strategy of treating interference as noise is equivalent to giving all of the power to the private message, i. e., setting $\alpha_1 = \alpha_2 = 1$; when the interference is meant to be fully decoded, the full power is given to the common message, i. e., $\alpha_1 = \alpha_2 = 0$. The TDMA scheme, on the other hand, is not necessarily contained in the simplified Han-Kobayashi region, since it requires a time-sharing random variable q with two states, the first one with probability λ, and the second one with probability $1 - \lambda$.

2.3.1.5 Sason's Scheme

Another simplified subregion of the complete Han-Kobayashi region was introduced by Sason in [12]. Here, time-sharing is allowed between two modes, i. e., the random variable q has two possible outcomes. However, in contrast to the pure TDMA strategy, in each time slot both transmitters are active. Instead of splitting the information into a private and a common part and jointly decoding the own private and both common messages at the receivers, a simple successive decoding scheme is employed: in the first transmission mode, the first user is prioritized and transmits at the highest possible rate; the second user must transmit at a rate that is low enough so that receiver one can decode the interference first and subtract it from the received signal. In the second transmission mode, the second user is prioritized, and the first user transmits at a sufficiently low rate.

For this scheme there are three free parameters that can be chosen from the interval $[0, 1]$: a time-sharing coefficient λ, and two power splitting coefficients β_1 and β_2. The first transmission mode occupies a fraction λ of the total time. In the first mode, user k uses a fraction β_k of its full power. In the second mode, which occupies a fraction of $1 - \lambda$ of the time, the users use $1 - \beta_k$ of their power. If $\beta_k = \lambda$, user k uses the same instantaneous power in both transmission modes.

The achievable region, which is rectangular for a given set of parameters, consequently can be described by

$$\begin{aligned} R_1 < {} & \lambda \log\left(1 + \frac{|h_{11}|^2\beta_1}{\lambda\sigma^2}\right) \\ & + (1-\lambda) \cdot \min\left\{\log\left(1 + \tfrac{|h_{11}|^2(1-\beta_1)}{|h_{12}|^2(1-\beta_2)+(1-\lambda)\sigma^2}\right), \log\left(1 + \tfrac{|h_{21}|^2(1-\beta_1)}{|h_{22}|^2(1-\beta_2)+(1-\lambda)\sigma^2}\right)\right\} \end{aligned} \tag{2.56}$$

$$\begin{aligned} R_2 < {} & (1-\lambda) \log\left(1 + \frac{|h_{22}|^2(1-\beta_2)}{(1-\lambda)\sigma^2}\right) \\ & + \lambda \cdot \min\left\{\log\left(1 + \tfrac{|h_{22}|^2\beta_2}{|h_{21}|^2\beta_1+\lambda\sigma^2}\right), \log\left(1 + \tfrac{|h_{12}|^2\beta_2}{|h_{11}|^2\beta_1+\lambda\sigma^2}\right)\right\}. \end{aligned} \tag{2.57}$$

Again, the complete achievable region with Sason's strategy is the union of these rectangles for all possible parameters λ, β_1, and β_2. It is contained in the complete Han-Kobayashi region, as it results from a number of simplifications that can only reduce the achievable region: the random variable q has only two outcomes; for the kth outcome, user k's power is given fully to the private message and the other user's power is given fully to the common message; and the receivers can only perform successive decoding. It is not necessarily contained in the simplified Han-Kobayashi region, however, since time-sharing is allowed.

The TDMA region is fully contained in Sason's achievable region, as it follows from setting $\beta_1 = 1$ and $\beta_2 = 0$ and varying λ. A similar case to the previously discussed scheme where one receiver decodes the interference and the other one treats it as noise is also contained by setting $\lambda = \beta_1 = \beta_2 = 1$; with the difference, however, that here only successive decoding is allowed.

2.3.1.6 Chong-Motani-Garg Region

For the Han-Kobayashi strategy it is required that the common message of the interfering transmitter is decodable with vanishing error probability, even though the contained information is subsequently discarded by the receiver. In [13], Chong, Motani, and Garg showed that by dropping this condition, the region achievable by any given Han-Kobayashi strategy with fixed input distributions and mapping functions can be slightly enlarged, while still being contained in the complete Han-Kobayashi region. Specifically, when the simplified Han-Kobayashi strategy is employed with fixed parameters α_1 and α_2, the conditions (2.38) and (2.40) on R_1 and R_2 can be dropped; the conditions on $R_1 + R_2$, $2R_1 + R_2$, and $R_1 + 2R_2$, on the other hand, remain the same, as does the result of taking the union over all α_1 and α_2.

2.3.2 Outer Bounds and Optimality Results

The best known achievable region for the Gaussian interference channel is the complete Han-Kobayashi region. So far no outer bound has been found that in general coincides with the boundary of the complete Han-Kobayashi region. As we will discuss in the following, however, for the case of *strong* interference, a fairly simple outer bound does coincide with a simple achievable scheme;

for other cases, some more complex outer bounds can be formulated, that at least come close to certain achievable schemes.

First of all, we can state the simple rectangular outer bound that follows from assuming that no interference is present:

$$R_1 < \log\left(1 + \frac{|h_{11}|^2}{\sigma^2}\right) \tag{2.58}$$

$$R_2 < \log\left(1 + \frac{|h_{22}|^2}{\sigma^2}\right). \tag{2.59}$$

Note that the corner points of this bound where one of the two rates is zero are always achievable by disabling one user.

2.3.2.1 Strong and Very Strong Interference

If $|h_{21}|^2 \geq |h_{11}|^2$ and $|h_{12}|^2 \geq |h_{22}|^2$, the interference is said to be *strong* and the strategy of fully decoding the interference achieves all rate pairs in the interior of the capacity region. This was proven independently in [10] and [14].

The outer bound argument is fairly intuitive: first, we set $h_{21} = 0$, which cannot reduce the capacity region since it only makes it easier for the second receiver to decode its message. Now, let us assume that we have an achievable rate pair of this new "one-sided" interference channel. By definition, the first receiver can decode its message with vanishing error probability, and can therefore also subtract it from the received signal, leaving only $h_{12}x_2 + n_1$. Also by definition, the second receiver can decode its message from $h_{22}x_2 + n_2$. Since $|h_{12}| \geq |h_{22}|$, receiver one must therefore also be able to fully decode the interference, i. e., any rate pair that is achievable in the one-sided interference channel, must be within the capacity region of the MAC seen by receiver one.

From (2.12), we thus obtain the upper sum rate bound $R_1 + R_2 < \log\left(1 + \frac{|h_{11}|^2+|h_{12}|^2}{\sigma^2}\right)$. In the same way, by setting $h_{12} = 0$, we can show that $R_1 + R_2 < \log\left(1 + \frac{|h_{22}|^2+|h_{21}|^2}{\sigma^2}\right)$. Therefore, (2.33) coincides with our outer bound. Also, the outer bounds (2.58) and (2.59) coincide with (2.31) and (2.32).

We can furthermore define the special case of *very strong* interference, in which the capacity region becomes rectangular. This is the case when the sum of the right-hand sides of (2.58) and (2.59) is less than or equal to the right-hand side of (2.33). The conditions for very strong interference can thus be shown to be

$$\frac{|h_{12}|^2}{|h_{22}|^2} \geq 1 + \frac{|h_{11}|^2}{\sigma^2} \tag{2.60}$$

$$\frac{|h_{21}|^2}{|h_{11}|^2} \geq 1 + \frac{|h_{22}|^2}{\sigma^2}. \tag{2.61}$$

In this regime, the capacity region notably is the same as if no interference was present at all, i. e., as if $h_{12} = h_{21} = 0$.

2.3.2.2 Sato Bound

In 1977, Sato [15] presented an outer bound on the sum rate $R_1 + R_2$ that is applicable regardless of the magnitude of the channel parameters. The outer bound is based on the observation that allowing

the receivers to cooperate in decoding the messages cannot reduce the capacity region. The capacity region with cooperative receivers depends on the correlation between the noise variables n_1 and n_2, whereas in the original interference channel the noise correlation clearly has no effect on the achievable rates. Therefore, even with the worst possible noise correlation coefficient, the capacity region with cooperative receivers must still be larger than that of the interference channel. The Sato bound is simply the sum rate that can be achieved in a system with cooperative receivers and the worst possible noise correlation.

With cooperative receivers, the scenario is a MAC with a vector-valued output random variable. The capacity region arguments are very similar to those for the scalar MAC, the expressions for the mutual information, however, are a bit different. For a given noise correlation, the sum rate must satisfy (cf. [5, Chapter 9])

$$R_1 + R_2 < \log\det\left(\mathbf{I} + \boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}^{-1}\boldsymbol{H}\right) \tag{2.62}$$

where

$$\boldsymbol{H} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{R} = \sigma^2 \cdot \begin{bmatrix} 1 & r \\ r^* & 1 \end{bmatrix} \tag{2.63}$$

and $r = \frac{1}{\sigma^2} \cdot \mathrm{E}\left[n_1 n_2^*\right]$ is the noise correlation coefficient. Note that $|r| \leq 1$, and that $|r| = 1$ yields infinite achievable sum rate (with the exception of a special case in which $\boldsymbol{H}$ is rank deficient, which we will ignore here). The worst possible noise correlation coefficient is defined by

$$\bar{r} = \arg\min_{r} \ \log\det\left(\mathbf{I} + \boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}^{-1}\boldsymbol{H}\right) \qquad \text{s.t.:} \quad |r| \leq 1. \tag{2.64}$$

The solution to this optimization problem is

$$\bar{r} = \frac{d_1 - \sqrt{d_1^2 - 4|d_2|^2}}{2d_2^*} \tag{2.65}$$

with

$$d_1 = \mathrm{tr}\left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\sigma^{-2}\right) + \det\left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\sigma^{-2}\right) \qquad \text{and} \tag{2.66}$$

$$d_2 = \left(h_{11}h_{21}^* + h_{12}h_{22}^*\right)\sigma^{-2}, \tag{2.67}$$

resulting in the outer bound

$$R_1 + R_2 < \log\left(1 + \frac{2|d_2|^2}{d_1 - \sqrt{d_1^2 - 4|d_2|^2}}\right). \tag{2.68}$$

The derivation of the expression (2.65) for $\bar{r}$ can be found in Appendix A1. Note that the Sato bound, in contrast to the other bounds in this section, depends on the complex phase of the channel coefficients in addition to their absolute value.

2.3.2.3 Carleial's Bound

The idea of another historically relevant outer bound, published by Carleial in 1983 [16], is to increase the magnitudes of the channel gains in a way that cannot reduce the capacity region until they fulfill the conditions for strong interference.

Let us assume, e. g., that $0 < |h_{21}|^2 < |h_{11}|^2$, i. e., the first transmitter's interference is "weak" in the sense that it does not fulfill the condition for strong interference. Now, we replace h_{22} by $\frac{|h_{11}|}{|h_{21}|}h_{22}$ and h_{21} by $\frac{|h_{11}|}{|h_{21}|}h_{21}$. We have thus increased h_{22} and h_{21} by a common factor while leaving the noise and transmit power constraint the same. This can clearly only be an improvement, as it is equivalent to decreasing the noise power at the second receiver.

Now, the channel coefficient from transmitter one to receiver two has the same magnitude as h_{11} and we can give an outer bound on the sum rate with the same argument as in the strong interference case; an achievable rate pair cannot be outside of the region of the MAC seen by receiver two:

$$R_1 + R_2 < \log\left(1 + \frac{\frac{|h_{11}|^2}{|h_{21}|^2}|h_{22}|^2 + |h_{11}|^2}{\sigma^2}\right). \tag{2.69}$$

Similarly, if $0 < |h_{12}|^2 < |h_{22}|^2$, we can also state the above with exchanged indices 1 and 2.

2.3.2.4 "One Bit" Bound

Recently, a new outer bound was formulated by Etkin, Tse, and Wang [17], which was remarkably proven to be no further than one bit (or a factor of two inside the logarithm) away from a certain achievable strategy, and therefore at most one bit from the true capacity region. The "one bit" bound is a "genie-aided" outer bound, i. e., it is based on providing the receivers with side information that they do not have in the actual channel model, but which yields manageable expressions for the capacity region. The derivation of the bound is rather lengthy; we therefore only give the result here and refer the reader to [17] for the proof, which relies on Fano's inequality [5] in a similar way as the converse of the channel coding theorem (AWGN).

For the "one bit" bound, we distinguish between two cases: *weak* interference and *mixed* interference. The interference is considered weak if $|h_{21}|^2 < |h_{11}|^2$ and $|h_{12}|^2 < |h_{22}|^2$. The resulting bound consists of the following expressions:

$$R_1 + R_2 < \log\left(1 + \frac{|h_{11}|^2}{\sigma^2}\right) + \log\left(1 + \frac{|h_{22}|^2}{|h_{21}|^2 + \sigma^2}\right) \tag{2.70}$$

$$R_1 + R_2 < \log\left(1 + \frac{|h_{22}|^2}{\sigma^2}\right) + \log\left(1 + \frac{|h_{11}|^2}{|h_{12}|^2 + \sigma^2}\right) \tag{2.71}$$

$$R_1 + R_2 < \log\left(1 + \frac{|h_{12}|^2}{\sigma^2} + \frac{|h_{11}|^2}{|h_{21}|^2 + \sigma^2}\right) + \log\left(1 + \frac{|h_{21}|^2}{\sigma^2} + \frac{|h_{22}|^2}{|h_{12}|^2 + \sigma^2}\right) \tag{2.72}$$

$$\begin{aligned} 2R_1 + R_2 < {} & \log\left(1 + \frac{|h_{11}|^2 + |h_{12}|^2}{\sigma^2}\right) + \log\left(1 + \frac{|h_{21}|^2}{\sigma^2} + \frac{|h_{22}|^2}{|h_{12}|^2 + \sigma^2}\right) \\ & + \log\left(\frac{|h_{11}|^2 + \sigma^2}{|h_{21}|^2 + \sigma^2}\right) \end{aligned} \tag{2.73}$$

$$\begin{aligned} R_1 + 2R_2 < {} & \log\left(1 + \frac{|h_{22}|^2 + |h_{21}|^2}{\sigma^2}\right) + \log\left(1 + \frac{|h_{12}|^2}{\sigma^2} + \frac{|h_{11}|^2}{|h_{21}|^2 + \sigma^2}\right) \\ & + \log\left(\frac{|h_{22}|^2 + \sigma^2}{|h_{12}|^2 + \sigma^2}\right). \end{aligned} \tag{2.74}$$

The interference is considered mixed, if either $|h_{12}|^2 \geq |h_{22}|^2$ or $|h_{21}|^2 \geq |h_{11}|^2$. Here, we give the expressions for $|h_{21}|^2 < |h_{11}|^2$ and $|h_{12}|^2 \geq |h_{22}|^2$; the description of the outer bound for the

other mixed interference case can be determined by exchanging the indices 1 and 2.

$$R_1 + R_2 < \log\left(1 + \frac{|h_{11}|^2}{\sigma^2}\right) + \log\left(1 + \frac{|h_{22}|^2}{|h_{21}|^2 + \sigma^2}\right) \tag{2.75}$$

$$R_1 + R_2 < \log\left(1 + \frac{|h_{11}|^2 + |h_{12}|^2}{\sigma^2}\right) \tag{2.76}$$

$$\begin{aligned} R_1 + 2R_2 < \log\left(1 + \frac{|h_{22}|^2 + |h_{21}|^2}{\sigma^2}\right) + \log\left(1 + \frac{|h_{12}|^2}{\sigma^2} + \frac{|h_{11}|^2}{|h_{21}|^2 + \sigma^2}\right) \\ + \log\left(1 + \frac{|h_{22}|^2}{|h_{12}|^2 + \sigma^2}\right). \end{aligned} \tag{2.77}$$

To show that the distance to the capacity region is at most one bit, the authors compare the outer bound with the achievable region obtained by using the Chong-Motani-Garg strategy with the following specific choice of power splitting parameters:

$$\alpha_1 = \min\left\{\frac{\sigma^2}{|h_{21}|^2}, 1\right\} \qquad \text{and} \qquad \alpha_2 = \min\left\{\frac{\sigma^2}{|h_{12}|^2}, 1\right\}. \tag{2.78}$$

The intuition behind this choice is that the private message should be at most as strong as the noise at the unintended receiver, in order to not be easily decodable; if the private message was easily decodable at the unintended receiver, redistributing the power in favor of the common message would be of advantage, as this would facilitate joint decoding at the unintended receiver.

The Chong-Motani-Garg region for one specific choice of α_1 and α_2 is defined by bounds on R_1, R_2, $R_1 + R_2$, $2R_1 + R_2$, and $R_1 + 2R_2$, similar to the above outer bound. With this description of the achievable region and the outer bound, it can be shown that for any pair (R_1, R_2) on the boundary of the Chong-Motani-Garg region, the pair $(R_1 + \log 2, R_2 + \log 2)$ is on or outside the above outer bound. Since the capacity region clearly increases in size with decreasing noise power σ^2, the one bit gap becomes less significant and this bound gives us an asymptotically accurate description of the capacity region for the case of vanishing noise power.

2.3.2.5 Recent Bounds by Kramer et Al. and Sum-Rate Optimality at Noisy Interference

In [18] and [19], a different set of outer bounds on the capacity region was proposed, which even are tighter than the "one bit" bound in some cases, but have not been proven to be within a fixed distance of the capacity region. The bounds are defined by rather complex expressions and are therefore omitted here.

A notable consequence from the bounds in [19] is that in a *noisy* interference regime, characterized by the condition

$$\frac{|h_{12}|}{|h_{11}|}|h_{21}|^2 + \frac{|h_{21}|}{|h_{22}|}|h_{12}|^2 \leq \sigma^2\left(1 - \frac{|h_{12}|}{|h_{11}|} - \frac{|h_{21}|}{|h_{22}|}\right), \tag{2.79}$$

treating interference as noise and employing Gaussian codebooks is sum-rate optimal, i. e., in this case $R_1 + R_2$ cannot be higher than the sum of the right-hand sides of (2.25) and (2.26).

2.3.2.6 Recent Bounds by Khandani et Al. and Sum-Rate Optimality at Mixed Interference

In [20], new bounds for the weak and mixed interference cases are derived that further tighten the "one bit" bound. In the same work, it is also shown that the scheme of fully decoding one user's

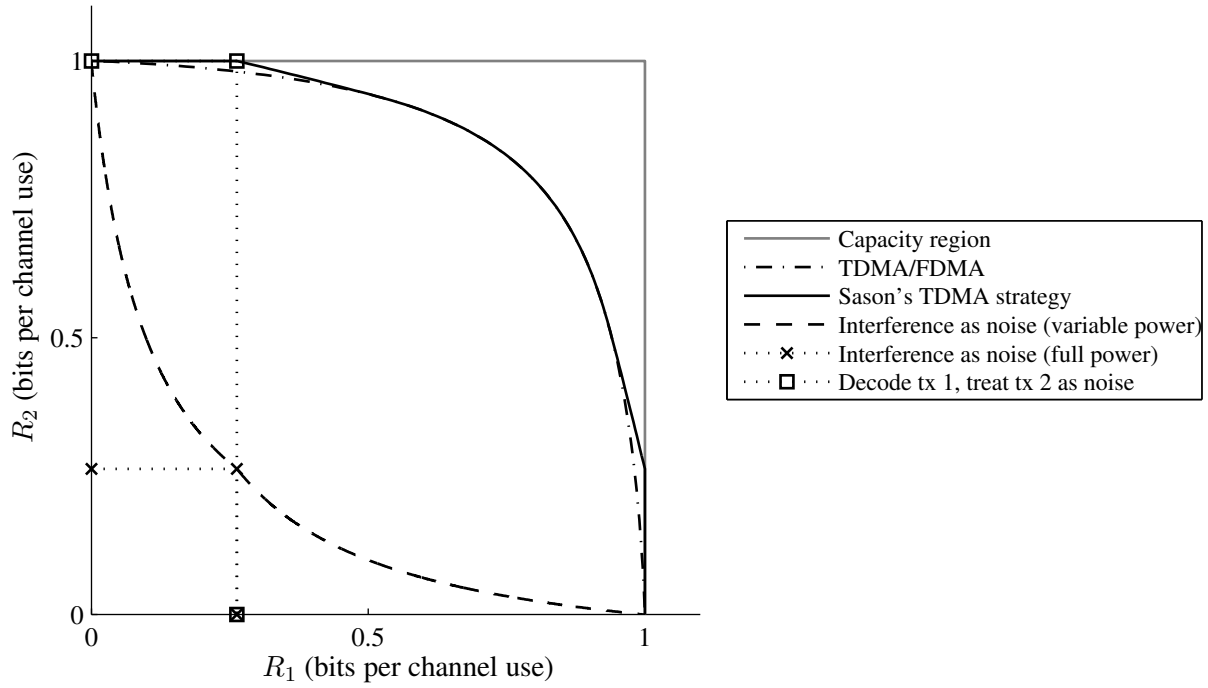


Figure 2.3: The capacity region and some suboptimal achievable schemes for a scenario with very strong interference. The squared channel magnitudes are $|h_{11}|^2 = |h_{22}|^2 = 1$ and $|h_{12}|^2 = |h_{21}|^2 = 4$, the noise power is $\sigma^2 = 1$. The strategy of fully decoding the interference achieves the capacity region (which is also identical to the Han-Kobayashi region).

interference and treating the other user's interference as noise while applying Gaussian codebooks is sum-rate optimal in the mixed interference case. If, e. g., $|h_{12}|^2 \geq |h_{22}|^2$ and $|h_{21}|^2 < |h_{11}|^2$, the strategy described at the end of Section 2.3.1.3 and resulting in the achievable region defined by (2.34)–(2.36) is optimal in terms of $R_1 + R_2$. The maximum sum rate in this case can be equivalently stated as:

$$R_1 + R_2 < \log\left(1 + \frac{|h_{11}|^2}{\sigma^2}\right) + \min\left\{\log\left(1 + \frac{|h_{12}|^2}{|h_{11}|^2 + \sigma^2}\right), \log\left(1 + \frac{|h_{22}|^2}{|h_{21}|^2 + \sigma^2}\right)\right\}. \quad (2.80)$$

2.3.3 Numerical Examples for Bounds on the Capacity Region

In this section we will present some numerical examples for the previously discussed inner and outer bounds in order to provide an impression of the wide range of possibilities for the qualitative behavior of the shape of the capacity region and its bounds.

The scenario in Figure 2.3 is a Gaussian interference channel with symmetric coefficients fulfilling the conditions for very strong interference (2.60) and (2.61). The capacity region is known to be rectangular and the same as if no interference was present at all. It can be achieved by fully decoding the interference first and then subtracting it from the received signal, which is also a special Han-Kobayashi strategy. Treating interference as noise, on the other hand, only achieves a small part of the capacity region, while the orthogonalization via TDMA or FDMA performs better, but also does not come close to achieving the optimal sum rate. The gain from allowing the extra degrees of freedom afforded by Sason's scheme turns out to be quite small.

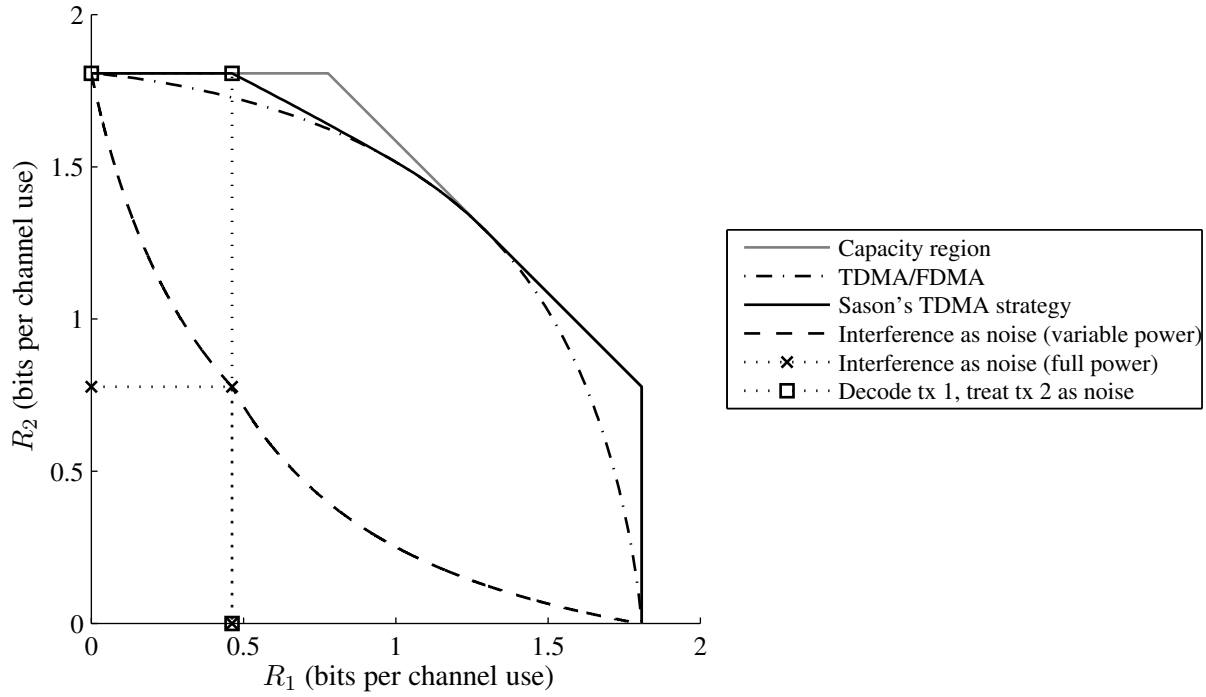


Figure 2.4: The capacity region and some suboptimal achievable schemes for a scenario with strong interference. The squared channel magnitudes are $|h_{11}|^2 = |h_{22}|^2 = |h_{21}|^2 = 1$ and $|h_{12}|^2 = 2.25$, the noise power is $\sigma^2 = 0.4$.

An asymmetric interference channel which has strong, but not very strong, interference is shown in Figure 2.4. The capacity region is pentagonal and can again be achieved by fully decoding the interference and therefore also by a specific simple Han-Kobayashi strategy. Here, though, TDMA also achieves the optimal sum rate, and Sason's scheme only misses a small section of the capacity region. Again, treating interference as noise is clearly suboptimal.

When the condition for noisy interference (2.79) is fulfilled, the capacity region is not known, but it is known that treating interference as noise is sum-rate optimal. As can be seen in Figure 2.5, the schemes not attempting to decode any of the interference (which are also contained in the simplified Han-Kobayashi region), come reasonably close to the best outer bound, while forcing the interference to be decodable results in a significant loss.

The scenarios in Figures 2.6, 2.7, and 2.8 fulfill neither the strong nor the noisy interference conditions, so that the boundary of the true capacity region may be anywhere between the tightest outer bound and the best achievable strategy. In Figure 2.6, the specific Chong-Motani-Garg achievable region, for which the one bit gap to the outer bound of [17] is proven, is also shown. It can furthermore be seen from the figures that the Sato bound can, but must not, outperform Carleial's bound (Figure 2.7 vs. Figures 2.6 and 2.8), that TDMA can be an improvement over the simplified Han-Kobayashi region (Figure 2.7), and that having one receiver decode its interference while the other one treats it as noise is sum-rate optimal in a scenario with mixed interference (Figure 2.8).

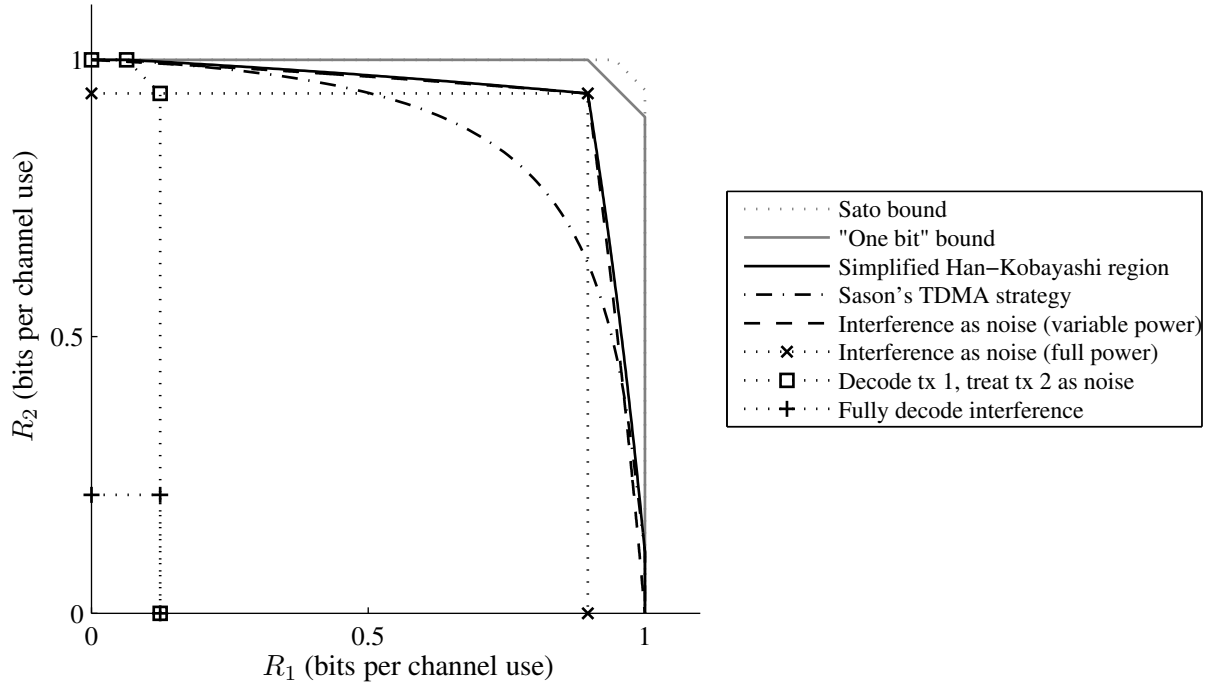


Figure 2.5: Some outer bounds and achievable schemes for a scenario with noisy interference. The channel coefficients are $h_{11} = h_{22} = 1$, $h_{12} = 0.4$, and $h_{21} = 0.3$, the noise power is $\sigma^2 = 1$. Treating interference as noise is sum-rate optimal in this case. Carleial's bound does not yield an improvement over the zero-interference outer bound.

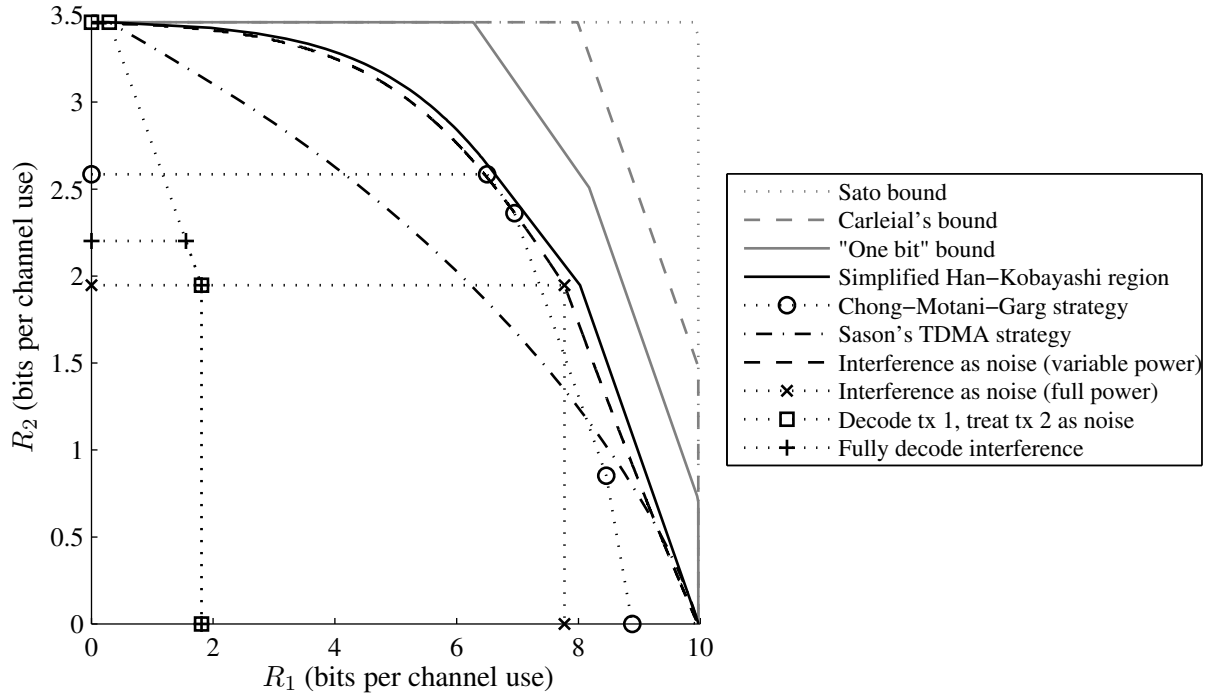


Figure 2.6: Some outer bounds and achievable schemes for a scenario with weak interference. The channel coefficients are $h_{11} = 10$, $h_{22} = 1$, $h_{12} = 0.6$, and $h_{21} = 0.5$, the noise power is $\sigma^2 = 0.1$. The Chong-Motani-Garg strategy is used with the fixed power splitting of (2.78).

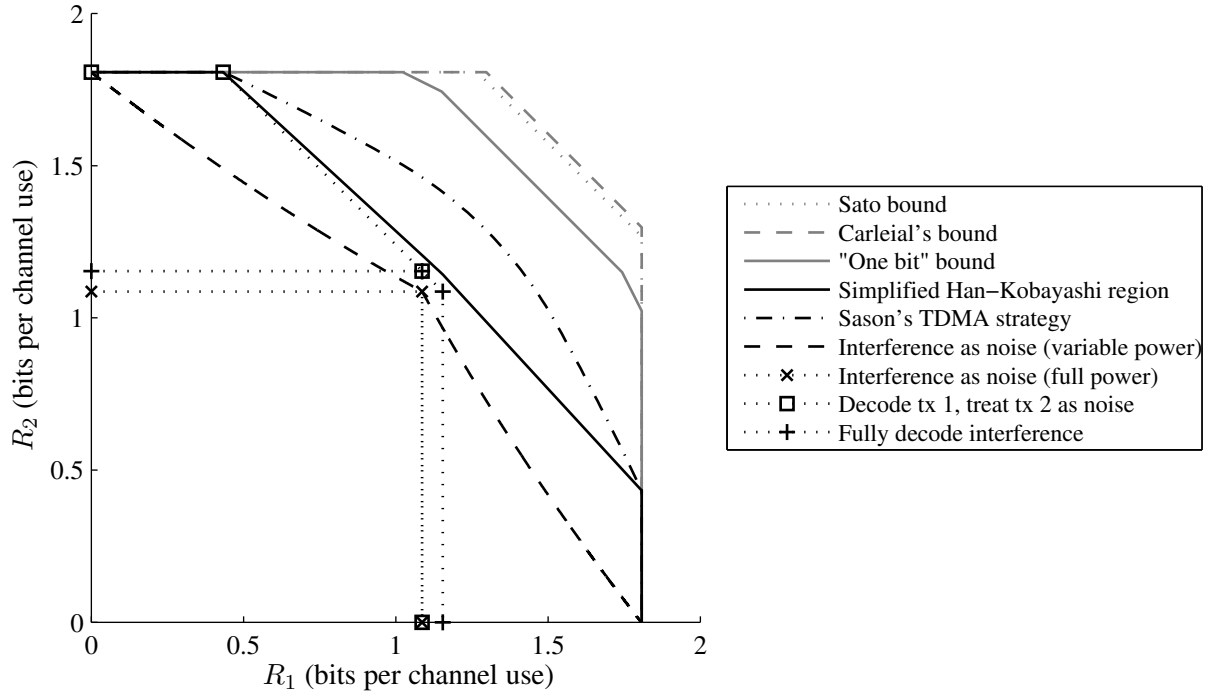


Figure 2.7: Another scenario with weak interference, in which TDMA yields a larger region than the simplified Han-Kobayashi scheme. The channel coefficients are $h_{11} = h_{22} = 1$ and $h_{12} = h_{21} = 0.7$, the noise power is $\sigma^2 = 0.4$.

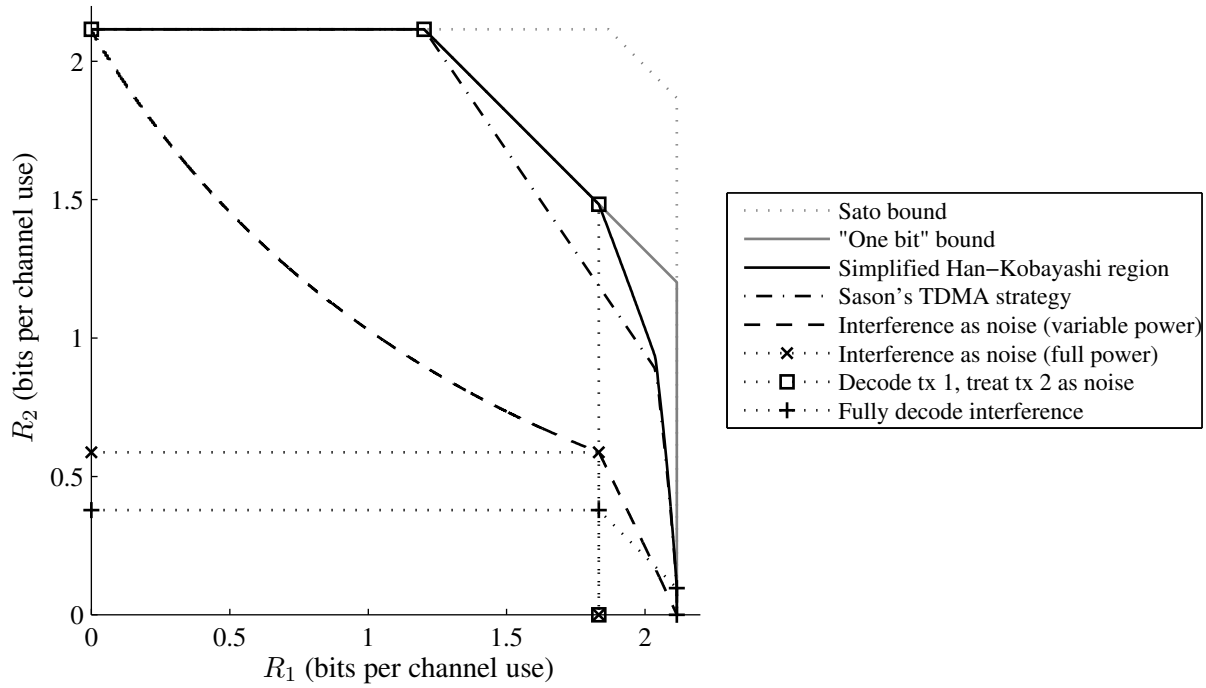


Figure 2.8: A scenario with mixed interference, in which it is sum-rate optimal to fully decode the first transmitter's signal at both receivers, while treating the interference caused by the second transmitter as noise. The channel coefficients are $h_{11} = h_{22} = 1$, $h_{12} = 0.3$ and $h_{21} = 1.3$, the noise power is $\sigma^2 = 0.3$. Carleial's bound coincides with the "one bit" bound in this case.

2.3.4 The Symmetric Interference Channel

We define the *symmetric* Gaussian interference channel as a Gaussian interference channel with $|h_{11}|^2 = |h_{22}|^2$ and $|h_{12}|^2 = |h_{21}|^2$. Examining the symmetric case allows us to more concisely investigate the effect of the "strength" of the interference, while disregarding the many complexities that arise, for example, in the mixed interference cases.

By simplifying (2.79) we observe that the symmetric interference channel is in the regime of noisy interference if

$$\sigma^2 \geq \frac{|h_{12}|^2}{\frac{|h_{11}|}{2|h_{12}|} - 1}. \tag{2.81}$$

This can clearly only be fulfilled if $|h_{11}|/|h_{12}| > 2$. Interference is strong, on the other hand, if $|h_{11}|^2 \leq |h_{12}|^2$ and very strong if, additionally,

$$\sigma^2 \geq \frac{|h_{11}|^2}{\frac{|h_{12}|^2}{|h_{11}|^2} - 1} \tag{2.82}$$

as can be seen from (2.60) and (2.61).

We now examine the question of how the capacity region, and, more specifically, the achievable sum rate, behaves over the interference strength $|h_{12}|^2$. For zero interference, i. e., $h_{12} = 0$, the capacity region is square and the achievable sum rate is given by

$$R_1 + R_2 < 2 \log \left(1 + \frac{|h_{11}|^2}{\sigma^2} \right). \tag{2.83}$$

When $|h_{12}|^2 \geq |h_{11}|^2 \left(1 + |h_{11}|^2/\sigma^2\right)$, so that (2.82) is fulfilled, the capacity region is identical to the zero-interference case. For values of $|h_{12}|^2$ inbetween, however, the achievable sum rate is lower, and can in general only be bounded from above and below.

Figures 2.9 and 2.10 show the influence of varying the interference power $|h_{12}|^2$ on the achievable sum rate of a symmetric Gaussian interference channel, for fixed noise powers $\sigma^2 = 0.1$ and $\sigma^2 = 10^{-4}$, respectively. Clearly, forcing the interference to be decodable is of great disadvantage in the noisy interference regime, while treating the interference as noise is strongly suboptimal for channels with strong or very strong interference. Allowing for the transmitters to not use the full power at least permits the case of only one user being active, which achieves half the sum rate at very strong interference. Furthermore, orthogonalizing the users with TDMA outperforms the simplified Han-Kobayashi scheme in some parts of the weak interference regime. It should finally be noted that the "one bit" bound can be up to two bits away from the best achievable scheme in terms of sum rate, as the one bit gap applies to each user's rate.

Interestingly, Figure 2.10 shows that the achievable sum rate at low noise power exhibits two minima and an intermediate peak in the regime of weak interference. This effect will be examined more closely in the following asymptotic analysis.

2.3.5 Asymptotic Results

As is the case with many other Gaussian channel scenarios in information theory, the interference channel lends itself to analysis of the case of asymptotically high signal-to-noise ratio (SNR), which we define here as $\sigma^{-2} \to \infty$.

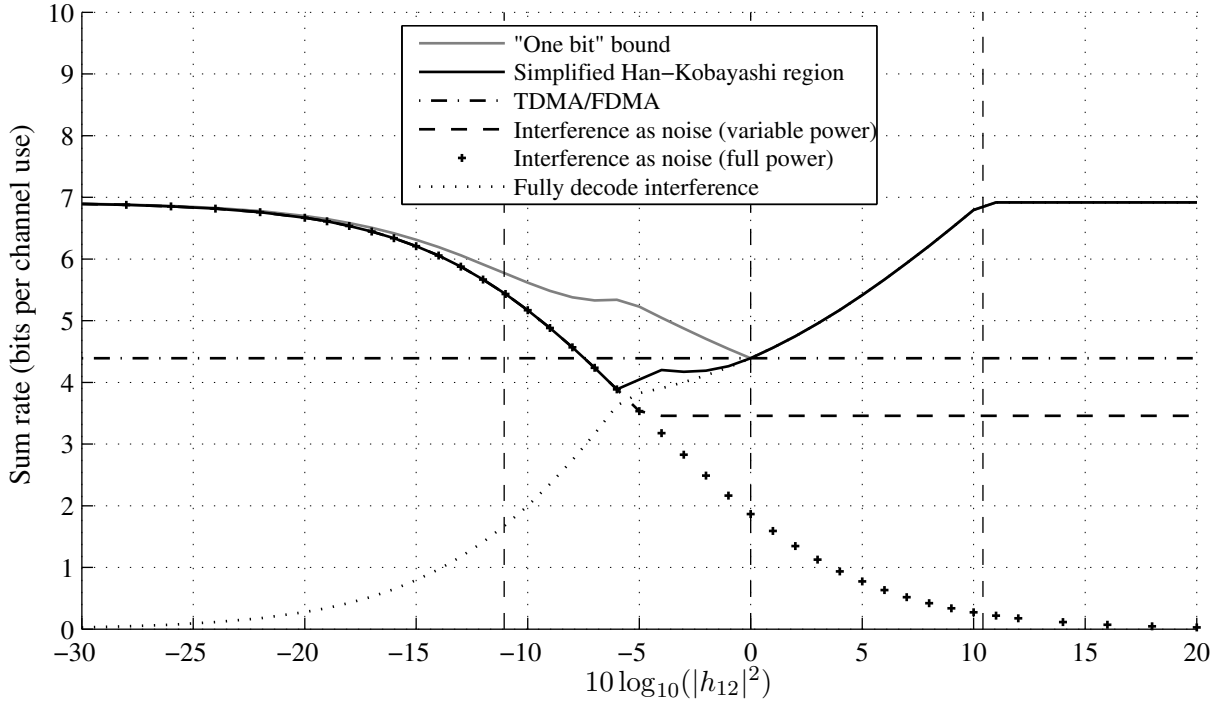


Figure 2.9: Upper and lower bounds on the sum capacity of a symmetric Gaussian interference channel with direct channel coefficients $h_{11} = h_{22} = 1$ and variable cross channel coefficients $h_{12} = h_{21}$. The noise power is $\sigma^2 = 0.1$. The three dashed vertical lines mark the boundaries of the noisy, strong, and very strong interference regimes, respectively.

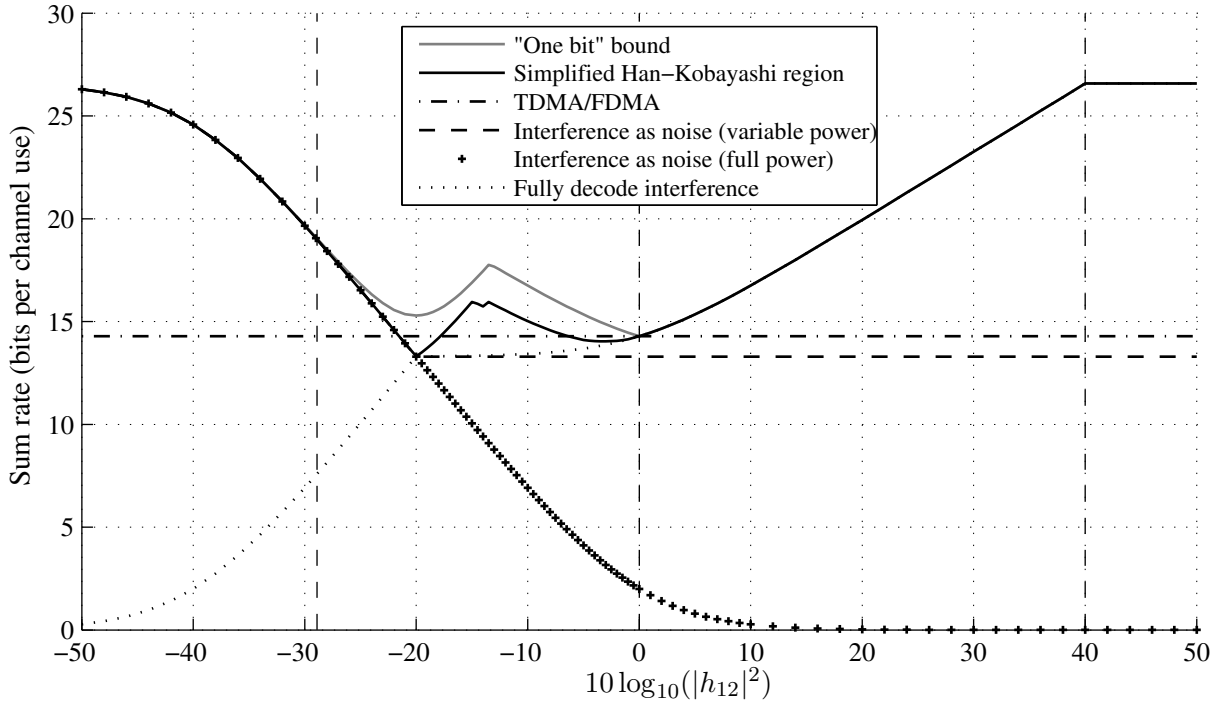


Figure 2.10: Upper and lower bounds on the sum capacity of the symmetric Gaussian interference channel as in Figure 2.9, but with noise power $\sigma^2 = 10^{-4}$.

First, we note that all expressions for capacity and lower or upper bounds on achievable rates encountered in this chapter are $\mathrm{O}\left(\log \sigma^{-2}\right)$ as $\sigma^{-2} \to \infty$, i. e., they grow logarithmically with the SNR. As a simple example, the capacity of the AWGN channel (2.7) approximately becomes

$$C \approx \log \sigma^{-2} + \log|h|^2 \tag{2.84}$$

when σ^{-2} becomes so large that the addition of one inside the logarithm can be neglected. For the MAC, we see that all three expressions (2.10)–(2.12) tend towards $\log \sigma^{-2} + c_i$ with some constants c_i that only depend on h_1 and h_2. When $\log \sigma^{-2}$ is much larger than all of the constants c_i, the MAC region is approximately triangular and is constrained for the most part by the bound on the sum rate (2.12).

In general, we define the sum capacity C_{sum} in any two-user system as the supremum of R_1+R_2 over all achievable rates and the *degrees of freedom* (DoF) D as the growth rate of the sum capacity over $\log \sigma^{-2}$:

$$D = \lim_{\sigma^{-2}\to\infty} \frac{C_{\mathrm{sum}}}{\log \sigma^{-2}}. \tag{2.85}$$

Both the AWGN channel and the Gaussian MAC therefore have one DoF. For a scenario with two parallel AWGN channels, which also can be viewed as a degenerate case of the interference channel with $h_{12} = h_{21} = 0$, we have two DoF, as C_{sum} is the sum of the right-hand sides of the expressions (2.58) and (2.59). The DoF can therefore also be interpreted as an equivalent number of parallel AWGN channels at high SNR.

For the DoF-analysis of non-degenerate interference channels, let us begin with the case of strong interference. The sum capacity in this case is given by the right-hand side of (2.33), and it is clear from (2.85) that $D = 1$. For interference channels with mixed or weak interference, we must again formulate lower and upper bounds on the DoF. If we use, e. g., Carleial's bound as an upper bound and the scheme of fully decoding interference as a lower bound, we see that again the limit is one in both cases. We therefore can state that the Gaussian two-user interference channel has one DoF.

The DoF can be illustrated as the asymptotic slope of the sum capacity when plotted over $\log \sigma^{-2}$. In Figure 2.11, we see that, even though the sum capacity is unknown as we are in a regime of weak interference, the upper and many of the lower bounds achieve the same slope, so that it is clear that the channel has one DoF.

The DoF do not capture the effects of noisy and very strong interference, though, as these regimes require σ^{-2} to be below a finite threshold, cf. (2.79) and (2.60). Therefore, the authors of [17] introduced the concept of *generalized degrees of freedom* (GDoF) for the symmetric case. The key idea of the GDoF analysis is to simultaneously examine the asymptotics of the sum capacity with respect to the SNR and the interference-to-noise ratio (INR). This is achieved by defining an exponential relationship between the SNR and the INR as $\sigma^{-2} \to \infty$, parametrized by the exponent α:

$$\frac{|h_{12}|^2}{\sigma^2} = \left(\frac{|h_{11}|^2}{\sigma^2}\right)^{\alpha} \quad \Leftrightarrow \quad |h_{12}|^2 = \left(|h_{11}|^2\right)^{\alpha} \left(\sigma^{-2}\right)^{\alpha-1}. \tag{2.86}$$

Thus, for $0 \leq \alpha < 1$ the cross channel magnitudes are decreased with decreasing noise power, for $\alpha = 1$ they remain constant (essentially simplifying the GDoF to the conventional DoF), and for $\alpha > 1$ are increased as the noise power decreases. By inserting (2.86) in (2.81) and (2.82), it can be seen that the noisy interference condition is asymptotically fulfilled for $\alpha \leq \frac{1}{3}$ and the very strong interference condition is asymptotically fulfilled for $\alpha \geq 2$.

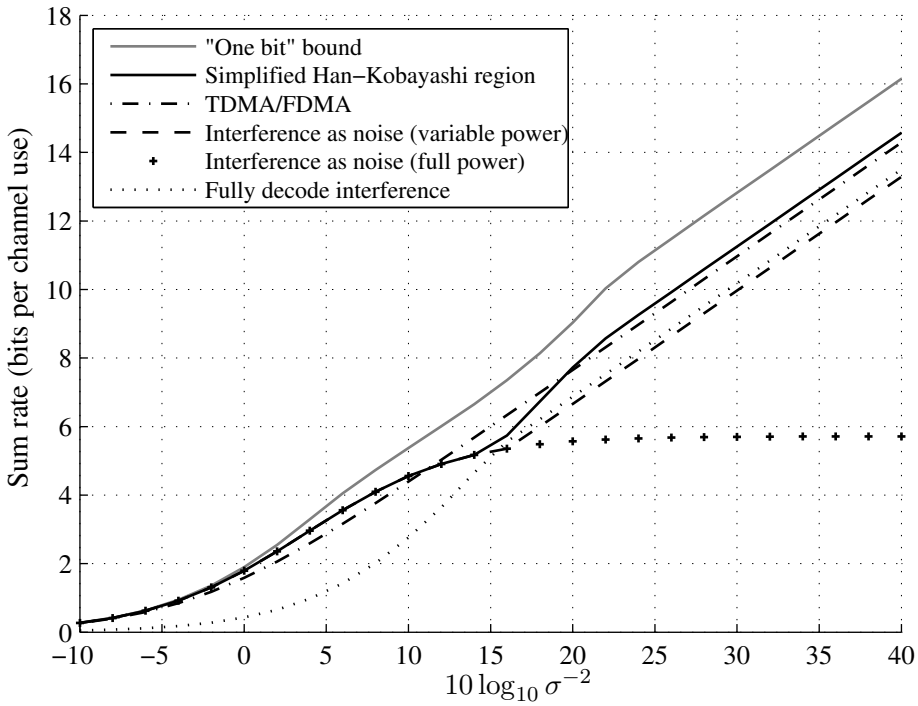


Figure 2.11: Upper and lower bounds on the sum capacity of a Gaussian interference channel with $h_{11} = h_{22} = 1$ and $h_{12} = h_{21} = 0.4$ for varying noise power σ^2.

Similar to the DoF, the GDoF are defined as the growth rate of the sum capacity over $\log \sigma^{-2}$, but here the channel gains, and therefore the sum capacity, depend on the parameter α via (2.86):

$$D_{\mathrm{G}}(\alpha) = \lim_{\sigma^{-2} \to \infty} \frac{C_{\mathrm{sum}}(\alpha)}{\log \sigma^{-2}}. \tag{2.87}$$

Again, using the known lower and upper bounds, the GDoF can be evaluated in a straightforward way, and the resulting expression is (cf. [17])

$$D_{\mathrm{G}}(\alpha) = \begin{cases} 2 - 2\alpha & 0 \leq \alpha < \frac{1}{2} \\ 2\alpha & \frac{1}{2} \leq \alpha < \frac{2}{3} \\ 2 - \alpha & \frac{2}{3} \leq \alpha < 1 \\ \alpha & 1 \leq \alpha < 2 \\ 2 & \alpha \geq 2. \end{cases} \tag{2.88}$$

A plot of the GDoF $D_{\mathrm{G}}(\alpha)$ is shown in Figure 2.12. A comparison with the sum rate bounds at high, but finite, SNR in Figure 2.10 shows that the GDoF capture the qualitative behavior of varying the cross channel magnitudes in symmetric interference channels very well. In particular, the GDoF analysis shows that within the regime of weak interference, i. e., $\alpha < 1$, an increase in interference power can be either beneficial, when $\frac{1}{2} \leq \alpha \leq \frac{2}{3}$, or detrimental, when $\alpha < \frac{1}{2}$ or $\alpha > \frac{2}{3}$. An intuition for this effect can perhaps be gained by means of the Han-Kobayashi strategy: increasing interference power on the one hand means that the common messages can be decoded more easily and thus can carry more useful information, on the other hand it means that the power of the private messages must be decreased, so that they do not create additional non-decodable

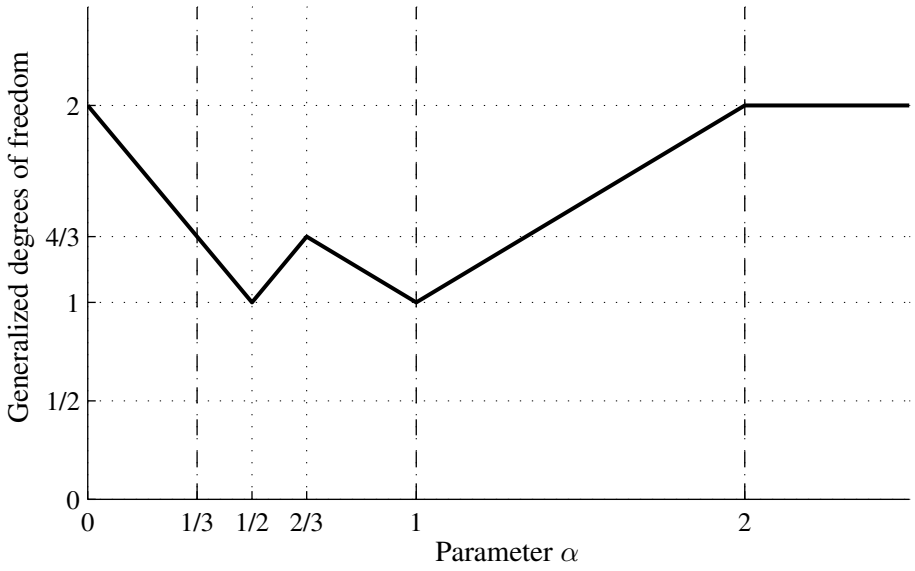


Figure 2.12: Generalized degrees of freedom of the two-user Gaussian interference channel over the interference exponent α. The three dashed vertical lines mark the boundaries of the noisy, strong, and very strong interference regimes, respectively.

noise at the unintended receivers. Which of these two effects is dominant, depends on which part of the weak interference regime we are examining.

We finally note that both DoF and GDoF can be defined on the individual users' rates instead of the sum capacity, leading to the notion of DoF or GDoF regions as the asymptotic shape of the capacity region. Furthermore, it is possible to extend the GDoF to cover general non-symmetric settings. This, however, requires three parameters, instead of the single parameter α, defining exponential relationships between the SNRs and INRs at both receivers, cf. [17].

2.4 The K-User Interference Channel

The Gaussian K-user interference channel is the extension of (2.22)–(2.24) to an arbitrary number of transmitter-receiver pairs K. It is characterized by the equations

$$y_1 = \sum_{k=1}^{K} h_{1k} x_k + n_1 \tag{2.89}$$

$$y_2 = \sum_{k=1}^{K} h_{2k} x_k + n_2 \tag{2.90}$$

$$\vdots$$

$$y_K = \sum_{k=1}^{K} h_{Kk} x_k + n_K \tag{2.91}$$

and the power constraints

$$\mathrm{E}\left[|x_1|^2\right] \leq 1, \quad \mathrm{E}\left[|x_2|^2\right] \leq 1, \quad \ldots \quad \mathrm{E}\left[|x_K|^2\right] \leq 1. \tag{2.92}$$

The noise variables $n_1, n_2, \ldots, n_K$ are complex Gaussian and have variance σ^2. The coefficient $h_{kj} \in \mathbb{C}$ describes the complex channel gain between transmitter j and receiver k. Again, receiver k must decode the message conveyed by x_k from the observation y_k with vanishing error probability. The codebooks of all interfering signals x_j with $j \neq k$ are known to receiver k.

It turns out that the capacity region of the K-user interference channel cannot be sufficiently analyzed by generalizing the achievable schemes and bounding techniques used in the two-user case. In particular, extending the idea of the Han-Kobayashi strategy to more than two users does not yield a close-to-optimal achievable region. While it is possible to formulate conditions for a very strong interference regime, in which the capacity region is identical to that of K parallel AWGN channels with gains $h_{11}, h_{22}, \ldots, h_{KK}$ [21], the general capacity region remains elusive. Significant research, however, has recently been devoted to characterizing the DoF of the K-user interference channel.

The main result is that, for *almost all* sets of channel coefficients h_{kj}, the number of DoF is $\frac{K}{2}$ [22, 23]. This result is remarkable for several reasons. First of all, it implies that adding users to the channel does not decrease the DoF per user, which is somewhat counter-intuitive. Every user always gets one half DoF (or "half the cake" [22]). Also, it shows that orthogonalizing the users via, e. g., TDMA, is not optimal. For both the MAC and the two-user interference channel (as well as many other Gaussian channel scenarios not discussed here, e. g., the multiple-input multiple-output (MIMO) broadcast channel), orthogonalization always is capable of achieving the optimal DoF. In the K-user interference channel, TDMA would yield one DoF, which is strictly less than $\frac{K}{2}$ for $K \geq 3$.

Furthermore, the achievability techniques leading to this result show that, contrary to previously discussed scenarios, random Gaussian codebooks are not DoF-optimal for $K \geq 3$. Instead, the notion of *interference alignment* is crucial. The transmit strategies and codebooks must be designed in a way that all interfering signals somehow occupy overlapping subspaces at each of the receivers. Possible techniques for creating such subspaces in which the interference can be aligned include lattice codes [21, 24], assuming non-causally known time-variations in the channel coefficients [22], or "real interference alignment" based on the properties of rational and irrational numbers [25, 23].

For the special case of the symmetric K-user IC, in which all direct channel coefficients are identical, i. e., $h_{11} = h_{22} = \ldots = h_{KK}$, and all cross channels h_{kj} with $k \neq j$ are identical, the GDoF have also been analyzed [26]. Here, the relationship between the direct and the cross channel gains is defined as

$$h_{12} = h_{11}^{\alpha} \sigma^{1-\alpha} \tag{2.93}$$

since the complex phase is significant, as opposed to the two-user case, cf. (2.86). The resulting expression for the GDoF is simply $\frac{K}{2} \cdot D_{\mathrm{G}}(\alpha)$, i. e., the GDoF from the two-user case, cf. (2.88), multiplied with $K/2$, with the exception of a singularity at $\alpha = 1$, where the GDoF of the K-user interference channel are one.

2.5 Results for Multi-Antenna Scenarios

In this section we examine Gaussian interference channels with vectors as channel input and output variables, corresponding to systems in which the transmitters and receivers have multiple antennas. Each entry of the input and output vectors is the symbol transmitted or received at one of the antennas. The power constraints are usually formulated as an upper limit on the sum of the powers of

the symbols transmitted at all antennas of one transmitter. The noise is assumed to be independent at all receive antennas with the same variance, which is not a loss of generality, as arbitrary noise covariance matrices can be covered by defining equivalent "whitened" channels, cf. Chapter 5.

2.5.1 Two Users

The two-user Gaussian multiple-input multiple-output (MIMO) interference channel with N_k antennas at transmitter k and M_k antennas at receiver k has input vectors $\boldsymbol{x}_k \in \mathbb{C}^{N_k}$ and output vectors $\boldsymbol{y}_k \in \mathbb{C}^{M_k}$, and is defined by the equations

$$\boldsymbol{y}_1 = \boldsymbol{H}_{11}\boldsymbol{x}_1 + \boldsymbol{H}_{12}\boldsymbol{x}_2 + \boldsymbol{n}_1 \tag{2.94}$$

$$\boldsymbol{y}_2 = \boldsymbol{H}_{22}\boldsymbol{x}_2 + \boldsymbol{H}_{21}\boldsymbol{x}_1 + \boldsymbol{n}_2 \tag{2.95}$$

and the power constraints

$$\mathrm{E}\left[\|\boldsymbol{x}_k\|_2^2\right] = \mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{x}_k\boldsymbol{x}_k^{\mathrm{H}}\right]\right) = \mathrm{tr}(\boldsymbol{Q}_k) \leq 1, \qquad \forall k \in \{1,2\} \tag{2.96}$$

where $\boldsymbol{Q}_1 \in \mathbb{C}^{N_1 \times N_1}$ and $\boldsymbol{Q}_2 \in \mathbb{C}^{N_2 \times N_2}$ are the input correlation matrices (or covariance matrices, assuming zero-mean input vectors). The channel matrices $\boldsymbol{H}_{kj} \in \mathbb{C}^{M_k \times N_j}$ contain the complex channel gains between all N_j antennas of transmitter j and all M_k antennas at receiver k. The noise vectors $\boldsymbol{n}_k \in \mathbb{C}^{M_k}$ are assumed to be zero-mean complex Gaussian with the covariance matrices

$$\mathrm{E}\left[\boldsymbol{n}_1\boldsymbol{n}_1^{\mathrm{H}}\right] = \mathrm{E}\left[\boldsymbol{n}_2\boldsymbol{n}_2^{\mathrm{H}}\right] = \sigma^2\mathbf{I}. \tag{2.97}$$

Many of the results from the scalar two-user interference channel can be generalized to the MIMO case with some effort. An achievable strategy based on the Han-Kobayashi scheme was presented in [27]. In [28], the "one bit" bound was formulated for a more general class of interference channels, which includes the multi-antenna case. The gap between the outer bound and the Han-Kobayashi region is shown to be one bit per receive antenna, i. e., for every rate pair (R_1, R_2) on the bound, the rate pair $(R_1 - M_1 \log 2, R_2 - M_2 \log 2)$ is proven to be achievable by a Han-Kobayashi strategy. Conditions for a strong interference regime where it is optimal to completely decode the interference are given in [29] for the case of either $N_1 = N_2 = 1$ or $M_1 = M_2 = 1$. Conditions for noisy, mixed, strong, and very strong interference regimes for general antenna configurations were proposed in [30, 31]. The resulting expressions for the sum capacity and capacity regions of the MIMO interference channel are not easily given in closed form, however, but mostly involve an optimization over all possible input covariance matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$.

The DoF of the two-user MIMO interference channel were analyzed in [32]. If all four channel matrices have full rank, the number of DoF is

$$\min\{M_1 + M_2, N_1 + N_2, \max\{M_1, N_2\}, \max\{M_2, N_1\}\} \tag{2.98}$$

and can be achieved by a scheme of transforming the channel into a number of parallel non-interfering scalar AWGN channels by means of linear pre- and post-processing of the input and output vectors, i. e., by spatial orthogonalization.

2.5.2 $K > 2$ Users

In the K-user Gaussian MIMO interference channel, the channel output of user k, for $k \in \{1, \ldots, K\}$, is defined as

$$\boldsymbol{y}_k = \sum_{j=1}^{K} \boldsymbol{H}_{kj}\boldsymbol{x}_j + \boldsymbol{n}_k \tag{2.99}$$

where $\boldsymbol{y}_k \in \mathbb{C}^{M_k}$, $\boldsymbol{n}_k \in \mathbb{C}^{M_k}$, and $\boldsymbol{x}_j \in \mathbb{C}^{N_j}$. The channel coefficients are contained in the matrices $\boldsymbol{H}_{kj} \in \mathbb{C}^{M_k \times N_j}$ for $(j,k) \in \{1,\ldots,K\}^2$, the transmit power constraints and noise covariances for all users are defined as in (2.96) and (2.97).

As for the K-user scalar interference channel, the capacity region and sum capacity of the K-user MIMO interference channel remain largely elusive. Some results have been formulated, however, concerning the DoF. In [33], it is shown when all transmitters have N antennas and all receivers have M antennas, the following number of DoF is achievable for *almost all* sets of channel matrices:

$$\frac{MN}{M+N}K. \tag{2.100}$$

Again, in order to achieve this number of DoF, random Gaussian codebooks are not sufficient and interference alignment is necessary. Also, the number of DoF is upper-bounded by (2.100), if additionally

$$K \geq \frac{M+N}{\gcd\{M,N\}} \tag{2.101}$$

is fulfilled, where $\gcd\{M,N\}$ is the greatest common divisor of M and N. Consequently, if $M = N$, the K-user MIMO interference channel almost always has exactly $KN/2$ DoF.

An interesting suboptimal technique for K-user MIMO interference channels is discussed in [34], where the possibilities of spatial orthogonalization are explored. Specifically, the goal is to design the transmit covariance matrices $\boldsymbol{Q}_k$ in such a way that at every receiver k the subspace spanned by the possible outcomes of the desired component $\boldsymbol{H}_{kk}\boldsymbol{x}_k$ is linearly independent of the subspace spanned by the possible outcomes of the interference component $\sum_{j\neq k} \boldsymbol{H}_{kj}\boldsymbol{x}_j$. Thus, desired signal and interference are orthogonalizable by a linear projection at the receiver, and specialized coding schemes are not necessary.

This technique of "spatial interference alignment" achieves the optimal $KN/2$ DoF for $K = 3$ and even-numbered N, but for more than three users is generally suboptimal in terms of DoF. It does, however, fall in the scope of the linear strategies discussed subsequently in this work and is therefore treated in detail in Chapter 5.

2.6 Linear Strategies in the Context of Information Theoretic Limits

In this chapter we gave an overview over what is currently known about the theoretical throughput limits of interference networks. Clearly, even in those cases where the limits are known, achieving them in a practical system is very challenging. It was always assumed, for instance, that some centralized entity is able to choose the best transmit strategy making use of perfect knowledge of all channel coefficients, which might not be easily accomplished in a system in which the users are only able to loosely cooperate. The transmit strategies (in particular the codebooks) of all transmitters must in turn also be made available to all receivers, so that they are able to subtract the decodable portions of the interference. This implies a considerable signaling overhead, especially when the channel conditions vary quickly, as is usually the case in wireless environments. In the following we investigate a particular class of strategies that is information theoretically suboptimal in most cases, but which alleviates some of these practical issues and significantly simplifies the process of channel coding and decoding.

These *linear* strategies are characterized by the following properties of the transmitters and receivers: the transmitted vectors $\boldsymbol{x}_k$ are obtained by applying a linear transformation, i. e., a matrix

multiplication, to a vector of *data symbols* $\boldsymbol{s}_k \in \mathbb{C}^{d_k}$, of which all entries are statistically independent and have unit variance. The statistical independence of the elements of the data symbol vector corresponds to d_k data streams being channel coded separately. The receiver, on the other hand, forms an estimate $\hat{\boldsymbol{s}}_k$ of the data symbol vector by applying a linear transformation (matrix multiplication) to the received vector $\boldsymbol{y}_k$. The individual data streams are then decoded from the elements of the vector $\hat{\boldsymbol{s}}_k$ separately, i. e., without knowledge of any of the codebooks of the interfering signals or of the other streams from the desired transmitter.

For the single-antenna interference channel, in which the channel inputs and outputs are scalar, the linear operation is simply the multiplication with a scalar. The linear receivers do not perform any sort of successive or joint decoding of parts of the interfering signals. The only linear strategy among the achievable schemes discussed in Section 2.3.1 is therefore that of treating interference as noise with variable power, cf. Section 2.3.1.1. Here, the linear operation at transmitter k is the multiplication of the unit-variance channel coded signal s_k with $\sqrt{p_k}$.

Even though the linear strategies do not achieve the full potential of general interference networks, their use is appealing for a number of reasons:

- The codebooks must not be known at the unintended receivers, thus greatly reducing the signaling overhead.
- The receivers are not required to perform joint or successive decoding of multiple messages, which can be computationally very demanding and/or add considerable decoding delay.
- Channel coding at the transmitters is straightforward, as the individual streams are encoded as if they were intended for transmission over a scalar channel, in contrast to the DoF-optimal alignment techniques, which require more complex coding schemes.
- The linear strategies achieve the optimal number of DoF for all two-user networks and for three-user networks with an even number of antennas at each terminal.
- They are even sum-rate optimal in situations in which the interference is weak compared to the desired signal (and the noise), which is not uncommon in practical systems. In cellular systems, for instance, the receivers are generally associated with a transmitter in close proximity, while the interferers are in neighboring cells, and therefore have lower channel gains.

Finally, we note that in the following we restrict ourselves to transmission strategies with a single data stream per user, i. e., where s_k is scalar. In the systems examined in Chapters 3 and 4, where the receivers only have a single antenna, this is fully sufficient, but in the MIMO case (cf. Chapter 5), this again is only a subset of all possible linear strategies, albeit one that already provides many interesting challenges.

3. Single-Antenna Interference Networks

The problem of finding the best linear strategy in single-input single-output (SISO) scenarios is a *power control* problem. Extensive research has been devoted to characterizing and efficiently achieving different kinds of optimality. A very detailed discussion can be found, e. g., in [35].

In this chapter we provide an overview of the results for the power control problem in SISO interference networks, again without claiming completeness. The purpose is to introduce the concepts of signal-to-interference-plus-noise ratio (SINR), utility functions, and different notions of optimality as well as the difficulties in achieving optimality. Many of these concepts are easily extended to multi-antenna interference networks, which are the focus of this work, and understanding the results for the SISO case is beneficial for the following chapters.

In the second part of this chapter, starting with Section 3.5, we discuss the aspect of *distributed* or *decentralized* optimization and introduce the distributed interference pricing algorithm, as it was proposed for the single-antenna case in [2]. A generalization of this algorithm will play an important role in the multi-antenna systems in the following chapters.

3.1 System Model

The system setup is identical to that of the information theoretic Gaussian K-user SISO interference channel discussed in Section 2.4: the K transmitter-receiver pairs (or *users*) are connected with scalar complex channel gains which model the frequency flat wireless channel. The channel coefficient between transmitter j and receiver k is $h_{kj} \in \mathbb{C}$ and is assumed to remain constant over time. We additionally assume that all channels between the transmitters and their intended receivers are non-zero, i. e., $h_{kk} \neq 0$ for all $k \in \{1, \ldots, K\}$.

The symbol transmitted by user k at a given time instant is the realization of the random variable x_k. The additive noise at receiver k is the realization of the random variable n_k. Thus, the received symbol of user k is the random variable

$$y_k = h_{kk}x_k + \sum_{\substack{j=1 \\ j\neq k}}^{K} h_{kj}x_j + n_k \qquad \forall k \in \{1, \ldots, K\} \tag{3.1}$$

where the first summand is the desired part, the second summand is the interference, and the third summand is the noise. The noise variable n_k is assumed to be identically distributed at each receiver k, as a zero-mean complex Gaussian variable with variance

$$\mathrm{E}\left[|n_k|^2\right] = \sigma^2 \qquad \forall k \in \{1, \ldots, K\} \tag{3.2}$$

where $\sigma^2 > 0$. We define the power p_k of the transmitted signal of user k to be the second moment of x_k, i. e.,

$$p_k = \mathrm{E}\left[|x_k|^2\right] \qquad \forall k \in \{1, \ldots, K\} \tag{3.3}$$

which is identical to the variance of x_k when $\mathrm{E}[x_k] = 0$, as will be assumed henceforth. Furthermore, note that $\mathrm{E}[x_k^* x_j] = 0$ for all $j \neq k$ since the transmitters are not able to cooperate. We

impose a unit power constraint and consequently only allow power allocations that fulfill

$$0 \leq p_k \leq 1 \qquad \forall k \in \{1, \ldots, K\}. \tag{3.4}$$

As discussed in the definition of the linear strategies at the end of the previous chapter, the receivers do not distinguish between interference and noise. A sensible quality metric for the communications link between transmitter k and receiver k therefore is the SINR γ_k:[1]

$$\gamma_k = \frac{|h_{kk}|^2 p_k}{\sum\limits_{j \neq k} |h_{kj}|^2 p_j + \sigma^2} \qquad \forall k \in \{1, \ldots, K\}. \tag{3.5}$$

We furthermore assign a utility function $u_k(\gamma_k)$ to each user k that quantifies the desirability of a certain SINR. A utility function in our definition must be strictly increasing, i. e., a higher SINR is always more desirable for a user than a lower SINR. Some common utility functions $u_k(\gamma_k)$ as well as their motivation will be discussed in Section 3.3.

It should be noted that the fact that the noise powers at all receivers are identical and all transmit powers are constrained in the same way does not present a loss of generality. A system with different noise and transmit powers can be easily transformed to meet the specifications with a normalization of the channel coefficients. To show this, let us consider a system in which the power of user k is constrained by P_k instead of one and the noise power at receiver k is σ_k^2 instead of σ^2; the transmit symbols in this system are denoted by x'_k, the noise variables by n'_k, and the channel coefficients by h'_{kj}. If we now define a second system with $x_k = \frac{1}{\sqrt{P_k}} x'_k$, $n_k = \frac{\sigma}{\sigma_k} n'_k$, and $h_{kj} = \frac{\sigma}{\sigma_k} \sqrt{P_j} h'_{kj}$, it is trivial to show that the specifications of equal noise power and unit transmit power constraints are met while the SINRs in both systems are identical.

3.2 The Utility Region

We define a *region* to be a set of K-tuples. The following three regions are relevant to the systems we are investigating:

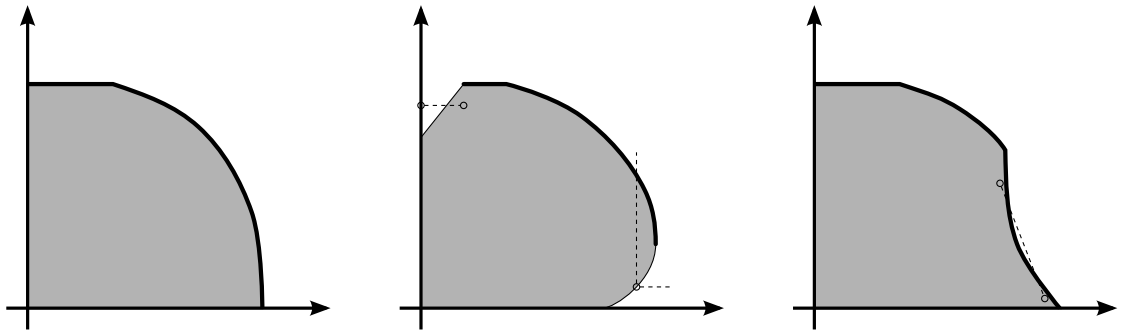
- The *power region,* which consists of all power K-tuples $(p_1, \ldots, p_K)$ that fulfill (3.4). We refer to these K-tuples as *feasible power allocations.*
- The *SINR region,* which consists of all SINR K-tuples $(\gamma_1, \ldots, \gamma_K)$ that result from a feasible power allocation via (3.5).
- The *utility region,* which consists of all utility K-tuples $(u_1(\gamma_1), \ldots, u_K(\gamma_K))$ that result from a feasible SINR K-tuple.

In the following we discuss some fundamental properties of these regions.

Definition 3.1. A region is *comprehensive,* if for any K-tuple $(r_1, \ldots, r_K)$ in the region all K-tuples $(\bar{r}_1, \ldots, \bar{r}_K)$ with $0 \leq \bar{r}_k \leq r_k$ for all $k \in \{1, \ldots, K\}$ are also in the region.

Definition 3.2. A region is *convex,* if for any two K-tuples in the region all convex combinations are also in the region. A convex combination of the two K-tuples $(\check{r}_1, \ldots, \check{r}_K)$ and $(\hat{r}_1, \ldots, \hat{r}_K)$ is any K-tuple $(\alpha\check{r}_1 + (1-\alpha)\hat{r}_1, \ldots, \alpha\check{r}_K + (1-\alpha)\hat{r}_K)$ with $\alpha \in [0, 1]$.

[1] The SINR metric does not capture all aspects of the quality of the communications link. For example, the shape of the PDFs of the desired portion of the received signal and the interference could be exploited to improve detection of the data symbols. For random Gaussian codebooks, however, the SINR is a sufficient description of the quality of a communications link.



(a) Comprehensive and convex

(b) Not comprehensive, but convex

(c) Comprehensive, but non-convex

Figure 3.1: Example regions that illustrate the comprehensiveness and convexity properties. In the middle example, it is possible to move towards an axis from a point in the region without remaining inside the region. In the right example, a line segment connecting two points is not entirely within the region. Boundary points that are strictly dominated by other points of the region and are therefore not part of the Pareto boundary are marked by a thinner line; the boundary points on the axes are also not part of the Pareto boundary.

Definition 3.3. The *Pareto boundary* of a region is the subset of K-tuples in the region that are not strictly dominated by any K-tuple in the region. A K-tuple $(r_1, \ldots, r_K)$ is said to strictly dominate another K-tuple $(\bar{r}_1, \ldots, \bar{r}_K)$ if $r_k > \bar{r}_k$ for all $k \in \{1, \ldots, K\}$.

Figure 3.1 illustrates comprehensiveness and convexity as well as the Pareto boundary for some example regions with $K = 2$. Comprehensiveness refers to the property that reducing one or more elements of a K-tuple in the region while remaining inside the positive quadrant will not lead outside the region. For a region to be convex, every line segment connecting two points of the region must lie entirely within the region. The Pareto boundary can be thought of as the top right boundary of the region. These concepts extend to higher dimensions, i. e., $K > 2$, in a straightforward way.

The power region consists of all power allocations $(p_1, \ldots, p_K)$ with $p_k \in [0, 1]$ for all $k \in \{1, \ldots, K\}$. It can be viewed as a unit hypercube in K-dimensional space. The power region is clearly both comprehensive and convex; its Pareto boundary consists of all power allocations in which at least one power p_k is equal to one.

The region of feasible power allocations is mapped to the region of achievable SINRs via (3.5). We observe that the following properties hold for the relationship between powers and SINR:

1) The SINR γ_k is increasing in p_k and non-increasing in p_j for all $j \neq k$.
2) If all powers are increased by a common factor, all SINRs are increased. Decreasing all powers by a common factor leads to a decrease of all SINRs.
3) Any change in power allocation that includes decreasing at least one power will lead to a decrease in at least one SINR. This can be seen from the fact that such a change in power allocation can always be achieved by first performing a decrease of all powers by a common factor and then increasing some but not all powers. The property then follows from the previous properties 2) and 1).

Proposition 3.1. *Every power allocation on the Pareto boundary of the power region is mapped to an SINR K-tuple on the Pareto boundary of the SINR region; conversely, every SINR K-tuple*

on the Pareto boundary of the SINR region can be achieved by a power allocation on the Pareto boundary of the power region.

Proof. For every power allocation on the Pareto boundary of the power region, at least one user has full power. We will assume that user k has full power, i. e., $p_k = 1$. A change of power allocation that does not violate the power constraints will either
1) include a decrease of one or more powers or
2) consist of only increases of one or more powers, but leave p_k unchanged.
In the first case at least one SINR will decrease due to the above property 3); in the second case γ_k cannot increase due to property 1). Therefore the resulting SINR K-tuple cannot strictly dominate the original SINR K-tuple, i. e., any SINR K-tuple resulting from a power allocation with at least one full power is on the Pareto boundary of the SINR region.

A power allocation not on the Pareto boundary of the power region, on the other hand, leads to an SINR K-tuple not on the Pareto boundary of the SINR region: if all of the powers are strictly smaller than one, it is possible to increase all powers by a common factor, which yields a strictly dominant SINR K-tuple due to 2) in the above properties. Therefore, an SINR K-tuple on the Pareto boundary of the SINR region cannot be achievable by a power allocation which is not on the Pareto boundary of the power region, but by definition must be achievable by some power allocation, from which it follows that the converse also holds. □

Proposition 3.2. *The SINR region is comprehensive.*

Proof. The proof is given in Appendix A2. □

Proposition 3.3. *The SINR region is non-convex, unless all cross channel coefficients h_{kj} with $j \neq k$ are zero.*

Proof. The proof is given in Appendix A3. □

The utility region consists of all feasible K-tuples $(u_1(\gamma_1), \ldots, u_K(\gamma_K))$. In contrast to the mapping from powers to SINRs, the elements of the SINR K-tuple are mapped separately to the elements of the utility K-tuple, i. e., u_k does not depend on γ_j with $j \neq k$.
Proposition 3.4. *The Pareto boundary of the power region is mapped to the Pareto boundary of the utility region.*

Proof. The proposition is an immediate consequence of the fact that the utility functions $u_k(\gamma_k)$ are strictly increasing. Therefore strict dominance in SINRs implies strict dominance in utilities and vice versa and Proposition 3.1 holds for utilities instead of SINRs as well. □

Proposition 3.5. *The utility region is comprehensive.*

Proof. Again, the proposition follows directly from the strict monotonicity of the utility functions and Proposition 3.2. □

3.2.1 Convexity of the Utility Region

The utility region is in general non-convex. As an example, consider the identity utility $u_k(\gamma_k) = \gamma_k$, where the utility region is the same as the SINR region and is therefore non-convex unless all cross channels are zero, cf. Proposition 3.3. It is, however, possible to formulate a sufficient condition for convexity of the utility region and thereby identify a special class of utility functions that guarantee a convex region. In the following we derive this condition along the lines of [35], where it is proven for a more general system model. We also note that a similar, but stricter, condition is derived in [2]. For the remainder of this section, we consider only utility functions that are twice differentiable.

Definition 3.4. A function $f(\boldsymbol{x})$ is *convex* in $\boldsymbol{x}$, if for any two vectors $\check{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}$ in the domain of f and any $\lambda \in [0, 1]$ the following holds:

$$f\left(\lambda\check{\boldsymbol{x}} + (1-\lambda)\hat{\boldsymbol{x}}\right) \leq \lambda f(\check{\boldsymbol{x}}) + (1-\lambda) f(\hat{\boldsymbol{x}}). \tag{3.6}$$

The function $f(\boldsymbol{x})$ is *concave* in $\boldsymbol{x}$ if $-f(\boldsymbol{x})$ is convex in $\boldsymbol{x}$, i. e., if "$\geq$" holds instead of "$\leq$" in above inequality.

An in-depth study of convex and concave functions as well as their properties can be found in [36, Chapter 3]. A notable result is that for a scalar, twice differentiable function convexity (or concavity) is equivalent to a non-negative (or non-positive) second derivative.

To begin with, we investigate the properties of the logarithmic SINR in the logarithmic powers. We define the logarithmic powers as

$$t_k = \log p_k \quad \Leftrightarrow \quad p_k = \mathrm{e}^{t_k} \qquad k \in \{1, \ldots, K\} \tag{3.7}$$

and the vector $\boldsymbol{t} = [t_1, \ldots, t_K]^{\mathrm{T}}$. By inserting (3.7) into (3.5) and taking the logarithm, we obtain the logarithmic SINR

$$w_k(\boldsymbol{t}) = \log \gamma_k = \log|h_{kk}|^2 + t_k - \log\left(\sum_{j\neq k} |h_{kj}|^2 \,\mathrm{e}^{t_j} + \sigma^2\right) \qquad k \in \{1, \ldots, K\}. \tag{3.8}$$

Lemma 3.1. *$w_k(\boldsymbol{t})$ is concave in $\boldsymbol{t}$.*

Proof. For any $\lambda \in (0, 1)$ and non-negative real a_i and b_i with $i \in \{1, \ldots, I\}$, the following holds:

$$\sum_{i=1}^{I} a_i^{\lambda} b_i^{1-\lambda} \leq \left(\sum_{i=1}^{I} a_i\right)^{\lambda} \left(\sum_{i=1}^{I} b_i\right)^{1-\lambda}. \tag{3.9}$$

This can be shown by defining $\boldsymbol{a} = \left[a_1^{\lambda}, \ldots, a_I^{\lambda}\right]^{\mathrm{T}}$ and $\boldsymbol{b} = \left[b_1^{1-\lambda}, \ldots, b_I^{1-\lambda}\right]^{\mathrm{T}}$ and applying Hölder's inequality (e. g., [37]):

$$\left|\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}\right| \leq \|\boldsymbol{a}\|_{\frac{1}{\lambda}} \|\boldsymbol{b}\|_{\frac{1}{1-\lambda}}. \tag{3.10}$$

Therefore, for any $\lambda \in (0,1)$

$$\begin{aligned}
&\lambda w_k\left(\check{\boldsymbol{t}}\right) + (1-\lambda) w_k\left(\hat{\boldsymbol{t}}\right) = \\
&= \log|h_{kk}|^2 + \lambda\check{t}_k + (1-\lambda)\hat{t}_k - \lambda\log\left(\sum_{j\neq k}|h_{kj}|^2\,\mathrm{e}^{\check{t}_j} + \sigma^2\right) - (1-\lambda)\log\left(\sum_{j\neq k}|h_{kj}|^2\,\mathrm{e}^{\hat{t}_j} + \sigma^2\right) \\
&= \log|h_{kk}|^2 + \lambda\check{t}_k + (1-\lambda)\hat{t}_k - \log\left(\sum_{j\neq k}|h_{kj}|^2\,\mathrm{e}^{\check{t}_j} + \sigma^2\right)^{\lambda}\left(\sum_{j\neq k}|h_{kj}|^2\,\mathrm{e}^{\hat{t}_j} + \sigma^2\right)^{1-\lambda} \qquad (3.11) \\
&\leq \log|h_{kk}|^2 + \lambda\check{t}_k + (1-\lambda)\hat{t}_k - \log\left(\sum_{j\neq k}|h_{kj}|^2\,\mathrm{e}^{\lambda\check{t}_j + (1-\lambda)\hat{t}_j} + \sigma^2\right) \qquad (3.12) \\
&= w_k\left(\lambda\check{\boldsymbol{t}} + (1-\lambda)\hat{\boldsymbol{t}}\right) \qquad (3.13)
\end{aligned}$$

and the condition for concavity is fulfilled. For the two remaining possibilities $\lambda = 0$ and $\lambda = 1$ it is clear that the condition is also fulfilled. □

Next, we express the utility by means of the logarithmic SINR and logarithmic powers:

$$u_k(\gamma_k) = u_k\left(\mathrm{e}^{w_k(\boldsymbol{t})}\right) \qquad k \in \{1,\ldots,K\}. \tag{3.14}$$

Lemma 3.2. *If*

$$c_k(\gamma_k) = -\frac{u_k''(\gamma_k)\gamma_k}{u_k'(\gamma_k)} \geq 1 \qquad \textit{with} \qquad u_k'(\gamma_k) = \frac{\partial u_k(\gamma_k)}{\partial\gamma_k} \qquad \textit{and} \qquad u_k''(\gamma_k) = \frac{\partial^2 u_k(\gamma_k)}{(\partial\gamma_k)^2} \tag{3.15}$$

holds for all $\gamma_k > 0$, then $u_k\left(\mathrm{e}^{w_k(\boldsymbol{t})}\right)$ is concave in $\boldsymbol{t}$.

Proof. $u_k\left(\mathrm{e}^x\right)$ is increasing in $\mathrm{e}^x = \gamma_k$ and therefore also increasing in x. Furthermore, with the chain rule the second derivative of $u_k\left(\mathrm{e}^x\right)$ w. r. t. x is:

$$\frac{\partial^2 u_k\left(\mathrm{e}^x\right)}{(\partial x)^2} = u_k''\left(\mathrm{e}^x\right)\mathrm{e}^{2x} + u_k'\left(\mathrm{e}^x\right)\mathrm{e}^x = u_k''(\gamma_k)\gamma_k^2 + u_k'(\gamma_k)\gamma_k. \tag{3.16}$$

Since a non-positive second derivative implies concavity, $u_k\left(\mathrm{e}^x\right)$ is concave in x for $c_k(\gamma_k) \geq 1$. If $u_k\left(\mathrm{e}^x\right)$ is concave in x, $u_k\left(\mathrm{e}^{w_k(\boldsymbol{t})}\right)$ is concave in $\boldsymbol{t}$ since $w_k(\boldsymbol{t})$ is concave in $\boldsymbol{t}$ and the composition of a concave increasing function with a concave function is again concave [36, Section 3.2.4]. □

Note that $c_k(\gamma_k)$ is a quantity used in economics to describe the *relative risk aversion* of a payoff function (e. g., [38]).

Theorem 3.1. *If $c_k(\gamma_k) \geq 1$ holds for all $\gamma_k > 0$ and for all $k \in \{1,\ldots,K\}$, the utility region is convex.*

Proof. Consider two feasible power allocations $(\check{p}_1,\ldots,\check{p}_K)$ and $(\hat{p}_1,\ldots,\hat{p}_K)$. The corresponding vectors of logarithmic power are $\check{\boldsymbol{t}}$ and $\hat{\boldsymbol{t}}$, and the resulting utility K-tuples are $(\check{u}_1,\ldots,\check{u}_K)$ and $(\hat{u}_1,\ldots,\hat{u}_K)$, respectively. Any non-trivial convex combination of the two utility K-tuples can

be expressed as $(\lambda\check{u}_1 + (1-\lambda)\hat{u}_1, \ldots, \lambda\check{u}_K + (1-\lambda)\hat{u}_K)$ with $\lambda \in (0,1)$. From Lemma 3.2 we know that

$$\lambda\check{u}_k + (1-\lambda)\hat{u}_k \leq u_k\left(\mathrm{e}^{w_k(\lambda\check{\boldsymbol{t}}+(1-\lambda)\hat{\boldsymbol{t}})}\right) \qquad \forall k \in \{1,\ldots,K\}. \tag{3.17}$$

Furthermore, we observe that the logarithmic power vector $\lambda\check{\boldsymbol{t}}+(1-\lambda)\hat{\boldsymbol{t}}$ corresponds to the power allocation $\left(\check{p}_1^{\lambda}\hat{p}_1^{1-\lambda}, \ldots, \check{p}_K^{\lambda}\hat{p}_K^{1-\lambda}\right)$, which is clearly feasible. Therefore, the convex combination of utilities $(\lambda\check{u}_1 + (1-\lambda)\hat{u}_1, \ldots, \lambda\check{u}_K + (1-\lambda)\hat{u}_K)$ is element-wise smaller or equal to the feasible utility K-tuple resulting from the power allocation $\left(\check{p}_1^{\lambda}\hat{p}_1^{1-\lambda}, \ldots, \check{p}_K^{\lambda}\hat{p}_K^{1-\lambda}\right)$. Due to the comprehensiveness of the utility and SINR region (cf. Propositions 3.2 and 3.5) it is clear that the convex combination must be feasible. Therefore the utility region is convex. □

Observation 3.1. The condition $c_k(\gamma_k) \geq 1$ can only be fulfilled for utility functions $u_k(\gamma_k)$ that are concave in γ_k since $u_k'(\gamma_k) > 0$ and $\gamma_k > 0$.
Observation 3.2. Another consequence of the condition is that $u_k(0) = -\infty$ since, as is shown in the proof of Lemma 3.2, the condition implies that $u_k\left(\mathrm{e}^x\right)$ is concave and increasing for $x \in \mathbb{R}$ and therefore must be unbounded for $x \to -\infty$.

3.3 Common Utility Functions

3.3.1 Achievable Rate Utility

$$u_k(\gamma_k) = \log\left(1+\gamma_k\right) = R_k. \tag{3.18}$$

If all interfering users employ random Gaussian codebooks and these codebooks are unknown to receiver k, i. e., the interference is equivalent to an additional Gaussian noise source, the link of user k can be viewed as an AWGN channel and the discussion on achievable code rates in Section 2.1 applies. Therefore, this utility describes the supremum of the code rates at which information can be reliably transmitted on the link of user k, cf. (2.7).

As $u_k(0) = 0$, the achievable rate utility does not belong to the class of utilities that guarantee a convex region (cf. Observation 3.2). This can be verified by taking the first and second derivatives of $u_k(\gamma_k)$ to obtain

$$c_k(\gamma_k) = \frac{\gamma_k}{1+\gamma_k} < 1. \tag{3.19}$$

Indeed, since for the two-user case the utility region coincides with the rate region achievable with the information theoretic scheme of treating interference as noise with variable power discussed in Section 2.3.1.1, examples for the shape of this utility region can be seen in Figures 2.3–2.8. Clearly, the utility region, in this case referred to as the *rate region,* can be non-convex.

We also refer to the achievable rate utility for user k as R_k. Because it is a measure for the link throughput, the rate utility is the focus of our algorithm design and numerical results, where we will generally use the logarithm with base 2 in order to quantify the throughput in bits per channel use (bpcu).

3.3.2 Logarithmic Utility

$$u_k(\gamma_k) = \log\gamma_k. \tag{3.20}$$

The logarithmic utility converges towards the achievable rate utility when $\gamma_k \gg 1$, i. e., when the SINR of user k is very high. For SINRs close to zero, on the other hand, the logarithmic utility is

strongly negative, whereas the achievable rate only yields non-negative values. We can therefore state that the logarithmic utility penalizes low SINRs more strongly.

By taking the first and second derivatives of $u_k(\gamma_k) = \log \gamma_k$ we obtain

$$c_k(\gamma_k) = 1. \tag{3.21}$$

Therefore the utility region is guaranteed to be convex.

3.3.3 Proportional Fair Rate Utility

$$u_k(\gamma_k) = \log\log(1+\gamma_k) = \log R_k. \tag{3.22}$$

If it is the overall goal to maximize the sum throughput of the system, it is possible that the optimal strategy involves setting the power of one or more users to zero, which may be considered "unfair" to those users. Ensuring that all users are treated equally, on the other hand, may lead to a strategy that does not efficiently utilize the potential of the system. The concept of *proportional fairness* aims at providing a compromise between these two extremes.

As it is defined in [39], a feasible rate K-tuple $(R_1, \ldots, R_K)$ is proportionally fair, if for every other feasible rate K-tuple $(R'_1, \ldots, R'_K)$

$$\sum_{k=1}^{K} \frac{R'_k - R_k}{R_k} \leq 0 \tag{3.23}$$

holds, i. e., if the sum of rates individually weighted with R_k^{-1} cannot be improved. Thus, a weak user's improvement is weighted strongly, while a strong user's degradation is given a lower weight.

It is shown in [39] that, if the rate region is convex, the proportional fair solution is unique and can be obtained by maximizing $\sum_k u_k(\gamma_k) = \sum_k \log R_k$ over all feasible power allocations. If the rate region is non-convex, which can be the case in our system, existence and uniqueness of a rate allocation fulfilling (3.23) is not guaranteed and maximizing $\sum_k \log R_k$ does not necessarily yield a proportional fair rate allocation [40]. Nonetheless, this utility can be used to obtain a rate allocation that presents a good compromise between maximizing sum throughput and total fairness.[2]

The proportional fair rate utility function again strongly penalizes low SINRs, i. e., $u_k(0) = -\infty$. Therefore, maximizing $\sum_k u_k(\gamma_k)$ will always yield a non-zero power allocation. The first and second derivatives of $u_k(\gamma_k) = \log\log(1+\gamma_k)$ are

$$u'_k(\gamma_k) = \frac{1}{(1+\gamma_k)\log(1+\gamma_k)} \qquad \text{and} \qquad u''_k(\gamma_k) = -\frac{1+\log(1+\gamma_k)}{(1+\gamma_k)^2(\log(1+\gamma_k))^2}. \tag{3.24}$$

Thus,

$$c_k(\gamma_k) = \frac{\gamma_k(1+\log(1+\gamma_k))}{(1+\gamma_k)\log(1+\gamma_k)} = \frac{\gamma_k \log(1+\gamma_k) + \gamma_k}{\gamma_k \log(1+\gamma_k) + \log(1+\gamma_k)}. \tag{3.25}$$

Since $\log(1+\gamma_k) < \gamma_k$ for $\gamma_k > 0$, the denominator is smaller than the numerator for positive SINRs and $c_k(\gamma_k) > 1$. With Theorem 3.1, the utility region is guaranteed to be convex.

[2] Another possible approach is to "convexify" the rate region by allowing time-sharing between different feasible power allocations. The proportional fair solution on the so-obtained *convex hull* of the rate region can then be determined, e. g., with a dual decomposition algorithm, cf. [41]. This approach is not further pursued in this work, as we do not allow for time-sharing in our system model.

A similar utility function

$$u_k(\gamma_k) = \begin{cases} \log(R_k - \bar{R}_k) & \text{if } R_k > \bar{R}_k, \\ -\infty & \text{otherwise} \end{cases} \tag{3.26}$$

can be used to determine the *Nash bargaining solution* (NBS) [42].[3] The NBS results from an axiomatic approach to modeling the outcome of a bargaining process, in which the users are faced with the threat of a "fallback" rate allocation $(\bar{R}_1, \ldots, \bar{R}_K)$ if the bargaining process fails. $(\bar{R}_1, \ldots, \bar{R}_K)$ may be the zero K-tuple, in which case the NBS utility is identical to the proportional fair rate utility, or any other pre-defined feasible rate allocation. Note that if $(\bar{R}_1, \ldots, \bar{R}_K)$ is on the Pareto boundary, there is always one user whose utility $u_k(\gamma_k)$ is $-\infty$. Again, the NBS is only defined for convex rate regions, and maximizing $\sum_k u_k(\gamma_k)$ may not lead to a solution fulfilling the NBS axioms in our system.

3.3.4 α-Fair Rate Utility

$$u_k(\gamma_k) = \frac{1}{1-\alpha} \left(\log(1+\gamma_k)\right)^{1-\alpha} = \frac{R_k^{1-\alpha}}{1-\alpha}. \tag{3.27}$$

The α-fair utility is defined for $\alpha \geq 0$ with $\alpha \neq 1$ [43]. It aims to provide a continuous tradeoff between sum throughput maximization and total fairness. With l'Hôpital's rule the proportional fair utility can be shown to be a natural continuation of the α-fair utility at $\alpha = 1$. For $\alpha = 0$ maximizing $\sum_k u_k(\gamma_k)$ is identical to maximizing the sum rate without taking fairness into consideration. As $\alpha \to \infty$, maximization of $\sum_k u_k(\gamma_k)$ approaches maxiziming the smallest element of the rate K-tuple, i. e., in the result all rates R_k will be equal and total fairness is guaranteed.

Again, the notion of α-fairness was originally only defined for convex rate regions. The use of this utility function for general regions is, however, also possible and is discussed in detail in [44].

The derivatives of $u_k(\gamma_k) = (1-\alpha)^{-1} \left(\log(1+\gamma_k)\right)^{1-\alpha}$ are

$$u_k'(\gamma_k) = \frac{1}{(1+\gamma_k)(\log(1+\gamma_k))^{\alpha}} \tag{3.28}$$

$$u_k''(\gamma_k) = -\frac{\alpha + \log(1+\gamma_k)}{(1+\gamma_k)^2(\log(1+\gamma_k))^{\alpha+1}} \tag{3.29}$$

and therefore

$$c_k(\gamma_k) = \frac{\gamma_k(\alpha + \log(1+\gamma_k))}{(1+\gamma_k)(\log(1+\gamma_k))} = \frac{\gamma_k \log(1+\gamma_k) + \alpha\gamma_k}{\gamma_k \log(1+\gamma_k) + \log(1+\gamma_k)}. \tag{3.30}$$

If $\alpha > 1$, the denominator of $c_k(\gamma_k)$ is again smaller than the numerator and the utility region is convex. If $\alpha < 1$, we can verify with l'Hôpital's rule that $\lim_{\gamma_k \to 0} c_k(\gamma_k) = \alpha$ and the sufficient condition for convexity of the utility region is not fulfilled.

[3] Note that $u_k(\gamma_k)$ is not strictly increasing in γ_k for $R_k < \bar{R}_k$, and therefore technically does not fall under our definition of a utility function. Many of the techniques discussed in the following can be applied nonetheless for this choice of $u_k(\gamma_k)$.

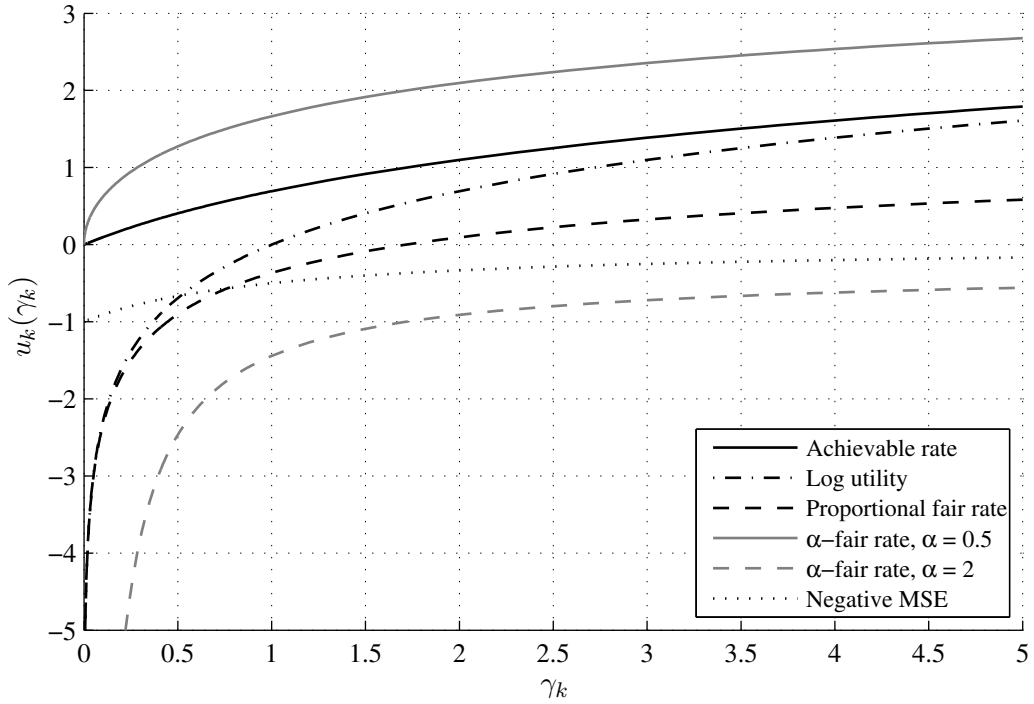


Figure 3.2: Discussed utility functions over the SINR γ_k

3.3.5 Negative Mean Squared Error

$$u_k(\gamma_k) = -\frac{1}{1+\gamma_k}. \tag{3.31}$$

Let us assume that the transmit symbol is formed as $x_k = \sqrt{p_k}s_k$, where s_k is the unit-variance data symbol, and that the receiver forms an estimate of the data symbol as $\hat{s}_k = g_k y_k$ using the scalar receiver coefficient $g_k \in \mathbb{C}$. The mean squared error (MSE) between the estimated and the true data symbol is defined as $\mathrm{E}\left[|\hat{s}_k - s_k|^2\right]$. The receiver coefficient g_k that minimizes the MSE can be shown to be

$$g_k = \frac{h_{kk}^* \sqrt{p_k}}{\sum\limits_{j=1}^{K} |h_{kj}|^2 p_j + \sigma^2} \tag{3.32}$$

and the MSE resulting from using the optimal g_k is

$$\mathrm{E}\left[|\hat{s}_k - s_k|^2\right] = \frac{\sum_{j\neq k}|h_{kj}|^2 p_j + \sigma^2}{\sum_j |h_{kj}|^2 p_j + \sigma^2} = \frac{1}{1+\gamma_k}. \tag{3.33}$$

As the MSE should be as small as possible, the negative MSE is used as a utility function.

For the negative MSE utility, $u_k(0) = -1$. With Observation 3.2 we can conclude that $u_k(\gamma_k)$ does not belong to the class of utility functions that guarantee a convex utility region. This can also be seen from

$$c_k(\gamma_k) = \frac{2\gamma_k}{1+\gamma_k}. \tag{3.34}$$

Figure 3.2 shows the behavior of the different discussed utility functions over the SINR. It can be seen that those utility functions that guarantee a convex region have a more pronounced

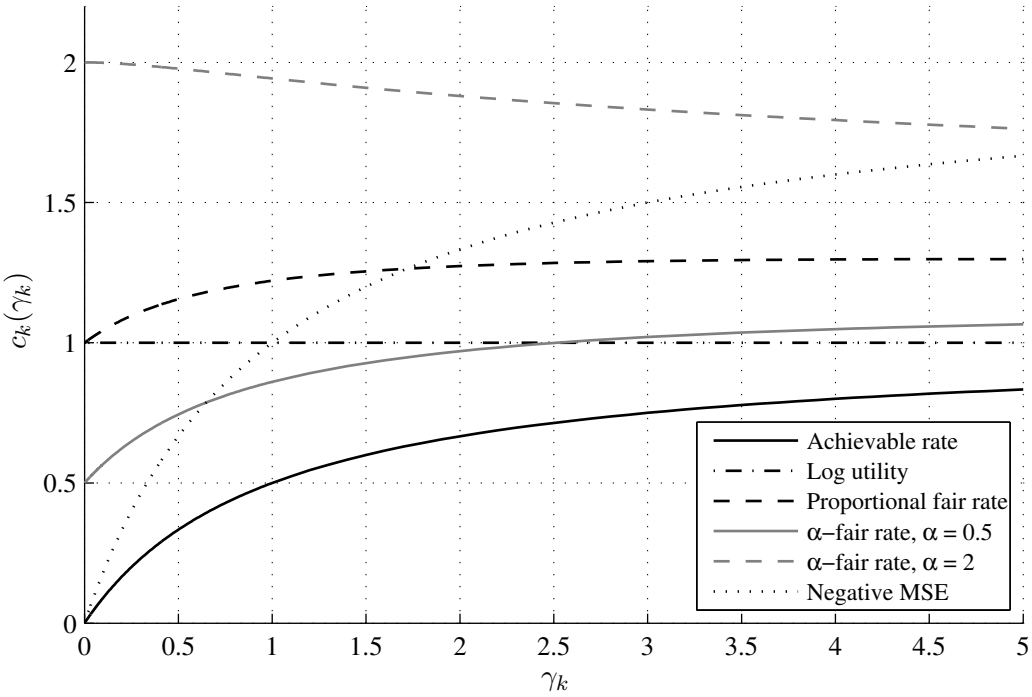


Figure 3.3: Value of $c_k(\gamma_k)$ over the SINR γ_k for the discussed utility functions

rightward curvature than the others. Figure 3.3 shows the value of $c_k(\gamma_k)$ for the same utility functions.

3.4 Concepts of Optimality

The K users' utility functions are competing objectives in the sense that it is generally not possible to maximize them simultaneously; it is therefore not possible to identify an undisputably optimal power allocation. In this section we will discuss some different criteria that are used to define optimality of a power allocation as well as the challenges in determining these power allocations.

3.4.1 The Selfish Solution or Nash Equilibrium

Let us assume that each user k is interested in maximizing its own utility function $u_k(\gamma_k)$ and is not concerned about the value of the other users' utility functions $u_j(\gamma_j)$ with $j \neq k$. Since the utility function $u_k(\gamma_k)$ of user k is increasing in p_k, in this case each user will transmit with full power, i. e., the resulting *selfish solution* is the power allocation $(p_1 = 1, \ldots, p_K = 1)$.

This setup also fits well into the framework of game theory: the players in the game are the K users and each user k can choose a strategy $p_k \in [0, 1]$. The strategies of all users $(p_1, \ldots, p_K)$ then determine the payoff for each user k, which is the utility function $u_k(\gamma_k)$.

A *Nash equilibrium* of a game is defined as a set of strategies for which any unilateral change of strategy by a single user cannot improve its own payoff. In our power control game, only the selfish solution is a Nash equlibrium: if p_k is strictly less than one, user k can improve its payoff by changing p_k to one; when all powers are one, on the other hand, every user k can only decrease its own payoff by changing p_k.

The selfish solution (or Nash equilibrium) is on the Pareto boundary of the power region and therefore also on the Pareto boundary of the SINR and utility regions. It is worth noting that if we assume fixed non-zero channel coefficients and let $\sigma^2 \to 0$, all SINRs γ_k are bounded from above when the selfish solution is employed. This suggests that if the noise power is low, the selfish solution may not be a very efficient strategy; in the numerical evaluation in Section 3.6 this will become apparent, in particular, for the achievable rate utility.

3.4.2 Altruistic Solutions

As a counterpart to the selfish solution, we define a power allocation to be *altruistic,* if every user k

1) cannot unilaterally improve the utility function $u_j(\gamma_j)$ of any user $j \neq k$ by changing p_k and
2) cannot improve its own utility $u_k(\gamma_k)$ by changing p_k without decreasing at least one other utility $u_j(\gamma_j)$ with $j \neq k$.

As an example, consider $K = 2$ and all four channel coefficients to be non-zero. The altruistic solutions in this case are $(p_1 = 1, p_2 = 0)$ and $(p_1 = 0, p_2 = 1)$: if both powers were non-zero, one user would always be able to improve the other user's utility by decreasing its own power; if one power was zero and the other less than one, both users would not be able to improve the other user's utility, but the active user would be able to improve its own utility without harming the inactive user.

In general, only power allocations with $p_k \in \{0, 1\}$ for all $k \in \{1, \ldots, K\}$ can be altruistic solutions. To show this, let us consider $0 < p_k < 1$: either all other utilities do not depend on p_k, in which case $u_k(\gamma_k)$ can be improved by setting $p_k = 1$ without harming the other users, or some other utility is decreasing in p_k, in which case this utility could be improved by setting $p_k = 0$. Furthermore, every user k that has full power $p_k = 1$ cannot experience any interference if the power allocation is an altruistic solution, i. e., the denominator in (3.5) consists only of σ^2.

When all channel coefficients are non-zero, the altruistic solutions are those power allocations in which exactly one p_k is one and all others are zero. If some cross channel coefficients h_{kj} with $k \neq j$ are zero, the set of altruistic solutions can be different. Consider, e. g., a system with $K = 3$ users and $h_{12} = h_{21} = 0$. Here, $(p_1 = 1, p_2 = 1, p_3 = 0)$ is an altruistic solution, but $(p_1 = 0, p_2 = 1, p_3 = 0)$ and $(p_1 = 1, p_2 = 0, p_3 = 0)$ are not.

In contrast to the selfish solution, at least one SINR γ_k grows without bound for $\sigma^2 \to 0$ if an altruistic solution is employed, so that user k has the full benefit of a channel with vanishing noise power. We will see in Section 3.6 that for the rate utility this can be a great advantage over the selfish solution.

3.4.3 Feasibility and the Power Minimization Problem

For some applications it might be necessary that every user is provided with a certain *quality of service,* which could be, e. g., a minimum rate necessary for operation. To this end, we assign each user k a utility target $\bar{u}_k$ and require that $u_k(\gamma_k) \geq \bar{u}_k$ for all $k \in \{1, \ldots, K\}$. Since there may be many power allocations that fulfill the utility targets, we define the optimal solution to be the power allocation that fulfills the utility targets of all users and at the same time minimizes the sum of the transmit powers $\sum_k p_k$.

It is of course also possible that the utility targets are outside the utility region, i. e., that they are *infeasible.* The first step, therefore, is to check the feasibility of the K-tuple of utility targets. We define $(\bar{\gamma}_1, \ldots, \bar{\gamma}_K)$ as the SINR K-tuple that corresponds to the utility targets; $\bar{\gamma}_k$ can be

determined by taking the inverse utility function of $\bar{u}_k$. To begin with, we can formulate a simple necessary and a simple sufficient condition for the feasiblity of $(\bar{\gamma}_1, \ldots, \bar{\gamma}_K)$.
Observation 3.3. The SINR allocation $(\bar{\gamma}_1, \ldots, \bar{\gamma}_K)$ can only be feasible if

$$\bar{\gamma}_k \leq \frac{|h_{kk}|^2}{\sigma^2} \qquad \forall k \in \{1, \ldots, K\}. \tag{3.35}$$

Observation 3.4. The SINR allocation $(\bar{\gamma}_1, \ldots, \bar{\gamma}_K)$ is feasible if

$$\bar{\gamma}_k \leq \frac{|h_{kk}|^2}{\sum\limits_{j \neq k} |h_{kj}|^2 + \sigma^2} \qquad \forall k \in \{1, \ldots, K\} \tag{3.36}$$

since this means that $(\bar{\gamma}_1, \ldots, \bar{\gamma}_K)$ is element-wise smaller or equal to the SINR K-tuple resulting from the selfish solution.

Assuming that feasibility cannot be ruled out with Observation 3.3, we proceed to determine the power allocation $(p_1, \ldots, p_K)$ that fulfills the SINR targets with equality. The process of determining a power allocation from an SINR K-tuple is examined in detail in the proof of Proposition 3.2; we therefore only give a brief description here and refer to Appendix A2 for the details.

First, all users with a target SINR of zero are removed from the system, as it is clear that their respective powers are also zero. Next, assuming that the remaining users have the indices $\{1, \ldots, K\}$, we define the matrix

$$\boldsymbol{A} = \begin{bmatrix} \frac{|h_{11}|^2}{\bar{\gamma}_1} & -|h_{12}|^2 & \cdots & -|h_{1K}|^2 \\ -|h_{21}|^2 & \frac{|h_{22}|^2}{\bar{\gamma}_2} & \cdots & -|h_{2K}|^2 \\ \vdots & \vdots & \ddots & \vdots \\ -|h_{K1}|^2 & -|h_{K2}|^2 & \cdots & \frac{|h_{KK}|^2}{\bar{\gamma}_K} \end{bmatrix} \tag{3.37}$$

and the vector $\boldsymbol{p} = [p_1, \ldots, p_K]^{\mathrm{T}}$ and compute the power allocation as

$$\boldsymbol{p} = \sigma^2 \boldsymbol{A}^{-1} \mathbf{1}. \tag{3.38}$$

If the matrix $\boldsymbol{A}$ is not invertible, the SINR K-tuple $(\bar{\gamma}_1, \ldots, \bar{\gamma}_K)$ cannot be feasible, since $\boldsymbol{A}$ is non-singular for all feasible SINR K-tuples, cf. Appendix A2; if $\boldsymbol{A}$ is invertible and one or more elements of the resulting vector $\boldsymbol{p}$ are negative or greater than one, the target SINRs are also infeasible; if every element of $\boldsymbol{p}$ is in the interval $[0, 1]$, the target SINRs are feasible and we have found the corresponding power allocation.

From the properties of the SINR we know that starting from $(p_1, \ldots, p_K)$ no power can be decreased without violating the SINR targets. Therefore, the power allocation that fulfills the SINR/utility targets with equality is also optimal in the sense that the total transmit power is minimized.

3.4.4 The SINR Balancing Problem

Another possible goal of an optimization of the power allocation is to ensure that all users achieve the same SINR and to maximize this common SINR, thus focusing on complete fairness. More generally, instead of requiring all SINRs to be identical, we can consider arbitrary fixed ratios of the SINRs to one another. This is called the *SINR balancing problem.*

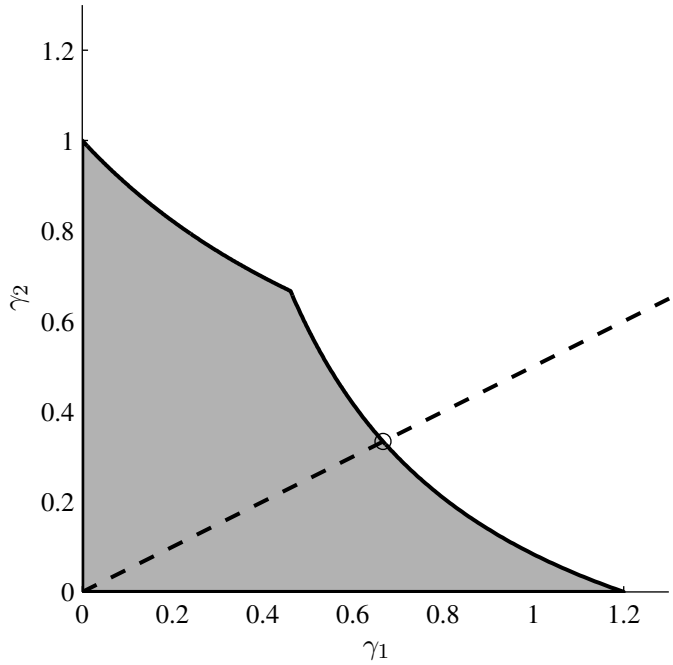


Figure 3.4: SINR region for $K = 2$ with $|h_{11}|^2 = 1.2$, $|h_{12}|^2 = 1.6$, $|h_{21}|^2 = 0.5$, and $|h_{22}|^2 = 1$. The noise power is $\sigma^2 = 1$. All points on the dashed line are SINR pairs of the form $(\beta\bar{\gamma}_1, \beta\bar{\gamma}_2)$, where $\bar{\gamma}_1 = 2$ and $\bar{\gamma}_2 = 1$. The circle marks the balancing solution.

Formally, for a given K-tuple $(\bar{\gamma}_1, \ldots, \bar{\gamma}_K)$, we would like to determine the highest real-valued scalar β for which $(\beta\bar{\gamma}_1, \ldots, \beta\bar{\gamma}_K)$ is a feasible SINR K-tuple. For example, with $\bar{\gamma}_k = 1$ for all $k \in \{1, \ldots, K\}$, we are searching for the highest SINR that can be achieved by all users simultaneously; if we set $\bar{\gamma}_1 = 2$, the resulting SINR of the first user will be twice as high as that of all other users.

We can think of all feasible SINR K-tuples that can be written as $(\beta\bar{\gamma}_1, \ldots, \beta\bar{\gamma}_K)$ as lying on a line segment starting at the origin, cf. Figure 3.4. It is intuitively clear that, as the SINR region is comprehensive, the solution to the SINR balancing problem must lie on the Pareto boundary of the SINR region and that there is a unique value β for which $(\beta\bar{\gamma}_1, \ldots, \beta\bar{\gamma}_K)$ is on the Pareto boundary.

A detailed discussion of the SINR balancing problem can be found in [35]. In the following we briefly derive a method to obtain the solution that requires only a fixed number of steps. We begin by characterizing the power allocations $(p_1, \ldots, p_K)$ that result in an SINR K-tuple of the form $(\beta\bar{\gamma}_1, \ldots, \beta\bar{\gamma}_K)$. From (3.5), the SINR of user k is

$$\beta\bar{\gamma}_k = \frac{|h_{kk}|^2 p_k}{\sum\limits_{j \neq k} |h_{kj}|^2 p_j + \sigma^2} \tag{3.39}$$

which we can rewrite as

$$\frac{\bar{\gamma}_k}{|h_{kk}|^2} \left(\sum_{j \neq k} |h_{kj}|^2 p_j + \sigma^2 \right) = \frac{1}{\beta} p_k. \tag{3.40}$$

By defining the matrix

$$\boldsymbol{C} = \begin{bmatrix} 0 & \bar{\gamma}_1 \frac{|h_{12}|^2}{|h_{11}|^2} & \cdots & \bar{\gamma}_1 \frac{|h_{1K}|^2}{|h_{11}|^2} \\ \bar{\gamma}_2 \frac{|h_{21}|^2}{|h_{22}|^2} & 0 & \cdots & \bar{\gamma}_2 \frac{|h_{2K}|^2}{|h_{22}|^2} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\gamma}_K \frac{|h_{K1}|^2}{|h_{KK}|^2} & \bar{\gamma}_K \frac{|h_{K2}|^2}{|h_{KK}|^2} & \cdots & 0 \end{bmatrix} \tag{3.41}$$

and the vectors

$$\boldsymbol{d} = \begin{bmatrix} \frac{\bar{\gamma}_1 \sigma^2}{|h_{11}|^2} \\ \vdots \\ \frac{\bar{\gamma}_K \sigma^2}{|h_{KK}|^2} \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_K \end{bmatrix} \tag{3.42}$$

we can write the relationship between all users' powers and SINRs in matrix-vector notation as

$$\boldsymbol{C}\boldsymbol{p} + \boldsymbol{d} = \beta^{-1}\boldsymbol{p}. \tag{3.43}$$

Since we are looking for a solution to this equation that lies on the Pareto boundary, we know that at least one power must be one. If we assume that $p_k = 1$ and define an extended power vector $\hat{\boldsymbol{p}} = [\boldsymbol{p}, 1]^{\mathrm{T}}$ and the matrix

$$\hat{\boldsymbol{C}}_k = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{d} \\ \boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{C} & \boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{d} \end{bmatrix} \tag{3.44}$$

where $\boldsymbol{e}_k$ is the kth column of the $K \times K$ identity matrix, we can write (3.43) as

$$\hat{\boldsymbol{C}}_k \hat{\boldsymbol{p}} = \beta^{-1}\hat{\boldsymbol{p}} \tag{3.45}$$

since

$$\boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{C}\boldsymbol{p} + \boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{d} = \beta^{-1}p_k = \beta^{-1}. \tag{3.46}$$

In words, $\hat{\boldsymbol{p}}$ is an eigenvector of $\hat{\boldsymbol{C}}_k$. Since we do not know which user k has full power in the solution to the balancing problem, we proceed as follows: for all $k \in \{1, \ldots, K\}$, we check whether there exists an eigenvector of $\hat{\boldsymbol{C}}_k$ with only non-negative elements and in particular with a positive kth and last element which corresponds to a positive eigenvalue. If such an eigenvector exists, it is then scaled so that $p_k = 1$. If all elements p_j with $j \neq k$ are in the interval $[0, 1]$, this vector contains the solution to the balancing problem.

As discussed above, it is clear from the properties of the SINR region that a unique power allocation exists that solves the balancing problem, and that this procedure will therefore always yield the solution. We note, however, that some statements on the existence and uniqueness of eigenvectors with non-negative entries for the matrices $\hat{\boldsymbol{C}}_k$ can also be made with the Perron-Frobenius theorem (e. g., [37, 45]).

If all users have identical utility functions and $\bar{\gamma}_1 = \ldots = \bar{\gamma}_K$, the utilities of all users will also be balanced. In general, however, the *utility balancing problem,* i. e., fixing the ratio of the users' utilities to one another, is considerably more difficult and cannot be solved in a straightforward way.

3.4.5 The Sum Utility Problem

If there are no hard target-SINR or fairness constraints, the goal from the point of view of the system operator is usually to maximize some sort of overall efficiency of the system. With properly designed utility functions, the overall efficiency of the system can be expressed as the sum of the utility functions of all users. Formally, we therefore wish to solve the optimization problem

$$\max_{p_1,\ldots,p_K} \sum_{k=1}^{K} u_k(\gamma_k) \qquad \text{s.t.:} \quad 0 \leq p_k \leq 1 \quad \forall k \in \{1,\ldots,K\}. \tag{3.47}$$

To begin with, it is clear that the solution to the sum utility problem must lie on the Pareto boundary of the utility region. Any utility K-tuple that is not on the Pareto boundary is by definition strictly dominated in terms of individual utilities, and therefore also in terms of sum utility, by another feasible utility K-tuple.

With the Lagrangian multipliers $\mu_1,\ldots,\mu_K$ and $\nu_1,\ldots,\nu_K$, we can formulate the *Karush-Kuhn-Tucker* (KKT) necessary conditions for optimality, cf. [36, Chapter 5]:

$$\sum_{j=1}^{K} \frac{\partial u_j(\gamma_j)}{\partial p_k} + \mu_k - \nu_k = 0 \qquad \forall k \in \{1,\ldots,K\} \tag{3.48}$$

$$\mu_k \geq 0 \qquad \forall k \in \{1,\ldots,K\} \tag{3.49}$$

$$\nu_k \geq 0 \qquad \forall k \in \{1,\ldots,K\} \tag{3.50}$$

$$\mu_k p_k = 0 \qquad \forall k \in \{1,\ldots,K\} \tag{3.51}$$

$$\nu_k (1 - p_k) = 0 \qquad \forall k \in \{1,\ldots,K\} \tag{3.52}$$

$$0 \leq p_k \leq 1 \qquad \forall k \in \{1,\ldots,K\}. \tag{3.53}$$

Proposition 3.6. *If the utility region is convex,* (3.47) *is equivalent to a concave maximization problem and the above KKT conditions are not only necessary, but also sufficient for global optimality.*

Proof. The proof is given in Appendix A4. □

For convex utility regions, the problem can thus be solved efficiently with a number of general purpose algorithms, e. g., projected gradient algorithms [46] or interior point methods [46, 36]. If the utility region is non-convex, on the other hand, solving the sum utility problem may be very difficult. In fact, it is shown in [47] that for the rate utility the problem is *NP-hard,* i. e., it belongs to a class of problems that is believed not to be solvable with a number of computations that grows only polynomially in the system dimensions. We can interpret the difficulty of solving the optimization problem as the potential existence of multiple local maxima, where small changes in the power allocation in any feasible direction lead to a decrease in sum utility. Each of the local maxima also fulfills the KKT conditions. As a consequence, gradient-based search algorithms can converge to one of the local optima and it is by no means guaranteed that the solution to the problem is found.

Some global optimization techniques do exist, however, that are guaranteed to find the solution to (3.47) with arbitrary precision for any utility function. A notable example is the *polyblock algorithm* [48], which is based on successively tightening an outer bound on the utility region by repeatedly projecting points that are outside the utility region onto the Pareto boundary, e. g.,

by solving balancing problems. The polyblock algorithm was specifically applied to our system model in [49]. Such algorithms are computationally very demanding, however, and their complexity grows rapidly with the system size. Therefore, their application is only feasible for small K.

3.4.5.1 Maximizing the Sum Rate

Due to its significance as the overall system throughput limit, maximization of the sum rate utility is the main focus of this work. With (3.18) and (3.5), the utility function for user k is

$$\begin{aligned} u_k(\gamma_k) = R_k = \log(1+\gamma_k) &= \log\left(1 + \frac{|h_{kk}|^2 p_k}{\sum_{j\neq k}|h_{kj}|^2 p_j + \sigma^2}\right) \\ &= \log\left(\sum_j |h_{kj}|^2 p_j + \sigma^2\right) - \log\left(\sum_{j\neq k}|h_{kj}|^2 p_j + \sigma^2\right). \end{aligned} \tag{3.54}$$

The derivative of the sum utility w. r. t. the power p_k of user k is

$$\sum_j \frac{\partial R_j}{\partial p_k} = \frac{|h_{kk}|^2}{\sum_i |h_{ki}|^2 p_i + \sigma^2} - \sum_{j\neq k}\left(\frac{|h_{jk}|^2}{\sum_{i\neq j}|h_{ji}|^2 p_i + \sigma^2} - \frac{|h_{jk}|^2}{\sum_i |h_{ji}|^2 p_i + \sigma^2}\right). \tag{3.55}$$

Note that the terms inside the brackets are non-negative since the two fractions have the same numerator, but the second denominator has one more summand than the first.

Since the rate region can be non-convex (cf. Section 3.3.1), the KKT conditions do not necessarily have a unique solution, and a power allocation satisfying the KKT conditions is not necessarily globally optimal. Furthermore, as noted previously, the sum rate maximization problem has been shown to be NP-hard [47].

Let us first assume a system with $K = 2$ users. The derivatives of the sum rate are

$$\frac{\partial(R_1+R_2)}{\partial p_1} = \frac{|h_{11}|^2}{|h_{11}|^2 p_1 + |h_{12}|^2 p_2 + \sigma^2} - \left(\frac{|h_{21}|^2}{|h_{21}|^2 p_1 + \sigma^2} - \frac{|h_{21}|^2}{|h_{21}|^2 p_1 + |h_{22}|^2 p_2 + \sigma^2}\right) \tag{3.56}$$

$$\frac{\partial(R_1+R_2)}{\partial p_2} = \frac{|h_{22}|^2}{|h_{21}|^2 p_1 + |h_{22}|^2 p_2 + \sigma^2} - \left(\frac{|h_{12}|^2}{|h_{12}|^2 p_2 + \sigma^2} - \frac{|h_{12}|^2}{|h_{11}|^2 p_1 + |h_{12}|^2 p_2 + \sigma^2}\right). \tag{3.57}$$

Now assume that the power allocation (p_1, p_2) fulfills the KKT conditions and that $0 < p_1 < 1$, i. e., none of the constraints on p_1 is active. Consequently, $\mu_1 = \nu_1 = 0$ and due to (3.48) the derivative in (3.56) must be zero. With the definitions

$$a_1 = \frac{|h_{11}|^2}{|h_{11}|^2 p_1 + |h_{12}|^2 p_2 + \sigma^2} \tag{3.58}$$

$$a_2 = \frac{|h_{21}|^2}{|h_{21}|^2 p_1 + |h_{22}|^2 p_2 + \sigma^2} \tag{3.59}$$

$$a_3 = \frac{|h_{21}|^2}{|h_{21}|^2 p_1 + \sigma^2} \tag{3.60}$$

it follows that $a_1 + a_2 = a_3$. The second derivative of the sum rate w. r. t. p_1, obtained by taking the derivative of (3.56) w. r. t. p_1, is $-a_1^2 - a_2^2 + a_3^2$. Since

$$-a_1^2 - a_2^2 + a_3^2 > -a_1^2 - 2a_1a_2 - a_2^2 + a_3^2 = -(a_1+a_2)^2 + a_3^2 = 0 \tag{3.61}$$

the first derivative is increasing in p_1 at the power allocation (p_1, p_2). Therefore such a power allocation cannot be a local maximum. Similarly, if $0 < p_2 < 1$, the power allocation cannot be a local maximum since the second derivative of the sum rate w. r. t. p_2 is positive whenever the first derivative is zero. Consequently, for $K = 2$ only the three power allocations $(p_1 = 1, p_2 = 0)$, $(p_1 = 0, p_2 = 1)$, $(p_1 = 1, p_2 = 1)$ can be sum-rate optimal and the sum rate maximization problem can be solved by comparing the sum rates of these three power allocations.

Unfortunately, this result is not generalizable to $K > 2$ users; for $K = 3$ users, scenarios can be constructed where in the optimal power allocation one power is strictly between zero and one. Then, finding power allocations that fulfill the KKT conditions does not seem to be straightforward. Even if it is known which constraints are active in the optimal power allocation, i. e., which users have zero power and which have full power, determining the powers of the remaining users involves solving a system of nonlinear (polynomial) equations.

It is, however, possible to characterize the optimal power allocations for sufficiently high and sufficiently low noise power σ^2. We begin with the case where σ^2 is large and define this as the *low-SNR regime*. The derivative of the sum rate (3.55) can also be stated as

$$\frac{\partial}{\partial p_k} \sum_j R_j = \frac{|h_{kk}|^2}{\sum_i |h_{ki}|^2 p_i + \sigma^2} - \sum_{j \neq k} \frac{|h_{jk}|^2 |h_{jj}|^2 p_j}{\left(\sum_i |h_{ji}|^2 p_i + \sigma^2\right) \left(\sum_{i \neq j} |h_{ji}|^2 p_i + \sigma^2\right)}. \tag{3.62}$$

The first (positive) term in this expression is clearly in $\mathrm{O}(\sigma^{-2})$ as $\sigma^2 \to \infty$; the remaining (negative) summands are in $\mathrm{O}(\sigma^{-4})$ and therefore approach zero more rapidly than the positive term. We can conclude that, for sufficiently high σ^2, all derivatives are greater than zero and the KKT conditions are only fulfilled by the power allocation $p_1 = \ldots = p_K = 1$. Therefore, in the low-SNR regime the selfish solution is optimal.

We define the case where σ^2 is low and σ^{-2} is asymptotically high, on the other hand, as the *high-SNR regime*. From (3.54), we observe that as $\sigma^{-2} \to \infty$

$$R_k = \begin{cases} \log(\sigma^{-2}) + \mathrm{O}(1) & \text{if} \quad \sum_{j \neq k} |h_{kj}|^2 p_j = 0 \quad \text{and} \quad p_k > 0 \\ \mathrm{O}(1) & \text{otherwise,} \end{cases} \tag{3.63}$$

i. e., R_k grows without bound as $\sigma^{-2} \to \infty$ only if the received interference power of user k is zero. Therefore, to maximize the sum rate in the high-SNR regime, it is necessary that there exists at least one user that does not experience any interference. In the case that all cross channel coefficients h_{kj} with $j \neq k$ are non-zero, it is only possible to remove interference for a user by setting the power of all other users to zero; thus, the optimal power allocation is to give the user k with the highest direct channel magnitude $|h_{kk}|$ the full power and all other users zero power, which also is an altruistic solution. If some cross channels are zero, the sum-rate optimal high-SNR power allocation is not necessarily an altruistic solution.

3.4.5.2 Maximizing the Sum Logarithmic Utility

We briefly investigate the sum utility problem with the logarithmic utility, as this is an example for a problem in which the KKT conditions are also sufficient. Furthermore, the logarithmic utility approaches the achievable rate when $\gamma_k \gg 1$ and therefore is also relevant in terms of system throughput. With (3.20) and (3.5), we can express the utility function as

$$u_k(\gamma_k) = \log(\gamma_k) = \log p_k + \log|h_{kk}|^2 - \log\left(\sum_{j \neq k} |h_{kj}|^2 p_j + \sigma^2\right). \tag{3.64}$$

Since a power close to zero results in a strongly negative utility, it is clear that the some utility can only be maximized if $p_k > 0$ for all $k \in \{1, \ldots, K\}$. Also, since the solution must lie on the Pareto boundary of the utility region, at least one power p_k is equal to one. The derivatives of the sum utility are

$$\frac{\partial}{\partial p_k} \sum_j u_j(\gamma_j) = \frac{1}{p_k} - \sum_{j \neq k} \frac{|h_{jk}|^2}{\sum_{i \neq j} |h_{ji}|^2 p_i + \sigma^2}. \tag{3.65}$$

For a system with $K = 2$ users, we can therefore state

$$\frac{\partial\big(u_1(\gamma_1) + u_2(\gamma_2)\big)}{\partial p_1} = \frac{1}{p_1} - \frac{|h_{21}|^2}{|h_{21}|^2 p_1 + \sigma^2} \tag{3.66}$$

$$\frac{\partial\big(u_1(\gamma_1) + u_2(\gamma_2)\big)}{\partial p_2} = \frac{1}{p_2} - \frac{|h_{12}|^2}{|h_{12}|^2 p_2 + \sigma^2}. \tag{3.67}$$

Since the positive part is always strictly greater than the negative part, both derivatives are always positive. Therefore the power allocation $(p_1 = 1, p_2 = 1)$ fulfills the KKT conditions and is globally optimal.

Again, for systems with $K > 2$ users, determining a power allocation that fulfills the KKT conditions is not straightforward. Even if it is known which users have full power, a nonlinear (polynomial) system of equations must be solved to determine the powers that are less than one.

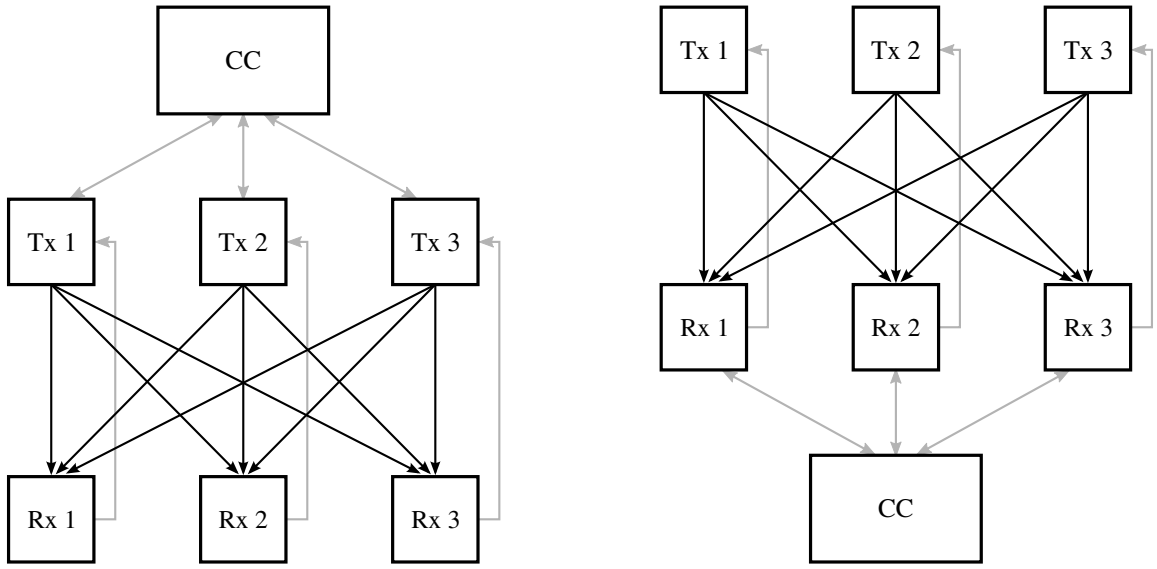
In the low-SNR regime, i. e., when σ^2 is sufficiently large, it can be seen from (3.65) that all derivatives are positive regardless of the power allocation and the selfish solution is optimal. In the high-SNR regime, on the other hand, an obvious simplification is not possible and the solution still requires solving a system of nonlinear equations.

3.5 Information Exchange and Distributed Optimization

So far, we have assumed that the solution fulfilling the desired optimality criterion can be computed with knowledge of all system parameters, i. e., the channel coefficients h_{kj} for all pairs $(k, j) \in \{1, \ldots, K\}^2$ and the noise power σ^2 as well as the utility functions $u_k(\gamma_k)$ for all $k \in \{1, \ldots, K\}$. In a wireless communications model with multiple autonomous transmitter-receiver pairs, however, not every node is able to estimate every channel coefficient. For example, in a system with $K = 3$ users, the first receiver might have knowledge of the channel coefficients h_{11}, h_{12}, and h_{13}, but not of h_{23} or h_{32}. In order to be able to determine the above optimal power allocations (with notable exception of the selfish solution), we therefore need to assume a *centralized computer* that first collects all necessary parameters, then determines the optimal power allocation, and finally informs the users which power they should use for transmission. For the exchange of information it is necessary that the centralized computer is connected to the users by means of a bi-directional signaling link.

In order to investigate the information exchange procedures more closely, we make the following assumptions:

- Receiver k is able to perfectly estimate all channel coefficients that contribute to its received signal y_k, i. e., h_{kj} for all $j \in \{1, \ldots, K\}$, in an initial pilot phase.
- The noise power σ^2 is universally known at all transmitters and receivers. To justify this assumption, we note that we could assume unit noise power and renormalize all channel coefficients h_{kj} to h_{kj}/σ, which would not change the users' SINRs.



(a) Connected transmitters

(b) Connected receivers

Figure 3.5: Communication links (black) and feedback/signaling links (gray) for the centralized computation model in a $K = 3$ user system.

- A feedback link from receiver k to transmitter k can be used to reliably signal channel coefficients or other system parameters.

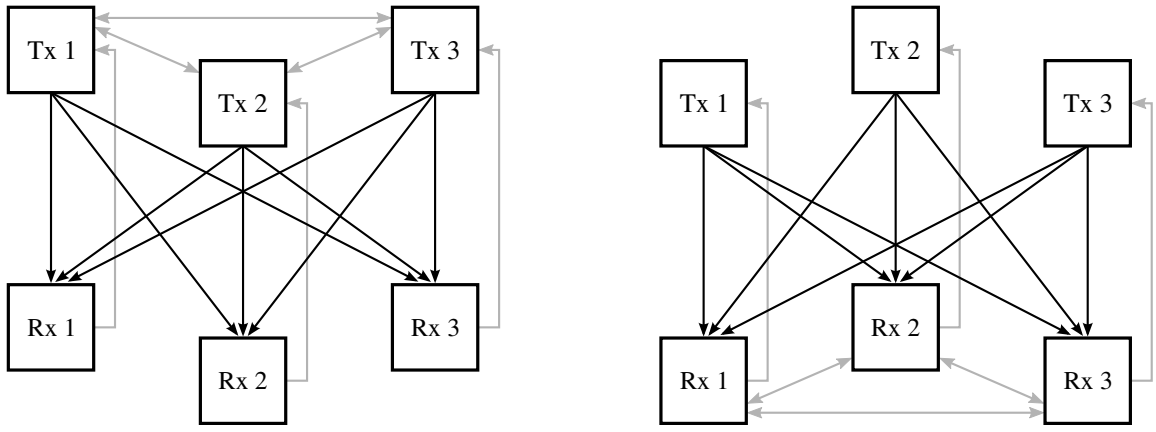
In out centralized model, the centralized computer can be either connected to the K transmitters or the K receivers via bi-directional signaling links, cf. Figure 3.5. If the centralized computer is on the transmitter side, information is exchanged in the following way:

1) After the channel estimation phase, each receiver k feeds the estimated channel coefficients h_{kj} for all $j \in \{1, \ldots, K\}$ back to transmitter k using the feedback link.
2) The transmitters send all channel coefficients to the centralized computer over the signaling links.
3) The centralized computer determines the transmit strategy.
4) The centralized computer informs all transmitters which power p_k they should use and which SINR γ_k they can expect, which is necessary for the choice of the coding scheme.

Similarly, if the centralized computer is on the receiver side, the procedure is as follows:

1) After the channel estimation phase, each receiver k sends the estimated channel coefficients h_{kj} for all $j \in \{1, \ldots, K\}$ to the centralized computer.
2) The centralized computer determines the transmit strategy.
3) The centralized computer informs receiver k of the power p_k and the resulting SINR γ_k.
4) Receiver k feeds back p_k and γ_k to transmitter k.

If a centralized computer is not available, we could assume that the users are able to exchange the channel coefficients and that then each user for himself can determine the optimal power allocation in the same way the centralized computer would. As discussed before, however, determining the optimal power allocation can be computationally very demanding, especially when K is large; performing the identical computations at every user is not only a waste of resources, it may also exceed the capabilities of the nodes in a mobile environment. We therefore introduce the *distributed model,* where each node only has very limited computational power and is not capable of solving



(a) Connected transmitters

(b) Connected receivers

Figure 3.6: Communication links (black) and feedback/signaling links (gray) for the distributed computation model in a $K = 3$ user system.

multi-variate optimization problems or systems of nonlinear equations. For the necessary information exchange between the users, we again assume either the transmitters or the receivers to be connected via bi-directional signaling links, cf. Figure 3.6.

In the distributed model, one-shot solutions where the users can determine their optimal power with one simple calculation are not possible in general. In the following we instead investigate communications procedures where, starting from some initial power allocation, the users make use of the signaling links to gradually improve the utility K-tuple with the goal of reaching some sort of optimality.

In addition to the three bullet points at the beginning of this section, we make another assumption for the iterative procedures discussed in the following:

- Receiver k can at any time perfectly estimate the current SINR γ_k.

We also introduce the super-script (ℓ) to denote the iteration number. Thus, the initial power allocation is $\left(p_1^{(0)}, \ldots, p_K^{(0)}\right)$, or in vector notation $\boldsymbol{p}^{(0)} = \left[p_1^{(0)}, \ldots, p_K^{(0)}\right]^{\mathrm{T}}$; after ℓ iterations, user k has power $p_k^{(\ell)}$ and SINR $\gamma_k^{(\ell)}$.

One iteration of a distributed power control algorithm typically consists of three steps:

1) Receiver k measures the current SINR $\gamma_k^{(\ell)}$.
2) Based on this measurement, information is exchanged between receiver k and transmitter k over the feedback link and if necessary between different users over the bi-directional signaling links.
3) User k computes the updated power $p_k^{(\ell+1)}$ as a function of $p_k^{(\ell)}$, $\gamma_k^{(\ell)}$, and the other locally available information.

While some distributed algorithms require these steps to be performed synchronously (in parallel) for all users $k \in \{1, \ldots, K\}$, others also allow for asynchronous (or sequential) updates, where only one user k updates its power in one iteration ℓ. The updates can be continued, e. g., for a predefined number of iterations, or until a termination criterion, such as convergence of the powers, has been reached.

We note that this procedure of alternatingly updating the transmit strategies and exchanging information is not the only conceivable concept of a decentralized algorithm. An alternative as-

sumption for a distributed scenario is that there are no signaling links between the users, but that instead the transmitters and receivers repeatedly switch roles, as in a *time-divison duplex* system, and that the channel coefficients are identical in both the forward and the reverse direction. In the reverse phase, the transmitters can then measure, e. g., the "reverse SINR" or the received signal power and use this information for the next transmit strategy update, cf. [35, 50]. Nonetheless, such algorithms can in general also be employed in our uni-directional communications scenario; the same information that would be measured by the transmitter in the reverse phase is simply communicated over the feedback and signaling links in our model.

In the following we give an overview over several distributed strategies and subsequently compare their features, drawbacks, and the necessary information exchange.

3.5.1 The Selfish Solution

The selfish solution can be achieved without any information exchange between the users. The transmitters know that they must use $p_k = 1$ regardless of the channel coefficients. It is, however, necessary for receiver k to measure the SINR γ_k and to feed this knowledge back to transmitter k, so that the proper coding strategy can be chosen.

3.5.2 The Noiseless Balancing Algorithm

This algorithm, proposed by Zander in 1992 [51], and thus one of the pioneering works in distributed power control, aims at balancing the SINRs for the noiseless case $\sigma^2 = 0$. We recall from Section 3.4.4 that the goal of the balancing problem is to find a feasible SINR K-tuple that is a scaled version of a non-zero target SINR K-tuple $(\bar{\gamma}_1, \ldots, \bar{\gamma}_K)$, i. e., that can be expressed as $(\beta\bar{\gamma}_1, \ldots, \beta\bar{\gamma}_K)$, with a maximized scaling factor β. From (3.43) it can be seen that in the noiseless case the resulting power allocation must fulfill

$$\boldsymbol{C}\boldsymbol{p} = \frac{1}{\beta}\boldsymbol{p} \tag{3.68}$$

i. e., the power vector $\boldsymbol{p}$ is an eigenvector with non-negative entries corresponding to a positive eigenvalue of the matrix $\boldsymbol{C}$ defined in (3.41).

To ensure convergence of the noiseless balancing algorithm, we must assume that the matrix $\boldsymbol{C}$ is *irreducible.*

Definition 3.5. A square matrix $\boldsymbol{C}$ is *irreducible* if there does not exist a permutation matrix $\boldsymbol{P}$ such that

$$\boldsymbol{P}\boldsymbol{C}\boldsymbol{P}^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{0} & \boldsymbol{G} \end{bmatrix} \tag{3.69}$$

where $\boldsymbol{E}$ and $\boldsymbol{G}$ are square matrices and a permutation matrix is a binary square matrix that has exactly one 1 in each row and each column and zeros elsewhere.

In our system model, irreducibility of $\boldsymbol{C}$ has the meaning that it is not possible to partition the K users into two non-empty sets, where the transmitters of the first set do not cause any interference at the receivers of the second set.

The update of the noiseless balancing algorithm must be done by all users synchronously and consists of two steps. First, a new power is determined as

$$\hat{p}_k^{(\ell+1)} = \left(c + \frac{\bar{\gamma}_k}{\gamma_k^{(\ell)}} \right) p_k^{(\ell)} \tag{3.70}$$

where $c > 0$ is any constant common to all users. Second, the power is rescaled as

$$p_k^{(\ell+1)} = \delta^{(\ell)} \hat{p}_k^{(\ell+1)} \tag{3.71}$$

with a factor

$$\delta^{(\ell)} = \left(\max_k \ \hat{p}_k^{(\ell+1)} \right)^{-1} \tag{3.72}$$

that is common to all users and ensures that the power constraint is fulfilled. As $\delta^{(\ell)}$ depends on the powers $\hat{p}_k^{(\ell+1)}$ of all users $k \in \{1, \ldots, K\}$, the values of the powers $\hat{p}_k^{(\ell+1)}$ must be exchanged between the users after the first of the two steps of the iteration.

By inserting the definition of the SINR $\gamma_k^{(\ell)}$ into the update rule, we can express the updated power of user k by means of the matrix $\boldsymbol{C}$ and the vector of previous powers $\boldsymbol{p}^{(\ell)}$:

$$\begin{aligned} p_k^{(\ell+1)} &= \delta^{(\ell)} \left(c + \frac{\bar{\gamma}_k}{\gamma_k^{(\ell)}} \right) p_k^{(\ell)} \\ &= \delta^{(\ell)} \left(c p_k^{(\ell)} + \frac{\bar{\gamma}_k}{|h_{kk}|^2} \left(\sum_{j \neq k} |h_{kj}|^2 p_j^{(\ell)} \right) \right) = \delta^{(\ell)} \boldsymbol{e}_k^{\mathrm{T}} \left(c\mathbf{I} + \boldsymbol{C} \right) \boldsymbol{p}^{(\ell)}. \end{aligned} \tag{3.73}$$

Therefore, the vector of updated powers is

$$\boldsymbol{p}^{(\ell+1)} = \delta^{(\ell)} \left(c\mathbf{I} + \boldsymbol{C} \right) \boldsymbol{p}^{(\ell)} = \prod_{i=0}^{\ell} \delta^{(i)} \cdot \left(c\mathbf{I} + \boldsymbol{C} \right)^{\ell+1} \boldsymbol{p}^{(0)}. \tag{3.74}$$

With the eigenvalue decomposition[4] $\boldsymbol{C} = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^{-1}$ it follows that

$$\boldsymbol{p}^{(\ell+1)} = \prod_{i=0}^{\ell} \delta^{(i)} \cdot \boldsymbol{U} \left(c\mathbf{I} + \boldsymbol{\Lambda} \right)^{\ell+1} \boldsymbol{U}^{-1} \boldsymbol{p}^{(0)}. \tag{3.75}$$

It is intuitive that, as ℓ becomes large, the elements of the diagonal matrix $(c\mathbf{I} + \boldsymbol{\Lambda})^{\ell+1}$ with the highest absolute value grow fastest, to the extent that all other elements become negligible. From the Perron-Frobenius theorem [37] we know that for a non-negative, irreducible matrix $\boldsymbol{C}$ there are potentially multiple eigenvalues that have the highest absolute value, i. e., that achieve the spectral radius, but that exactly one of these is real and positive. Therefore the dominating diagonal element of $c\mathbf{I} + \boldsymbol{\Lambda}$ is unique. The Perron-Frobenius theorem also tells us that the corresponding eigenvector has only positive elements. Finally, assuming that the ith eigenvalue is dominant, $\boldsymbol{e}_i^{\mathrm{T}} \boldsymbol{U}^{-1} \boldsymbol{p}^{(0)}$ is non-zero for any positive initialization vector $\boldsymbol{p}^{(0)}$, as is shown in detail in [51]. Therefore, $\boldsymbol{U} \left(c\mathbf{I} + \boldsymbol{\Lambda} \right)^{\ell+1} \boldsymbol{U}^{-1} \boldsymbol{p}^{(0)}$ in the limit is an eigenvector of $\boldsymbol{C}$ with positive entries, i. e., as $\ell \to \infty$

$$\boldsymbol{C} \boldsymbol{p}^{(\ell)} = \frac{1}{\beta} \boldsymbol{p}^{(\ell)} \tag{3.76}$$

with $\beta > 0$, which is the noiseless balancing solution. The factor $\prod_{i=0}^{\ell} \delta^{(i)}$ ensures that after each iteration the power constraints are met.

[4] The argument can be extended to the case where $\boldsymbol{C}$ is not diagonalizable by using the Jordan canonical form [37] instead of the eigenvalue decomposition. The important point is that the Perron-Frobenius theorem tells us that the dominating diagonal entry of $c\mathbf{I} + \boldsymbol{\Lambda}$ is unique.

In [51], the algorithm is proposed with $c = 1$. The constant c could also be chosen differently without affecting the outcome of the algorithm. Both a very high value of c and a value close to zero can slow down convergence, however: in the former case the relative difference of the entries of $c\mathbf{I} + \boldsymbol{\Lambda}$ becomes small, while in the latter case the entries corresponding to the multiple eigenvalues of $\boldsymbol{C}$ that achieve the spectral radius have similar absolute values.

If noise is present, convergence towards (3.76) can be achieved by replacing the first part of the update by

$$\hat{p}_k^{(\ell+1)} = \left(c + \frac{\bar{\gamma}_k}{\gamma_k^{(\ell)}} \right) p_k^{(\ell)} - \frac{\bar{\gamma}_k \sigma^2}{|h_{kk}|^2} \tag{3.77}$$

to compensate for the additional noise term in (3.73). The outcome of the algorithm in this case, however, is still the *noiseless* SINR balancing solution, i. e., the solution to (3.68) instead of (3.43).

It should finally be noted that, if the matrix $\boldsymbol{C}$ is not irreducible, the algorithm is not guaranteed to converge to the balanced solution.

The information exchange of the noiseless balancing algorithm will be analyzed more closely in Section 3.5.6. We would like to point out, though, that the first part of the power update requires only the current value of the own SINR and no information from other users' links. Only the second step, in which the powers are scaled uniformly to meet the power constraints, raises the need for communication between the users.

3.5.3 The Feasibility/Power Minimization Algorithm

In 1993, Foschini and Miljanic [52] proposed a similar algorithm that aims at fulfilling a non-zero target SINR K-tuple $(\bar{\gamma}_1, \ldots, \bar{\gamma}_K)$ with equality and also accounts for the noise power σ^2, thus solving the feasibility/power minimization problem, cf. Section 3.4.3. However, the unit power constraints cannot be incorporated in this algorithm, i. e., the resulting powers may be larger than one. Again, for convergence it is necessary to assume irreducibility of the matrix $\boldsymbol{C}$.

The power update is similar to that of the noiseless balancing algorithm with $c = 0$, but consists only of one step and does not contain a scaling factor:

$$p_k^{(\ell+1)} = \frac{\bar{\gamma}_k}{\gamma_k^{(\ell)}} p_k^{(\ell)}. \tag{3.78}$$

The update is such that in every iteration the users change their power to achieve their target SINR with equality assuming that the other users' powers remain constant. Clearly, when all target SINRs have been reached, the update does not change the powers.

With the matrix and vectors defined in Section 3.4.4, we can express the updated power of user k in terms of the vector of previous powers

$$p_k^{(\ell+1)} = \frac{\bar{\gamma}_k}{|h_{kk}|^2} \left(\sum_{j \neq k} |h_{kj}|^2 p_j^{(\ell)} + \sigma^2 \right) = \boldsymbol{e}_k^{\mathrm{T}} \left(\boldsymbol{C}\boldsymbol{p}^{(\ell)} + \boldsymbol{d} \right). \tag{3.79}$$

Therefore, the vector of updated powers is

$$\boldsymbol{p}^{(\ell+1)} = \boldsymbol{C}\boldsymbol{p}^{(\ell)} + \boldsymbol{d} = \boldsymbol{C}^2 \boldsymbol{p}^{(\ell-1)} + \boldsymbol{C}\boldsymbol{d} + \boldsymbol{d} = \ldots = \boldsymbol{C}^{\ell+1} \boldsymbol{p}^{(0)} + \left(\mathbf{I} + \boldsymbol{C} + \boldsymbol{C}^2 + \ldots + \boldsymbol{C}^{\ell} \right) \boldsymbol{d}. \tag{3.80}$$

If the spectral radius of $\boldsymbol{C}$ is strictly smaller than one, $\boldsymbol{C}^{\ell+1}$ approaches the null matrix for large ℓ and we can apply the geometric series to the sum within the brackets, i. e.,

$$\lim_{\ell\to\infty} \boldsymbol{p}^{(\ell+1)} = (\mathbf{I}-\boldsymbol{C})^{-1}\,\boldsymbol{d} = (\mathbf{I}-\boldsymbol{C})^{-1} \cdot \begin{bmatrix} \frac{\bar{\gamma}_1}{|h_{11}|^2} & 0 & \cdots & 0 \\ 0 & \frac{\bar{\gamma}_2}{|h_{22}|^2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \frac{\bar{\gamma}_K}{|h_{KK}|^2} \end{bmatrix} \cdot \sigma^2\mathbf{1} = \sigma^2\boldsymbol{A}^{-1}\mathbf{1} \quad (3.81)$$

which is, with (3.37) and (3.38), the power allocation that achieves the SINR targets $(\bar{\gamma}_1,\ldots,\bar{\gamma}_K)$ and therefore the solution to the feasibility/power minimization problem without power constraints.

If the spectral radius of $\boldsymbol{C}$ is one or larger, however, the power allocation does not converge and one or more powers grow without bound. It is shown in [52] that if the SINR targets are feasible, the spectral radius of $\boldsymbol{C}$ is always below one. For the rigorous proof we refer to [52] and give an intuitive argument for this result in the following: if the SINR targets are feasible with noise power σ^2, they are also feasible with noise power $\alpha\sigma^2$ for $0<\alpha<1$; furthermore, when the noise power is decreased in this way, the target SINR K-tuple is surely not on the Pareto boundary, as it can be met with equality by multiplying all powers with the factor α. In the case of asymptotically low noise, i. e., for $\alpha\to 0$,

$$\boldsymbol{C}\boldsymbol{p} = \frac{1}{\beta}\boldsymbol{p} \quad (3.82)$$

defines the balancing solution for the target K-tuple $(\bar{\gamma}_1,\ldots,\bar{\gamma}_K)$, as was shown in the previous section. Since the SINR targets $(\bar{\gamma}_1,\ldots,\bar{\gamma}_K)$ are feasible and not on the Pareto boundary for $\alpha\to 0$ and $(\beta\bar{\gamma}_1,\ldots,\beta\bar{\gamma}_K)$ by definition is on the Pareto boundary, the resulting β must be larger than one. Finally, the Perron-Frobenius theorem tells us that $1/\beta$ is the spectral radius of $\boldsymbol{C}$.

The feasibility/power minimization algorithm does not require any information exchange between the users: only the current value of the own SINR is required for the power update. Also, the updates need not be performed in parallel (even though this was assumed for the convergence proof), i. e., the target SINRs are also reached if only some users update their powers in a given iteration. The powers that fulfill the target SINRs can be arbitrary large, however. Furthermore, if it turns out that the target SINRs are infeasible even without the power constraints, the powers will continue to grow without bound with every iteration, which is highly impractical.

As a remedy, the power update could be modified so that the power is "capped" at one, i. e.,

$$p_k^{(\ell+1)} = \min\left\{\frac{\bar{\gamma}_k}{\gamma_k^{(\ell)}}p_k^{(\ell)},\,1\right\}. \quad (3.83)$$

As a result, the power allocation would fulfill the power constraints after each update and still could converge to the SINR targets if they are feasible. If the SINR targets are infeasible, the algorithm would converge to some power allocation on the Pareto boundary.

3.5.4 The Distributed Gradient Projection Algorithm

In [35], an algorithm is proposed that optimizes the sum utility by means of a projected gradient search. Clearly, when the problem (3.47) has more than one local optimum, a gradient-based algorithm does not necessarily converge to the global optimum, but might, depending on the initialization and the step size, end up in another local optimum and thus not find the solution to (3.47).

Nonetheless, a good local optimum will yield superior performance to most other power allocations; also, determining a local optimum with manageable computational effort may be the only reasonable option when searching for the global optimum is computationally infeasible.

In a centralized projected gradient algorithm, a feasible initial power allocation $\boldsymbol{p}^{(0)}$ is chosen and the power vector is then iteratively updated according to

$$\boldsymbol{p}^{(\ell+1)} = \Pi\left(\boldsymbol{p}^{(\ell)} + \delta \cdot \frac{\partial}{\partial \boldsymbol{p}} \sum_j u_j(\gamma_j)\bigg|_{\boldsymbol{p}=\boldsymbol{p}^{(\ell)}}\right) \tag{3.84}$$

where δ is the *step size* and $\Pi(\cdot)$ is the *projection.* The step size δ is crucial for convergence of the algorithm: it can either be a constant or vary in the course of the iterations; if it changes over the iterations, the values of δ may follow a pre-defined sequence or be determined adaptively [46]. Too small steps in general mean that convergence takes very long, while too large steps can cause convergence to fail altogether. For the following distributed gradient projection algorithm it is assumed that δ is a constant.

The projection $\Pi(\cdot)$ ensures that the new power allocation fulfills the power constraints. It returns the feasible power vector that is closest in terms of Euclidean norm to the argument of the projection [35]. In our case, the projection is a very straightforward operation: every element of the argument that is greater than one is set to one, every negative element is set to zero, and the others are not changed.

From the point of view of user k the power update therefore is

$$\hat{p}_k^{(\ell+1)} = p_k^{(\ell)} + \delta \cdot \frac{\partial}{\partial p_k} \sum_j u_j(\gamma_j)\bigg|_{\boldsymbol{p}=\boldsymbol{p}^{(\ell)}} \tag{3.85}$$

$$p_k^{(\ell+1)} = \min\left\{\max\left\{\hat{p}_k^{(\ell+1)}, 0\right\}, 1\right\}. \tag{3.86}$$

Making use of the chain rule, we express the derivative as

$$\begin{aligned}\frac{\partial}{\partial p_k} \sum_j u_j(\gamma_j)\bigg|_{\boldsymbol{p}=\boldsymbol{p}^{(\ell)}} &= u_k'\left(\gamma_k^{(\ell)}\right) \frac{|h_{kk}|^2}{\sum_{i\neq k}|h_{ki}|^2 p_i^{(\ell)} + \sigma^2} \\ &\quad - \sum_{j\neq k} u_j'\left(\gamma_j^{(\ell)}\right) \frac{|h_{jj}|^2 p_j^{(\ell)}}{\left(\sum_{i\neq j}|h_{ji}|^2 p_i^{(\ell)} + \sigma^2\right)^2} |h_{jk}|^2 \\ &= u_k'\left(\gamma_k^{(\ell)}\right) \frac{\gamma_k^{(\ell)}}{p_k^{(\ell)}} - \sum_{j\neq k} \pi_j^{(\ell)} |h_{jk}|^2\end{aligned} \tag{3.87}$$

where we defined

$$\begin{aligned}\pi_j^{(\ell)} &= u_j'\left(\gamma_j^{(\ell)}\right) \frac{|h_{jj}|^2 p_j^{(\ell)}}{\left(\sum_{i\neq j}|h_{ji}|^2 p_i^{(\ell)} + \sigma^2\right)^2} = u_j'\left(\gamma_j^{(\ell)}\right) \frac{{\gamma_j^{(\ell)}}^2}{|h_{jj}|^2 p_j^{(\ell)}} \\ &= -\frac{\partial u_j(\gamma_j)}{\partial\left(|h_{jk}|^2 p_k\right)}\bigg|_{\boldsymbol{p}=\boldsymbol{p}^{(\ell)}} \qquad \text{for any } k \neq j.\end{aligned} \tag{3.88}$$

As can be seen from the second line, the quantity $\pi_j^{(\ell)}$ can also be interpreted as the negative derivative of the utility of user j w. r. t. the interference power received from *any* undesired transmitter,

i. e., $\pi_j^{(\ell)}$ is the marginal decrease in the utility of user j when its interference power is marginally increased. In the interference pricing algorithm presented in the next section, the quantity $\pi_j^{(\ell)}$ will have special significance and will be termed the *interference price* of user j.

We can now write the power update of user k (prior to the projection) as

$$\hat{p}_k^{(\ell+1)} = p_k^{(\ell)} + \delta \left(u_k' \left(\gamma_k^{(\ell)} \right) \frac{\gamma_k^{(\ell)}}{p_k^{(\ell)}} - \sum_{j \neq k} \pi_j^{(\ell)} |h_{jk}|^2 \right) \tag{3.89}$$

and observe that in addition to the current value of the own SINR and own power, knowledge of the cross channel magnitudes $|h_{jk}|^2$ and the current values of $\pi_j^{(\ell)}$ for all $j \neq k$ is required, which must be obtained by exchanging information with the other users. The cross channel coefficients only need to be exchanged once during the initialization of the algorithm; the prices $\pi_j^{(\ell)}$, however, change with each iteration and must therefore be exchanged after each iteration. We furthermore note that all users must update their power synchronously, so that the updated power vector can be expressed as (3.84).

For the purpose of clarifying which information is necessary for the update, we have expressed all terms as functions of the SINRs $\gamma_k^{(\ell)}$ as far as possible. If the power of user k after iteration ℓ is zero, however, the SINR is also zero and the quantity $\gamma_k^{(\ell)}/p_k^{(\ell)}$ cannot be evaluated. Instead, the expression $|h_{kk}|^2 / \left(\sum_{j \neq k} |h_{kj}|^2 p_j^{(\ell)} + \sigma^2 \right)$ must be used, which we assume can just as easily be measured by receiver k as the SINR $\gamma_k^{(\ell)}$.

Convergence of the distributed gradient projection algorithm can be guaranteed if the step size δ is sufficiently small. As noted previously, if δ is too small, convergence may require very many iterations. It is possible to determine an optimal step size, but this requires bounding the norm of the Hessian matrix of the sum utility [35, 46], and is therefore not practical if the nodes have limited processing power. Therefore, heuristics are needed for a good choice of the step size δ, which is an open problem that is not further pursued here.

We finally note that in [35] a procedure of switching the roles of transmitters and receivers is proposed by which the transmitters can estimate $\sum_j \pi_j^{(\ell)} |h_{jk}|^2$ without communicating with the other users over signaling links. Also, a very similar algorithm can be derived by operating on the logarithmic powers t_k instead of p_k.

3.5.5 The Interference Pricing Algorithm

As an extension to the selfish solution, where each user k maximizes its own utility $u_k(\gamma_k)$, the term "pricing" has been generally applied to schemes where each user k maximizes its own utility minus a cost term that is linear in the own power, i. e., $u_k(\gamma_k) - \pi \cdot p_k$, cf. [53, 54]. The additional cost term can be seen as a heuristic to incorporate the negative effects that an increase of p_k has on other users' utilities into the payoff function of user k, or as a modification of the utility function to encourage more "social" behavior of the users, while maintaining the principle that each user is responsible only for maximizing its own payoff. A good choice of values for the prices has been regarded generally as a matter of heuristics and the prices were usually assumed to be constant over iterations and/or users.

In [2], it was shown that by properly adapting the prices in each iteration, the pricing approach can be used to find a solution candidate for the sum utility problem. In each iteration of the *inter-*

ference pricing algorithm user k solves its own power control problem

$$p_k^{(\ell+1)} = \arg\max_{p_k} \; u_k(\gamma_k)\Big|_{p_j=p_j^{(\ell)}\;\forall j\neq k} - \sum_{j\neq k} \pi_j^{(\ell)} |h_{jk}|^2 p_k \qquad \text{s.t.:} \quad 0 \leq p_k \leq 1 \tag{3.90}$$

with the *interference prices* $\pi_j^{(\ell)}$ as defined in (3.88). This problem is an inexpensive line search and can even be solved in closed form for many utility functions. Furthermore, if $u_k(\gamma_k)$ is concave in γ_k (and consequently in p_k as well), (3.90) is a concave maximization problem, and thus does not have multiple local optima.

In the following we show that if a power allocation is stationary in the interference pricing algorithm, i. e., if $p_k^{(\ell+1)} = p_k^{(\ell)}$ for all $k \in \{1, \ldots, K\}$, it fulfills the KKT conditions of the sum utility problem. To begin with, note that stationarity of the pricing algorithm implies that $p_k^{(\ell)}$ fulfills the KKT conditions of the problem (3.90). With the Lagrangian multipliers μ_k and ν_k for the power constraints, the KKT conditons of (3.90) are

$$\frac{\partial}{\partial p_k} u_k(\gamma_k)\Big|_{\boldsymbol{p}=\boldsymbol{p}^{(\ell)}} - \sum_{j\neq k} \pi_j^{(\ell)} |h_{jk}|^2 + \mu_k - \nu_k = 0 \tag{3.91}$$

$$\mu_k \geq 0 \tag{3.92}$$

$$\nu_k \geq 0 \tag{3.93}$$

$$\mu_k p_k^{(\ell)} = 0 \tag{3.94}$$

$$\nu_k \left(1 - p_k^{(\ell)}\right) = 0 \tag{3.95}$$

$$0 \leq p_k^{(\ell)} \leq 1. \tag{3.96}$$

Considering that by definition (3.88)

$$\frac{\partial}{\partial p_k} u_j(\gamma_j)\Big|_{\boldsymbol{p}=\boldsymbol{p}^{(\ell)}} = -\pi_j^{(\ell)} |h_{jk}|^2 \tag{3.97}$$

and that the stationarity conditions must be fulfilled for all users, it is clear that the stationary power allocation $\boldsymbol{p}^{(\ell)}$ fulfills the conditions (3.48)–(3.53) and that any power allocation fulfilling the KKT conditions of the sum utility problem is a stationary power allocation of the interference pricing algorithm.

In contrast to the previously discussed distributed projected gradient algorithm, the pricing updates need not necessarily be performed synchronously. The result that a stationary point is a solution candidate of the sum utility problem holds regardless of the order in which the users have performed their updates, or even whether they have at some point used outdated information in their updates. Whether the algorithm is guaranteed to converge to such a stationary point, is a different question. In [2], it is proven using results from the theory of *supermodular games* that, if the utility functions $u_k(\gamma_k)$ are such that $c_k(\gamma_k) \in [1, 2]$ for all feasible SINR K-tuples and all $k \in \{1, \ldots, K\}$, the interference pricing algorithm converges to the global optimum. Again, it is neither required that the users perform their updates in parallel or in any defined order, nor that they always use the most up-to-date interference prices; it is only necessary that all powers and interference prices are updated "regularly". Note, however, that while for the logarithmic utility convergence is thereby guaranteed, the rate utility does not fall into this class of utility functions, as for the rate utility $c_k(\gamma_k) < 1$.

In the following we sketch an alternative convergence argument that is given in [55]; this argument applies to a somewhat broader set of utility functions that also includes the rate utility. The conditions that every utility function $u_k(\gamma_k)$ must fulfill for the convergence argument are:

- $u_k(\gamma_k)$ is concave in p_k, i. e., $u_k''(\gamma_k) \leq 0$ and thus $c_k(\gamma_k) \geq 0$ for all feasible SINRs.
- $u_k(\gamma_k)$ is convex in p_j for all $j \neq k$, i. e., $\partial^2 u_k(\gamma_k)/(\partial p_j)^2 \geq 0$ for all $j \neq k$. With

$$\frac{\partial^2 u_k(\gamma_k)}{(\partial p_j)^2} = u_k''(\gamma_k) \left(\frac{\partial \gamma_k}{\partial p_j}\right)^2 + u_k'(\gamma_k) \frac{\partial^2 \gamma_k}{(\partial p_j)^2} \tag{3.98}$$

  and

$$\frac{\partial \gamma_k}{\partial p_j} = -\frac{\gamma_k^2 |h_{kj}|^2}{|h_{kk}|^2 p_k} \qquad \text{and} \qquad \frac{\partial^2 \gamma_k}{(\partial p_j)^2} = \frac{2\gamma_k^3 |h_{kj}|^4}{|h_{kk}|^4 p_k^2} \tag{3.99}$$

  for $j \neq k$, it can be seen that this condition is equivalent to $c_k(\gamma_k) \leq 2$ for all feasible SINRs.
- The magnitude of the second derivative of the utilitiy does not approach zero anywhere in the interval of feasible SINRs, i. e., $|u_k''(\gamma_k)| \geq C > 0$ for all feasible SINRs.

Thus, we are essentially allowing utility functions with $c_k(\gamma_k) \in [0, 2]$ with the small caveat that the second derivative may not vanish anywhere. The rate utility fulfills these conditions, as the second derivative approaches zero only for inifinite SINR, which is not feasible.

For the following convergence argument, we assume that in iteration ℓ only user k updates its power, and that the interference prices are up-to-date. By adding some terms that do not depend on p_k to the utility function, the power update of the interference pricing algorithm can be written as

$$p_k^{(\ell+1)} = \arg\max_{p_k} \; a_k^{(\ell)}(p_k) \qquad \text{s. t.:} \quad 0 \leq p_k \leq 1 \tag{3.100}$$

with

$$a_k^{(\ell)}(p_k) = u_k(\gamma_k)\Big|_{p_j = p_j^{(\ell)} \; \forall j \neq k} + \sum_{j \neq k} \left(u_j\left(\gamma_j^{(\ell)}\right) - \pi_j^{(\ell)} |h_{jk}|^2 \left(p_k - p_k^{(\ell)}\right)\right). \tag{3.101}$$

The function $a_k^{(\ell)}(p_k)$ is an approximation of the sum utility, where the utility summands of users $j \neq k$ were replaced by their first order Taylor approximation centered at the current power $p_k^{(\ell)}$. Therefore, we have

$$a_k^{(\ell)}\left(p_k^{(\ell)}\right) = \sum_j u_j\left(\gamma_j^{(\ell)}\right) \tag{3.102}$$

i. e., at for $p_k = p_k^{(\ell)}$ the approximation is exact. Since the summands $u_j(\gamma_j)$ with $j \neq k$ are convex in p_k, linearization of $u_j(\gamma_j)$ in p_k leads to an under-estimation of the true sum utility, i. e.,

$$a_k^{(\ell)}(p_k) \leq \sum_j u_j(\gamma_j)\Big|_{p_i = p_i^{(\ell)} \; \forall i \neq k} \tag{3.103}$$

and we can state our chain of arguments for convergence:

- Since the update consists of maximizing $a_k^{(\ell)}(p_k)$,

$$a_k^{(\ell)}\left(p_k^{(\ell+1)}\right) \geq a_k^{(\ell)}\left(p_k^{(\ell)}\right) \tag{3.104}$$

  and therefore the value of the sum utility is non-decreasing over the iterations.

- The sum utility is bounded from above, therefore it converges over the iterations, i. e., the changes in sum utility from one iteration to the next become arbitrarily small.
- Since the sum utility is under-estimated by the approximation $a_k^{(\ell)}(p_k)$, the change in the approximation resulting from updating the power also becomes arbitrarily small, i. e., for any $\varepsilon > 0$ there exists an ℓ_0 such that

$$a_k^{(\ell)}\left(p_k^{(\ell+1)}\right) - a_k^{(\ell)}\left(p_k^{(\ell)}\right) \leq \varepsilon \qquad \forall \ell \geq \ell_0. \tag{3.105}$$

- An arbitrarily small change in $a_k^{(\ell)}(p_k)$ implies an arbitrarily small change in power p_k from iteration ℓ to iteration $\ell + 1$ since the magnitude of the second derivative of the utilities is bounded from below. The rigorous proof for this part of the argument is rather tedious and not very instructive, and is therefore omitted here. Intuitively it is clear that an arbitrarily small change of $a_k^{(\ell)}(p_k)$ can only be accompanied by a much larger change in powers if the function $a_k^{(\ell)}(p_k)$ is arbitrarily "flat" around the maximum, which would require the second derivative of $a_k^{(\ell)}(p_k)$ to vanish, which in turn would require that $u_k''(\gamma_k)$ approaches zero.

Therefore, all powers converge and the power allocation reaches a stationary point.

Similar to the distributed gradient projection algorithm, the interference pricing algorithm requires that the squared channel magnitudes $|h_{jk}|^2$ for all $j \in \{1, \ldots, K\}$ are known to user k, which necessitates an initial channel information exchange phase. Then, in order to perform a power update, the current interference prices $\pi_j^{(\ell)}$ for all $j \neq k$ must be known to user k. A detailed analysis of the necessary amount of information exchange is given in Section 3.5.6.

3.5.5.1 Update Rules for the Sum Rate Problem

As discussed in Section 3.4.5.1, the derivatives of the rate of user k w. r. t. the different powers are

$$\frac{\partial R_k}{\partial p_k} = \frac{\gamma_k}{(1+\gamma_k)p_k} = \frac{1}{\frac{p_k}{\gamma_k} + p_k} \tag{3.106}$$

$$\frac{\partial R_k}{\partial p_j} = -\frac{\gamma_k^2 |h_{kj}|^2}{(1+\gamma_k)|h_{kk}|^2 p_k} \qquad \text{for} \quad j \neq k \tag{3.107}$$

where the second line can also be obtained by applying the chain rule and using (3.99). The interference price of user k can thus be computed as

$$\pi_k^{(\ell)} = \frac{{\gamma_k^{(\ell)}}^2}{\left(1+\gamma_k^{(\ell)}\right)|h_{kk}|^2 p_k^{(\ell)}}. \tag{3.108}$$

For the power update, we first set the derivative of the utility minus cost to zero:

$$\frac{\partial R_k}{\partial p_k} = \sum_{j \neq k} \pi_j^{(\ell)} |h_{jk}|^2. \tag{3.109}$$

Keeping in mind that p_k/γ_k does not depend on p_k, the power $\hat{p}_k^{(\ell+1)}$ that fulfills this condition is

$$\hat{p}_k^{(\ell+1)} = \frac{1}{\sum_{j \neq k} \pi_j^{(\ell)} |h_{jk}|^2} - \frac{p_k^{(\ell)}}{\gamma_k^{(\ell)}}. \tag{3.110}$$

If $\hat{p}_k^{(\ell+1)}$ is negative (or larger than one), the solution to (3.90) is zero (or one) since R_k, and therefore also R_k minus the linear cost terms, is concave in p_k. The power update can thus be completed by the projection

$$p_k^{(\ell+1)} = \min\left\{\max\left\{\hat{p}_k^{(\ell+1)}, 0\right\}, 1\right\}. \tag{3.111}$$

Special attention must be paid to two cases where the evaluation of (3.110) is problematic: first, if $p_k^{(\ell)}$ is zero, we must use the expression $\left(\sum_{j\neq k}|h_{kj}|^2 p_j^{(\ell)} + \sigma^2\right)/|h_{kk}|^2$ instead of $p_k^{(\ell)}/\gamma_k^{(\ell)}$, as discussed below (3.89). Second, if $\sum_{j\neq k}\pi_j^{(\ell)}|h_{jk}|^2$ is zero, the right-hand side of (3.110) is infinite and it is clear from the optimization problem (3.90) that $p_k^{(\ell+1)}$ is one.

3.5.5.2 Update Rules for the Sum Logarithmic Utility Problem

Similar to the previous section, with the derivatives

$$\frac{\partial u_k(\gamma_k)}{\partial p_k} = \frac{1}{p_k} \tag{3.112}$$

$$\frac{\partial u_k(\gamma_k)}{\partial p_j} = -\frac{\gamma_k|h_{kj}|^2}{|h_{kk}|^2 p_k} \qquad \text{for} \quad j \neq k \tag{3.113}$$

we can compute the interference prices as

$$\pi_k^{(\ell)} = \frac{\gamma_k^{(\ell)}}{|h_{kk}|^2 p_k^{(\ell)}} \tag{3.114}$$

and perform the power update as

$$\hat{p}_k^{(\ell+1)} = \frac{1}{\sum_{j\neq k}\pi_j^{(\ell)}|h_{jk}|^2} \tag{3.115}$$

$$p_k^{(\ell+1)} = \min\left\{\hat{p}_k^{(\ell+1)}, 1\right\}. \tag{3.116}$$

3.5.6 Comparison of the Distributed Algorithms

Table 3.1 gives an overview of some of the main properties of the discussed distributed algorithms. For the information exchange analysis in Table 3.2, we assumed that the transmitters are connected via bi-directional signaling links, as is depicted in Figure 3.6(a). Also, each transmitter k is able to evaluate its utility function $u_k(\gamma_k)$ and the first derivative $u_k'(\gamma_k)$, but need not have any knowledge of the other users' utility functions. The first column shows which information must be fed back from every receiver to its transmitter over the feedback link once at the beginning of the algorithm; the information in the second column is signaled from every transmitter k to every transmitter $j \neq k$ once. The third and fourth column contain the information that is fed back from receivers to transmitters and signaled among the transmitters at every iteration.

If the receivers are connected and perform the computation of the power update, as in Figure 3.6(b), the information exchange requirements are very similar. The main difference is that the receivers must additionally feed back the powers $p_k^{(\ell)}$ to their transmitters after each update.

Algorithm	Requires synchronous updates?	Issues
Selfish solution	No	—
Noiseless balancing	Yes	balancing solution for $\sigma^2 = 0$; convergence requires irreducibility
Feasibility/power min.	No	does not allow for upper limits on power; diverges if SINR targets are infeasible; convergence requires irreducibility
Projected gradient	Yes	choice of a good step size is not straightforward
Interference pricing	No	no convergence guarantees for $c_k(\gamma_k) > 2$

Table 3.1: Overview of the discussed distributed algorithms

Algorithm	Rx $k \to$ Tx k once	Tx $k \to$ Tx j once	Rx $k \to$ Tx k per iteration	Tx $k \to$ Tx j per iteration
Selfish solution	γ_k	—	—	—
Noiseless balancing	$\lvert h_{kk}\rvert^2$	—	$\gamma_k^{(\ell)}$	$\hat{p}_k^{(\ell)}$
Feasibility/power min.	—	—	$\gamma_k^{(\ell)}$	—
Projected gradient	$\lvert h_{kj}\rvert^2\ \forall j$	$\lvert h_{kj}\rvert^2$	$\gamma_k^{(\ell)}$	$\pi_k^{(\ell)}$
Interference pricing	$\lvert h_{kj}\rvert^2\ \forall j$	$\lvert h_{kj}\rvert^2$	$\gamma_k^{(\ell)}$	$\pi_k^{(\ell)}$

Table 3.2: Information exchange requirements of the distributed algorithms assuming that the transmitters compute the updates and are connected via signaling links, cf. Figure 3.6(a).

3.6 Numerical Evaluation of the Distributed Power Control Algorithms

In this section we give some examples demonstrating the typical convergence behavior of the discussed algorithms and then proceed to evaluate the average performance of the projected gradient and interference pricing algorithms in terms of sum utility for a simple independent and identically distributed (i. i. d.) Gaussian channel model. We begin with the special case of $K = 2$ users where the trajectories of the discussed algorithms can be visualized in the SINR and rate regions.

3.6.1 Numerical Examples for $K = 2$ Users

In Figures 3.7–3.9, a typical two-user SINR region is plotted for three different values of the noise power σ^2. When the noise power decreases, the shape of the region becomes "more hyperbolic", and the SINR of the active user in an altruistic solution becomes much higher than the SINRs achievable by the selfish solution.

The balancing target in all three scenarios is such that the first user has twice the SINR of the second user; the balancing solution, i. e., the point on the Pareto boundary that fulfills the balancing target, is marked with a circle. The noiseless balancing algorithm, however, converges towards the balancing solution assuming that $\sigma^2 = 0$, which we mark with a triangle in the plots. For low values of σ^2 the noiseless balancing solution and the balancing solution almost coincide, while

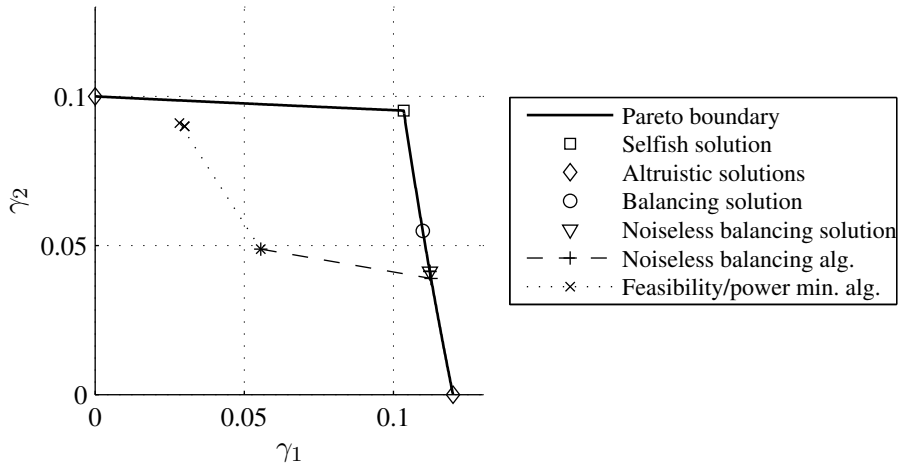


Figure 3.7: SINR region for the same channel as in Figure 3.4, but with $\sigma^2 = 10$. The algorithms are initialized with the power allocation $(p_1 = 0.5, p_2 = 0.5)$, the SINR target of the feasibility algorithm is $(0.03, 0.09)$, the balancing target is $(2, 1)$.

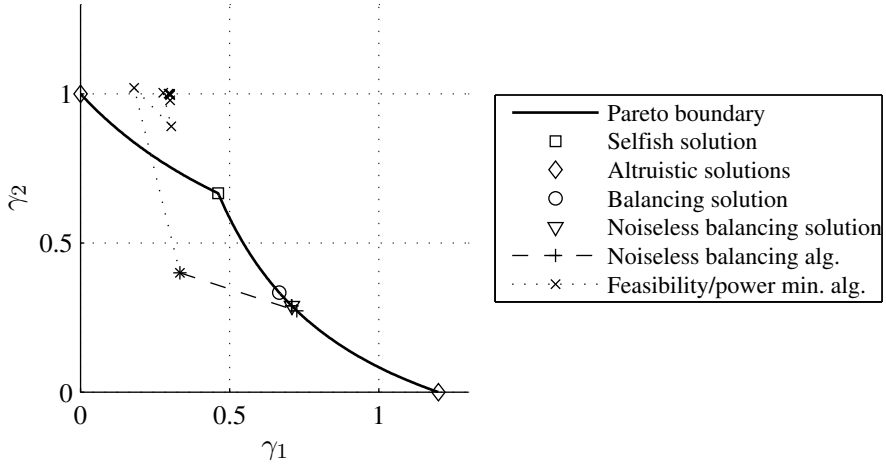


Figure 3.8: Scenario as in Figure 3.7 with $\sigma^2 = 1$ and feasibility target $(0.3, 1)$.

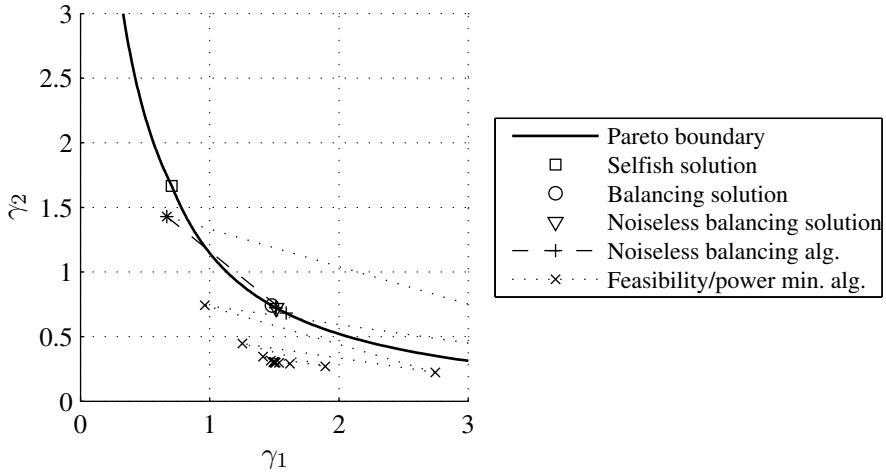


Figure 3.9: Scenario as in Figure 3.7 with $\sigma^2 = 0.1$ and feasibility target $(1.5, 0.3)$.

for higher noise powers there is a significant gap. In the plotted scenarios, the noiseless balancing algorithm reaches its goal in very few iterations.

The convergence performance of the feasibility algorithm, which was implemented with parallel updates, depends more heavily on the value of the noise power. In Figure 3.7, the target is reached in two iterations, while in Figure 3.9, where $\sigma^2 = 0.1$, more than ten iterations are needed. Note that in Figure 3.8 an infeasible SINR target was chosen. The algorithm converges towards it, nonetheless, thereby violating the transmit power constraints. It is also possible that an SINR target is infeasible even with infinite transmit power; in such a case (which is not included in our examples) the algorithm would simply diverge.

In Figures 3.10–3.12, the same scenarios are plotted in the rate domain instead of the SINR domain. Again it can be seen that the non-convexity of the region is more pronounced for lower noise powers. As discussed in Section 3.4.5.1, the sum-rate optimal rate pair can be either the selfish solution or one of the altruistic solutions. Since the sum rate is equal for all points on a straight line with slope -1, the optimal rate pair can be graphically determined as the point where the region is "touched" by such a line of equal sum rate. In the figures the sum-rate optimal point is marked by a circle and the supporting line of equal sum rate is visualized as a dotted line.

The pricing algorithm was implemented with parallel updates, i. e., both users update their powers synchronously and then exchange the updated prices. In Figures 3.10 and 3.11, the pricing algorithm converges to the optimal selfish selfish solution immediately. With lower noise power in the scenario in Figure 3.12, on the other hand, several iterations are needed to disable the second user. This scenario is also an example for a case in which the projected gradient and the pricing algorithm happen to converge to two different local optima. In this particular case the pricing algorithm finds the global optimum, but this does not appear to be a general trend.

The trajectories of the projected gradient algorithm in Figures 3.10 and 3.11 also give an intuition for the difficulty of choosing the best step size. While in the case $\sigma^2 = 10$ the algorithm seems to take unnecessarily small steps, for $\sigma^2 = 1$ the optimum is found with a single iteration, even though $\delta = 1$ was used in both cases.

In Figure 3.13, the achievable region for the same channel with $\sigma^2 = 1$ can again be seen, but this time in the logarithmic utility domain. The utility region is clearly convex, and, as discussed in Section 3.4.5.2, the selfish soltution is optimal in terms of sum utility.

3.6.2 Numerical Examples for the Convergence Behavior for $K = 4$ Users

In Figures 3.14 and 3.15, we show how the sum rate evolves over the iterations for a given channel with $K = 4$ users. The coefficients of the channel are

$$\begin{array}{llll} |h_{11}|^2 = 0.8 & |h_{12}|^2 = 0.8 & |h_{13}|^2 = 2.5 & |h_{14}|^2 = 2.2 \\ |h_{21}|^2 = 0.2 & |h_{22}|^2 = 0.8 & |h_{23}|^2 = 0.4 & |h_{24}|^2 = 0.5 \\ |h_{31}|^2 = 0.4 & |h_{32}|^2 = 0.4 & |h_{33}|^2 = 0.4 & |h_{34}|^2 = 2.6 \\ |h_{41}|^2 = 2.6 & |h_{42}|^2 = 1.8 & |h_{43}|^2 = 0.4 & |h_{44}|^2 = 0.2. \end{array} \tag{3.117}$$

Again, parallel updates were used for the interference pricing algorithm. All algorithms were terminated when the Euclidean norm of the difference of the updated vector of powers $\boldsymbol{p}^{(\ell+1)}$ and the previous power vector $\boldsymbol{p}^{(\ell)}$ was below 10^{-4}.

The figures clearly show that choosing a small step size δ for the projected gradient algorithm leads to slower convergence. Also, the scenario in Figure 3.14 is an example where the step size has an influence on which local optimum is reached: with $\delta = 10$ (and with the interference pricing

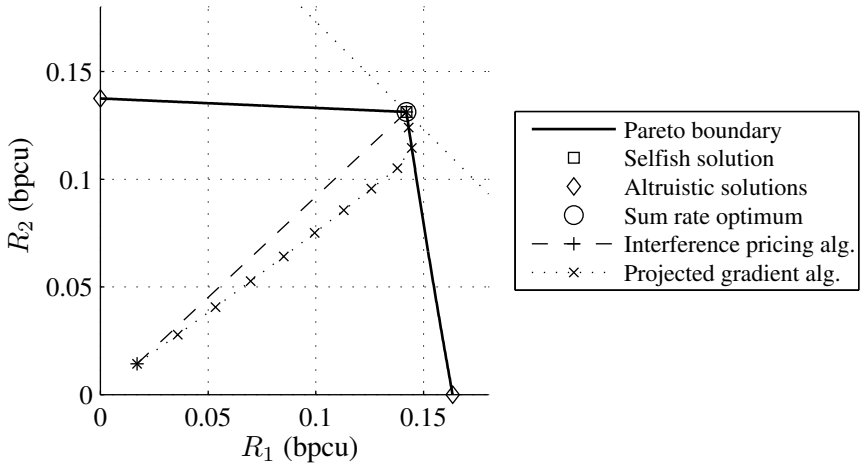


Figure 3.10: Rate region for the scenario in Figure 3.7. The algorithms are initialized with the power allocation $(p_1 = 0.1, p_2 = 0.1)$, the step size of the projected gradient algorithm is $\delta = 1$. All rate pairs on the dotted line have the optimal sum rate.

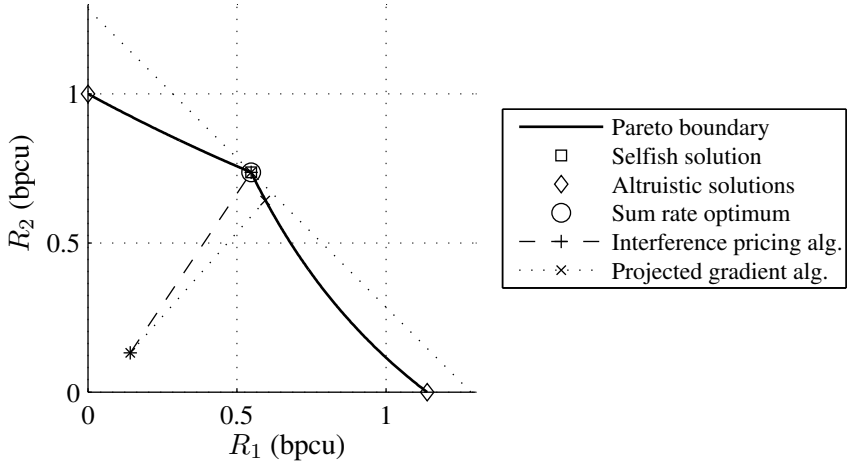


Figure 3.11: Scenario as in Figure 3.10 with $\sigma^2 = 1$.

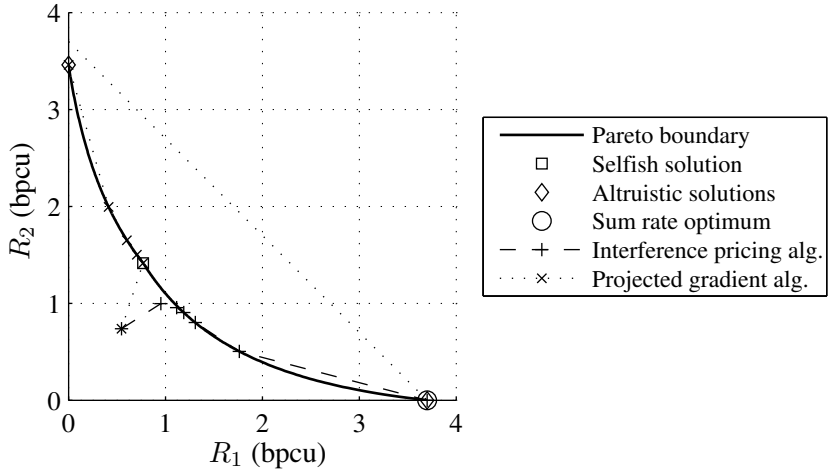


Figure 3.12: Scenario as in Figure 3.10 with $\sigma^2 = 0.1$.

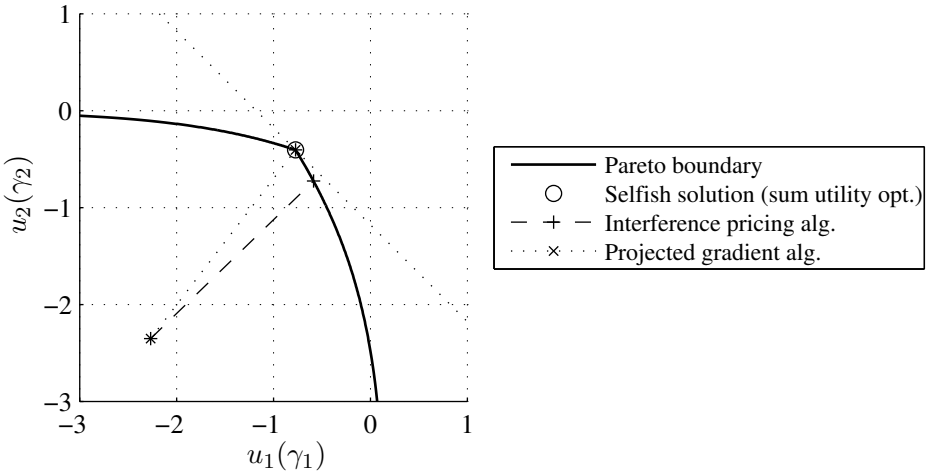


Figure 3.13: Scenario as in Figure 3.11 with logarithmic utility instead of rate.

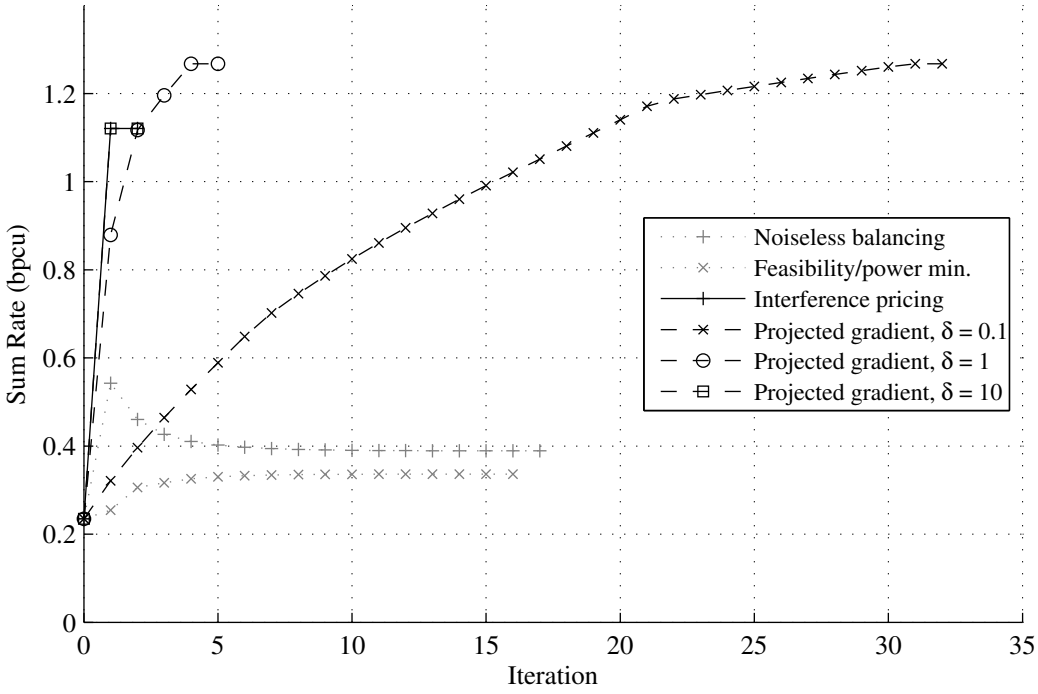


Figure 3.14: Convergence of the sum rate over iterations for the channel in (3.117) with $\sigma^2 = 1$. The initial power allocation for all algorithms is $p_k = 0.1$ for all $k \in \{1, 2, 3, 4\}$. The SINR target for both the feasibility and balancing algorithms is $\bar{\gamma}_k = 0.06$ for all $k \in \{1, 2, 3, 4\}$.

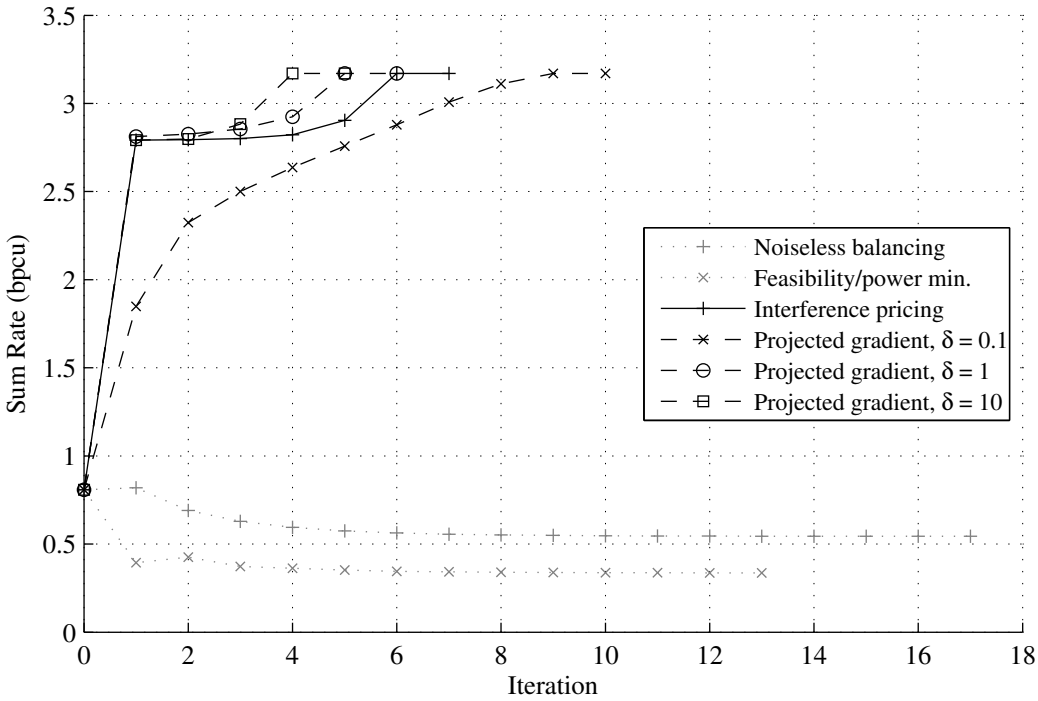


Figure 3.15: Scenario as in Figure 3.14 with $\sigma^2 = 0.1$.

algorithm) an inferior local optimum is found compared to the lower step sizes. Again, though, it cannot be generally stated that small step sizes yield better performance.

For the same scenario, but with logarithmic instead of rate utility, Figure 3.16 shows that the step sizes $\delta = 1$ and $\delta = 10$ lead to visible oscillations and are therefore not sufficiently small to ensure convergence, whereas the pricing algorithm converges without problems.

3.6.3 Average Sum Utility Performance in a Gaussian Channel Model

While we have so far only presented exemplary channel realizations, it is of course of interest how the algorithms aimed at maximizing the sum utility perform on average. To this end, we must first define a channel model: in the following we assume that all channel coefficients h_{kj} are chosen independently with a zero-mean unit-variance complex Gaussian distribution, i. e., the real and imaginary parts of all h_{kj} are Gaussian with mean zero and variance one half, and are uncorrelated.

For the results in Figures 3.17–3.20 the initial power allocation was $p_k = 0$ for all $k \in \{1, \ldots, K\}$. The iterative algorithms were run until either the Euclidean norm of the change in the power vectors was below 10^{-4} or 1000 iterations were reached. In addition to the distributed algorithms, we included the performance of two simple centralized schemes: for the plot labeled "exhaustive user selection" the best of the $2^K - 1$ non-zero power allocations with $p_k \in \{0, 1\}$ for all $k \in \{1, \ldots, K\}$ was used; also, the best out of the K altruistic solutions, in which exactly one user has power one and all others have power zero, is shown.

Figures 3.17 and 3.18 show that for both $K = 4$ and $K = 10$ users the pricing algorithm and the projected gradient algorithm have very similar sum-rate performance, and that the influence of the step size δ on the performance is marginal. Also, it can be seen that the distributed algorithms do not generally find the global optimum, as they are outperformed by the exhaustive user selec-

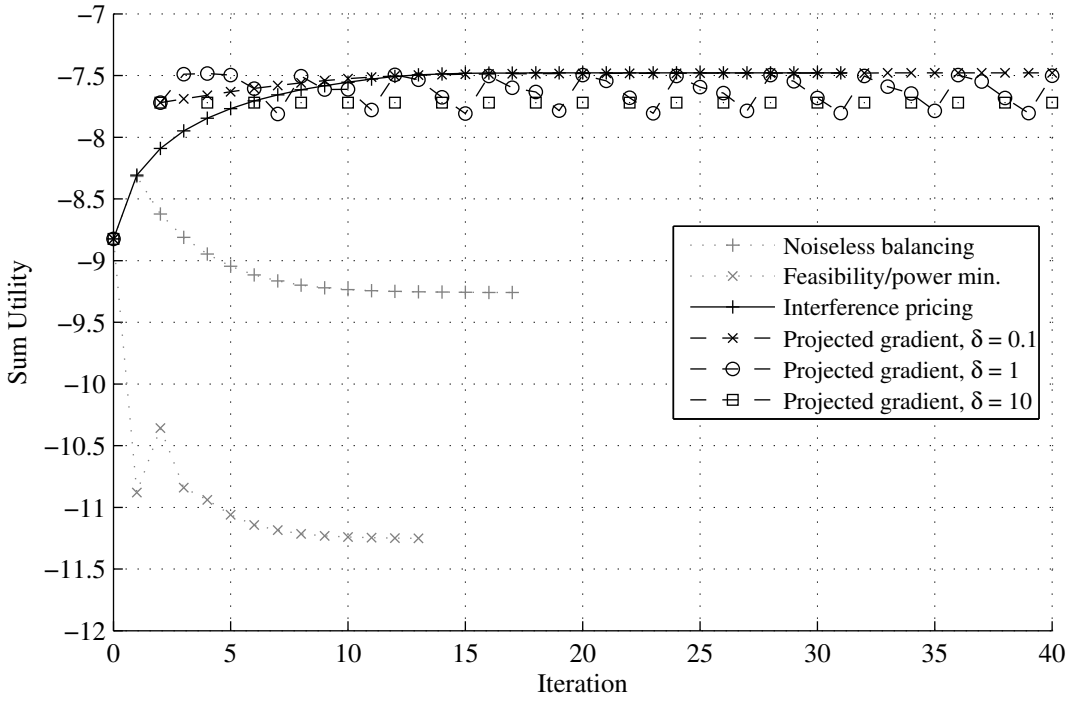


Figure 3.16: Scenario as in Figure 3.15 with logarithmic utility. Discontiuities in the plots for the projected gradient algorithm are a result of the sum utility being $-\infty$ after certain updates in which one or more users are given zero power.

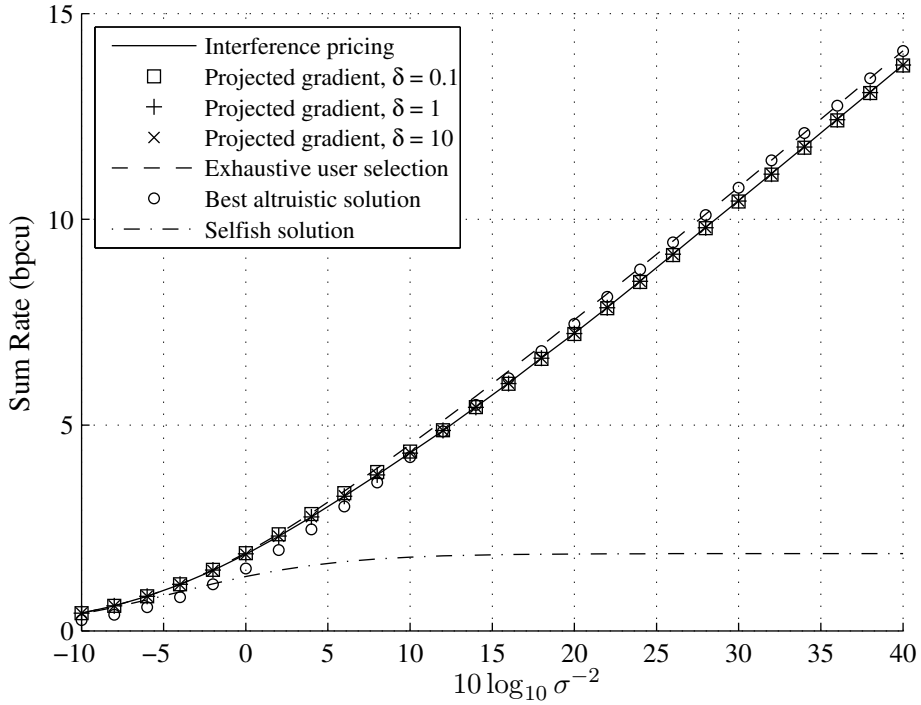


Figure 3.17: $K = 4$ users, sum rate averaged over 1000 channel realizations.

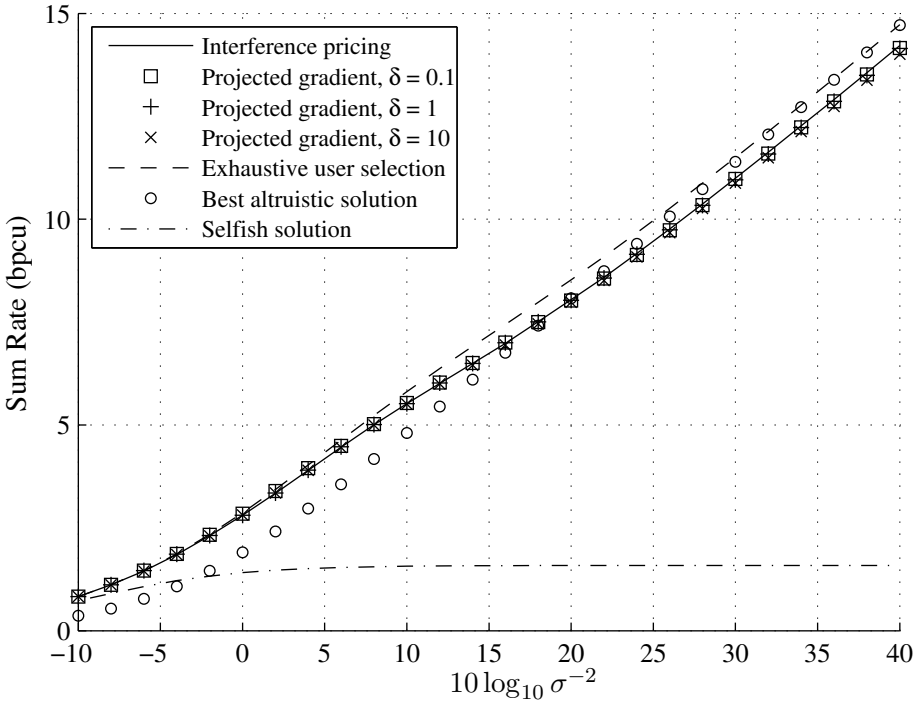


Figure 3.18: $K = 10$ users, sum rate averaged over 1000 channel realizations.

tion scheme. All schemes except for the selfish solution, however, achieve an asymptotic slope of approximately 3.3 bpcu per 10 dB increase of the inverse noise power. This is a consequence of the fact that when all cross channels are non-zero, the optimal sum rate in bpcu grows proportionally to $\log_2(\sigma^{-2}) = \log_2(10) \cdot \log_{10}(\sigma^{-2})$ as $\sigma^{-2} \to \infty$, cf. (3.63).

For the results in Figure 3.19 the channel model was modified so that the direct channels h_{kk} have unit variance, whereas the cross channels h_{kj} with $k \neq j$ have a variance of only 0.01. Here, it appears to be crucial that the projected gradient step size is sufficiently small in order to be able to reach the performance of the pricing algorithm.

For the logarithmic utility (Figure 3.20) the pricing algorithm is proven to find the global optimum and it appears that the step size $\delta = 0.1$ is sufficiently small so that the projected gradient algorithm also converges towards the global optimum. For $\delta = 0.5$, on the other hand, the plot disappears for inverse noise powers above 5 dB since the average sum utility is minus infinity. This indicates that in at least one channel realization the projected gradient algorithm failed to converge as a result of the step size being too large.

Finally, in Figure 3.21 we show the effect of choosing different initial power allocations on the sum-rate performance. In addition to the previously used initialization with zero power, the selfish solution was used as a starting point as well as a "random" power allocation where each user starts with a random power drawn from a uniform distribution between zero and one. These three initialization schemes have only a marginal difference in performance. A fourth scheme, where $p_1^{(0)} = 1$ and all other powers are set to zero initially, on the other hand, performs far worse, presumably because the initial allocation lies in the vicinity of a local optimum that is in most cases not globally optimal.

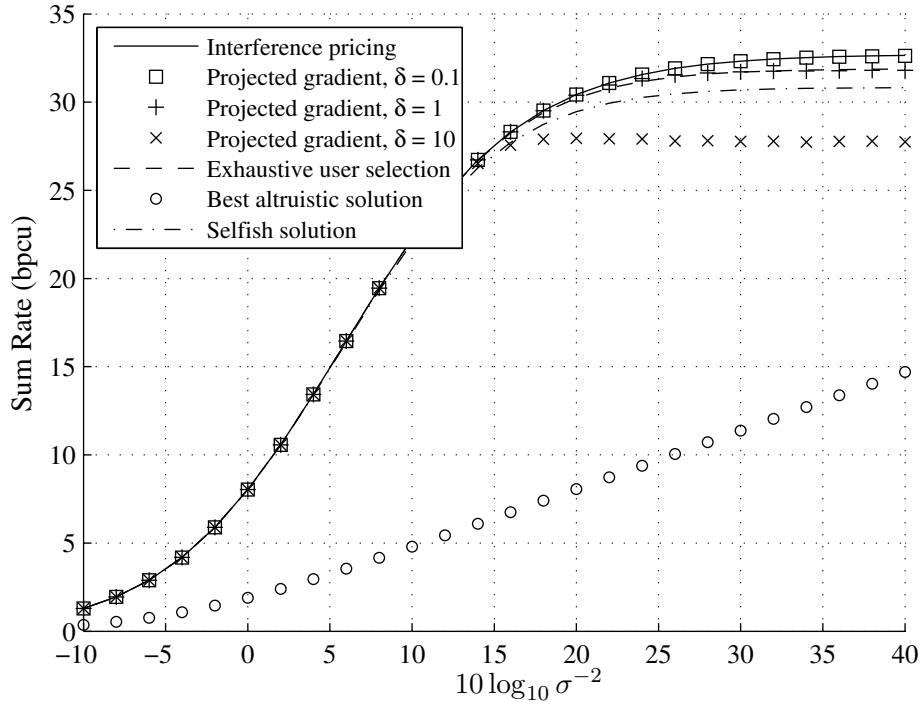


Figure 3.19: $K = 10$ users, sum rate averaged over 1000 channel realizations of a channel model where the variance of the cross channels is 0.01 and the variance of the direct channels is 1.

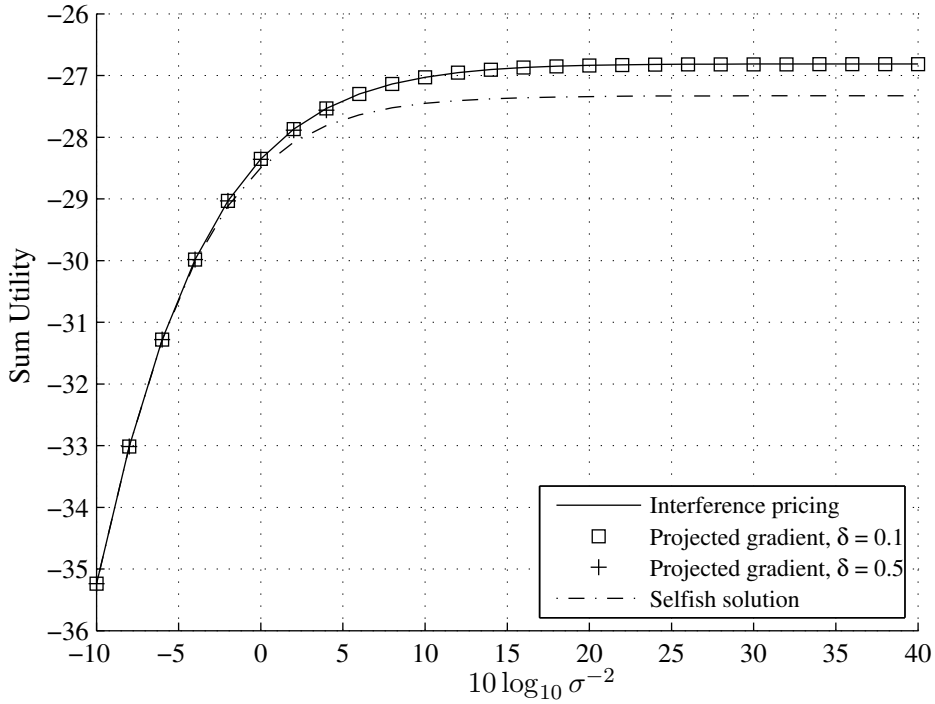


Figure 3.20: $K = 10$ users, sum logarithmic utility averaged over 1000 channel realizations. All channel coefficients have unit variance.

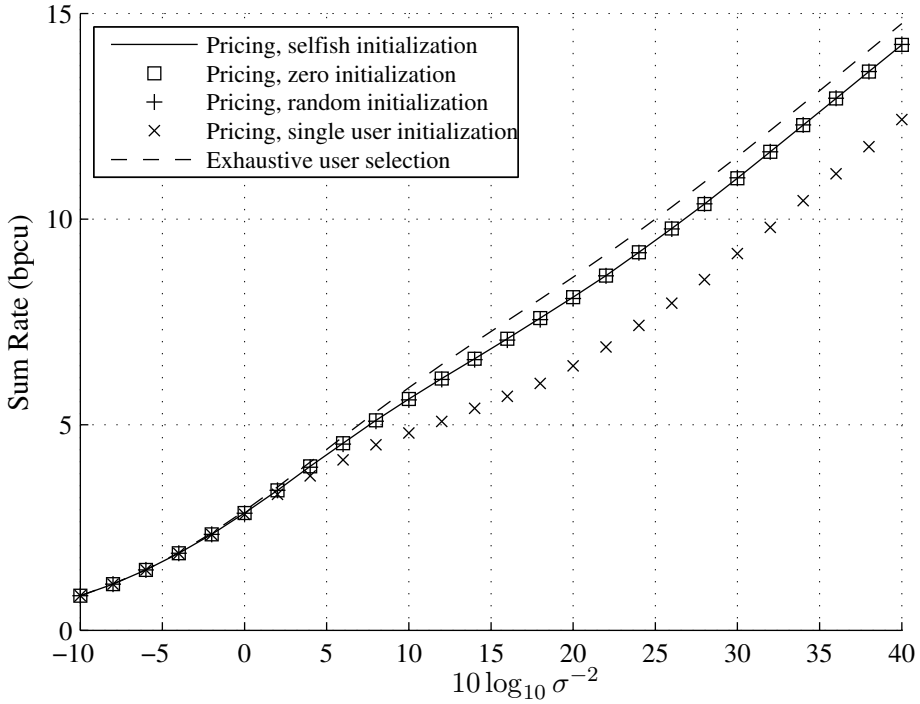


Figure 3.21: Scenario as in Figure 3.18, but with different initial power allocations.

4. MISO Interference Networks

If the transmitting nodes in a wireless interference network have more than one antenna, the transmit strategy consists of more than simply choosing a suitable transmit power; the signals transmitted from the different antennas can, e. g., be given different gains and phase shifts, thus making it possible to generate a specific spatial characteristic of the composite transmitted signal, a technique known as *beamforming*. As a simple example, consider a system with two users and two antennas at each transmitter: the transmitters can form their signals such that the individual signals from the two antennas exactly cancel out at the respective unintended receivers. Assuming that the signals do not happen to also cancel out at the intended receivers, we have effectively created two parallel non-interfering links, over which data can be transmitted with full power.

The additional possibilities of designing the transmit strategies to perform this type of *spatial interference cancellation,* however, make the mathematical structure of the underlying problems considerably more complex. In contrast to the SISO case, for multiple-input single-output (MISO) networks simple conditions for the convexity of the utility region remain elusive; similarly, closed-form solutions for the problems of deciding whether a given strategy is on the Pareto boundary of the utility region or whether a given SINR K-tuple is feasible have not been found. Finally, determining the globally optimal transmit strategy for the sum utility problem appears to be computationally infeasible in most cases; the focus of the algorithm design instead lies on good average performance, reliable convergence behavior, and computationally simple update procedures suitable for distributed implementation.

In this chapter we give an overview of some analytical results concerning optimal transmit strategies for MISO interference networks and proceed to discuss and numerically evaluate a number of different algorithms aimed at maximizing the sum utility.

4.1 System Model

The main difference to the system model of the previous chapter is that we now allow the transmitting nodes to have more than one antenna; we denote the number of antennas of transmitter k as N_k. Each receiving node in our system has one antenna. The complex gains of the channels from each of the N_j antennas of transmitter j to the antenna of receiver k are collected in the vector $\boldsymbol{h}_{kj} \in \mathbb{C}^{N_j}$. We furthermore require that the direct channels are non-zero, i. e., $\boldsymbol{h}_{kk} \neq \mathbf{0}$ for all $k \in \{1, \ldots, K\}$.

With the vector $\boldsymbol{x}_k \in \mathbb{C}^{N_k}$ containing the symbols transmitted from the N_k antennas of transmitter k, the received symbol of receiver k is

$$y_k = \boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{x}_k + \sum_{j \neq k} \boldsymbol{h}_{kj}^{\mathrm{T}} \boldsymbol{x}_j + n_k \qquad \forall k \in \{1, \ldots, K\} \tag{4.1}$$

where, again, the noise variance

$$\mathrm{E}\left[|n_k|^2\right] = \sigma^2 \qquad \forall k \in \{1, \ldots, K\} \tag{4.2}$$

with $\sigma^2 > 0$ is equal at all receivers. The transmit symbol vector $\boldsymbol{x}_k \in \mathbb{C}^{N_k}$ has mean zero and is uncorrelated between transmitters, i. e., $\mathrm{E}[\boldsymbol{x}_k] = \boldsymbol{0}$ and $\mathrm{E}[\boldsymbol{x}_j \boldsymbol{x}_k^{\mathrm{H}}] = \boldsymbol{0}$ for $j \neq k$. The transmit covariance matrix of user k is

$$\boldsymbol{Q}_k = \mathrm{E}\left[\boldsymbol{x}_k \boldsymbol{x}_k^{\mathrm{H}}\right] \in \mathbb{C}^{N_k \times N_k} \qquad \forall k \in \{1, \ldots, K\}. \tag{4.3}$$

Note that any valid covariance matrix $\boldsymbol{Q}_k$ must be positive semi-definite. We impose a unit power constraint on the sum of the powers across the antennas of a single user, which with (4.3) results in the condition

$$\mathrm{tr}\left(\boldsymbol{Q}_k\right) \leq 1 \qquad \forall k \in \{1, \ldots, K\}. \tag{4.4}$$

Since the power of the portion of the received symbol y_k that results from transmitter j is

$$\mathrm{E}\left[|\boldsymbol{h}_{kj}^{\mathrm{T}} \boldsymbol{x}_j|^2\right] = \boldsymbol{h}_{kj}^{\mathrm{T}} \boldsymbol{Q}_j \boldsymbol{h}_{kj}^{*} \qquad \forall (k, j) \in \{1, \ldots, K\}^2 \tag{4.5}$$

the SINR of user k is

$$\gamma_k = \frac{\boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{Q}_k \boldsymbol{h}_{kk}^{*}}{\sum\limits_{j \neq k} \boldsymbol{h}_{kj}^{\mathrm{T}} \boldsymbol{Q}_j \boldsymbol{h}_{kj}^{*} + \sigma^2} \qquad \forall k \in \{1, \ldots, K\}. \tag{4.6}$$

As for single-antenna systems, cf. Section 3.1, different per-user transmit power constraints and noise powers can be accomodated by appropriately renormalizing the channel coefficient vectors.

An important class of transmit strategies in MISO interference networks is *single-stream* beamforming, where the transmit signal $\boldsymbol{x}_k$ is the result of multiplying a unit-variance scalar data symbol s_k with a constant beamforming vector $\boldsymbol{v}_k$, i. e.,

$$\boldsymbol{x}_k = \boldsymbol{v}_k \cdot s_k \qquad \forall k \in \{1, \ldots, K\}. \tag{4.7}$$

The covariance matrices, which in this case have at most rank one, can be written as

$$\boldsymbol{Q}_k = \boldsymbol{v}_k \boldsymbol{v}_k^{\mathrm{H}} \qquad \forall k \in \{1, \ldots, K\}. \tag{4.8}$$

Since $\mathrm{tr}\left(\boldsymbol{Q}_k\right) = \|\boldsymbol{v}_k\|_2^2$, the unit power constraint is

$$\|\boldsymbol{v}_k\|_2^2 \leq 1 \qquad \forall k \in \{1, \ldots, K\}. \tag{4.9}$$

The SINR of user k expressed by means of the beamforming vectors is

$$\gamma_k = \frac{|\boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k|^2}{\sum\limits_{j \neq k} |\boldsymbol{h}_{kj}^{\mathrm{T}} \boldsymbol{v}_j|^2 + \sigma^2} \qquad \forall k \in \{1, \ldots, K\}. \tag{4.10}$$

4.2 The Utility Region

In the SISO case, the region of feasible transmit strategies, or power region, consists of K-tuples of real scalars; the SINR region and the utility region are also sets of K-tuples of real scalars, and the mapping between feasible power K-tuples and feasible SINR or utility K-tuples was shown to be bijective, cf. (3.38) and Appendix A2. In the MISO case, on the other hand, the transmit strategies

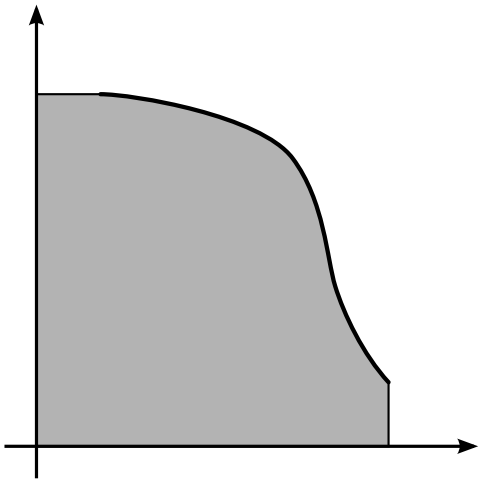


Figure 4.1: Example region in which the strong and weak Pareto boundary are not identical: sections of the boundary belonging to the weak Pareto boundary, but not to the strong Pareto boundary, are marked by a thinner line.

are K-tuples of covariance matrices $(\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_K)$, or, if we allow only single-stream beamforming, K-tuples of beamforming vectors $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K)$; the region of feasible transmit strategies therefore has a higher dimensionality than the SINR or utility region, which are again sets of K-tuples of real scalars $(\gamma_1, \ldots, \gamma_K)$ or $(u_1(\gamma_1), \ldots, u_K(\gamma_K))$. Consequently, there must exist SINR or utility K-tuples that can be achieved by more than one transmit strategy. This fundamental difference to the SISO case makes it considerably more difficult to analyze the properties of the SINR and utility region for MISO interference networks. Nonetheless, a number of interesting observations have been made, which we discuss in the following.

Definition 4.1. The *strong Pareto boundary* of a region is the subset of K-tuples in the region that are not dominated by any other K-tuple in the region with strict dominance in at least one component. Therefore, a K-tuple $(r_1, \ldots, r_K)$ is on the strong Pareto boundary if there does not exist another K-tuple $(\bar{r}_1, \ldots, \bar{r}_K) \neq (r_1, \ldots, r_K)$ in the region with $\bar{r}_k \geq r_k$ for all $k \in \{1, \ldots, K\}$.

The difference between the strong Pareto boundary and the previously utilized (weak) Pareto boundary, cf. Definition 3.3, is illustrated in Figure 4.1: instead of requiring that no other point of the region strictly dominates a boundary point, we now require that no other point dominates in at least one component with the other components being equal. Consequently, sections of the weak Pareto boundary that run parallel to one or more of the axes are not part of the strong Pareto boundary.

Proposition 4.1. *The SINR region and the utility region are comprehensive.*

Proof. To begin with, we assume a fixed K-tuple of positive semi-definite matrices $(\boldsymbol{Q}'_1, \ldots, \boldsymbol{Q}'_K)$ where $\mathrm{tr}(\boldsymbol{Q}'_k) = 1$ for all $k \in \{1, \ldots, K\}$ and define our covariance matrices as $\boldsymbol{Q}_k = p_k \boldsymbol{Q}'_k$ with $0 \leq p_k \leq 1$ for all $k \in \{1, \ldots, K\}$. From (4.6) and (3.5) it follows that the region of SINRs or utilities achievable by varying the powers p_k is identical to that of a SISO scenario with squared channel magnitudes $|h_{jk}|^2 = \boldsymbol{h}_{jk}^{\mathrm{T}} \boldsymbol{Q}'_k \boldsymbol{h}_{jk}^*$ for all $(j, k) \in \{1, \ldots, K\}^2$, which is comprehensive, cf. Propositions 3.2 and 3.5. The complete SINR or utility region for the MISO interference network clearly is the union of such SISO regions over all valid K-tuples $(\boldsymbol{Q}'_1, \ldots, \boldsymbol{Q}'_K)$ and is therefore also comprehensive. In the same way, comprehensiveness can be shown for the case of single-stream beamforming. □

Proposition 4.2 (Cf. [56, 57, 58])**.** *The SINR region and the utility region with the restriction of single-stream beamforming* (4.7) *are the same as without this restriction, i. e., with covariance matrices that may have a rank higher than one.*

Although highly intuitive, as having to decode multiple independent streams of information from a single scalar signal seems to be an unnecessary burden for the receiver, this result is not immediately evident from the expressions (4.6) and (4.10). It was proven independently in [56, 57, 58] with some effort; the approach in all three proofs is to first allow covariance matrices $\boldsymbol{Q}_k$ with arbitrary rank and to then show that for points on the strong Pareto boundary the rank of the resulting covariance matrices is at most one. Due to the comprehensiveness of the regions, identical strong Pareto boundaries imply that the regions are identical. As a consequence of Proposition 4.2, we can limit our discussion to single-stream beamforming without losing optimality.

Completely characterizing the weak or strong Pareto boundary of the SINR or utility region for a MISO interference network is not straightforward. In particular, for a given transmit strategy no simple method is known to determine whether the resulting SINR or utility K-tuple is on the Pareto boundary of the region. In [56, 57, 58], however, different necessary, but not sufficient, conditions were derived for strong Pareto optimality. In the following we discuss the necessary conditions from [58], from which some useful conclusions can be drawn.

Proposition 4.3 (Cf. [58])**.** *Any beamformer $\boldsymbol{v}_k$ resulting in an SINR K-tuple on the strong Pareto boundary of the SINR region (and consequently resulting in a utility K-tuple on the strong Pareto boundary of the utility region) must fulfill*

$$\boldsymbol{Z}_k \boldsymbol{v}_k = \mu \boldsymbol{v}_k \tag{4.11}$$

where

$$\boldsymbol{Z}_k = \lambda_{kk} \boldsymbol{h}_{kk}^* \boldsymbol{h}_{kk}^{\mathrm{T}} - \sum_{j \neq k} \lambda_{jk} \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}}, \tag{4.12}$$

μ is the highest eigenvalue of $\boldsymbol{Z}_k$, and

$$\sum_{j=1}^{K} \lambda_{jk} = 1 \qquad \textit{with} \qquad \lambda_{jk} \geq 0 \quad \forall j \in \{1, \ldots, K\}. \tag{4.13}$$

Furthermore, $\boldsymbol{v}_k = \boldsymbol{0}$ if $\mu < 0$ and $\|\boldsymbol{v}_k\|_2^2 = 1$ if $\mu > 0$.

The proof draws on the concept of *power gain regions:* the power gain from transmitter k to receiver j is defined as $|\boldsymbol{h}_{jk}^{\mathrm{T}} \boldsymbol{v}_k|^2$; the power gain region of user k is the set of K-tuples $(|\boldsymbol{h}_{1k}^{\mathrm{T}} \boldsymbol{v}_k|^2, \ldots, |\boldsymbol{h}_{Kk}^{\mathrm{T}} \boldsymbol{v}_k|^2)$ that are achievable with a feasible beamformer $\boldsymbol{v}_k$. Clearly, user k must choose its beamformer $\boldsymbol{v}_k$ such that $|\boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k|^2$ is high, whereas $|\boldsymbol{h}_{jk}^{\mathrm{T}} \boldsymbol{v}_k|^2$ for $j \neq k$ is low. In particular, a beamformer $\boldsymbol{v}_k$ resulting in a power gain K-tuple such that $|\boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k|^2$ can be further increased and all $|\boldsymbol{h}_{jk}^{\mathrm{T}} \boldsymbol{v}_k|^2$ with $j \neq k$ can be further decreased at the same time cannot result in a Pareto optimal SINR K-tuple, since $\boldsymbol{v}_k$ can be changed to increase all SINRs. Therefore, only a certain section of the boundary of the K power gain regions comes into question for transmit strategies resulting in SINR K-tuples on the Pareto boundary of the SINR region. This section of the boundary of the power gain region of user k is characterized by above condition.

Some remarks on the result stated in Proposition 4.3:

- As can be observed from inserting (4.12) into (4.11), the beamformer $\boldsymbol{v}_k$ of user k is a linear combination of the conjugate channel vectors between transmitter k and all receivers $j \in$

$\{1, \ldots, K\}$, i. e.,

$$\boldsymbol{v}_k = \sum_{j=1}^{K} \xi_{jk} \boldsymbol{h}_{jk}^* \tag{4.14}$$

with $\xi_{jk} \in \mathbb{C}$. This result was previously proven in [59].

- If $\boldsymbol{h}_{kk}$ is linearly independent of the space spanned by all $\boldsymbol{h}_{jk}$ with $j \neq k$, the highest eigenvalue of $\boldsymbol{Z}_k$ is non-negative regardless of the parameters $\lambda_{1k}, \ldots, \lambda_{Kk}$. For this case it was additionally shown in [58] that $\|\boldsymbol{v}_k\|_2^2 = 1$ is a necessary condition for strong Pareto optimality. Furthermore, in this case it is possible for transmitter k to choose a beamformer $\boldsymbol{v}_k$ such that $\boldsymbol{h}_{jk}^{\mathrm{T}} \boldsymbol{v}_k = 0$ for all $j \neq k$ while $\boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k \neq 0$, i. e., data can be transmitted to the desired receiver without causing interference to the undesired receivers. We refer to a beamformer fulfilling these conditions as a *zero-forcing* (ZF) solution and state that, if zero-forcing is possible for user k, it is always optimal for user k to transmit with full power.
- In the special case of $K = 2$ users with $N_1 \geq 2$ and $N_2 \geq 2$ transmit antennas as well as linearly independent channel vectors $\boldsymbol{h}_{11}$ and $\boldsymbol{h}_{21}$, and linearly independent channel vectors $\boldsymbol{h}_{22}$ and $\boldsymbol{h}_{12}$, it follows from the previous remark that both users transmit at full power; also, the necessary condition can be further simplified:

$$\boldsymbol{v}_k = \frac{\lambda_k \boldsymbol{v}_k^{\mathrm{MF}} + (1 - \lambda_k) \boldsymbol{v}_k^{\mathrm{ZF}}}{\left\| \lambda_k \boldsymbol{v}_k^{\mathrm{MF}} + (1 - \lambda_k) \boldsymbol{v}_k^{\mathrm{ZF}} \right\|_2} \qquad \forall k \in \{1, 2\} \tag{4.15}$$

with $\lambda_k \in [0, 1]$, where

$$\boldsymbol{v}_k^{\mathrm{MF}} = \frac{\boldsymbol{h}_{kk}^*}{\|\boldsymbol{h}_{kk}^*\|_2} \qquad \text{and} \qquad \boldsymbol{v}_k^{\mathrm{ZF}} = \frac{\left(\mathbf{I} - \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}} / \|\boldsymbol{h}_{jk}\|_2^2\right) \boldsymbol{h}_{kk}^*}{\left\| \left(\mathbf{I} - \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}} / \|\boldsymbol{h}_{jk}\|_2^2\right) \boldsymbol{h}_{kk}^* \right\|_2} \quad \text{with } j \neq k. \tag{4.16}$$

With the beamformer $\boldsymbol{v}_k^{\mathrm{MF}}$ user k maximizes its own SINR, whereas $\boldsymbol{v}_k^{\mathrm{ZF}}$ causes no interference at the unintended transmitter; therefore, the optimal strategies can be viewed as trading off between selfishness and altruism. A derivation of the expressions for $\boldsymbol{v}_k^{\mathrm{MF}}$ and $\boldsymbol{v}_k^{\mathrm{ZF}}$ as well as a more detailed discussion of the concepts of selfishness and altruism is given in the following section.

- Due to the constraint on $\sum_j \lambda_{jk}$, the set of allowed K-tuples $(\lambda_{1k}, \ldots, \lambda_{kK})$ has $K - 1$ "independent" parameters. Consequently, the number of independent parameters in the conditions on all K beamformers is $K(K-1)$ if it is clear that all beamformers use full power, and up to K^2 if $\|\boldsymbol{v}_k\|_2^2$ is also variable for all users. The Pareto boundary of the SINR or utility region, on the other hand, is a $K-1$-dimensional manifold in $\mathbb{R}^K$. It is therefore intuitive that only a small subset of those transmit strategies fulfilling the necessary conditions is actually on the strong Pareto boundary.

As will be shown in Section 4.5, examples can be found for SINR regions that are convex and non-convex. Similarly, the utility region with the achievable rate utility function can be convex or non-convex. Moreover, in contrast to the SISO case, it is possible for the utility region with the logarithmic utility function to be non-convex in MISO interference networks. A simple sufficient condition on the utility functions that guarantees a convex utility region, as was given for the SISO case in Theorem 3.1, is not known for scenarios with multiple antennas.

4.3 Optimal Transmit Strategies

4.3.1 The Selfish Solution or Nash Equilibrium

If user k is concerned only about maximizing its own utility, it must choose the beamformer $\boldsymbol{v}_k$ that maximizes its SINR γ_k. As can be seen in (4.10), γ_k is increasing in $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2$ and does not depend on $\boldsymbol{v}_k$ in any other way. Therefore, the selfish beamformer of user k fulfills

$$\boldsymbol{v}_k^{\mathrm{MF}} = \arg\max_{\boldsymbol{v}_k} |\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2 \qquad \text{s.t.:} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1. \tag{4.17}$$

It is straightforward to show that the resulting beamformer is

$$\boldsymbol{v}_k^{\mathrm{MF}} = \frac{1}{\|\boldsymbol{h}_{kk}^*\|_2} \cdot \boldsymbol{h}_{kk}^* \tag{4.18}$$

i. e., the linear filter "matched" to the direct channel $\boldsymbol{h}_{kk}$ (hence the super-script "MF"). The beamformer $\boldsymbol{v}_k^{\mathrm{MF}}$ fulfills the necessary condition for strong Pareto optimality (4.11): the corresponding parameter values are $\lambda_{kk} = 1$ and $\lambda_{jk} = 0$ for all $j \neq k$.

When all users $k \in \{1, \ldots, K\}$ apply their respective selfish beamformers $\boldsymbol{v}_k^{\mathrm{MF}}$, we refer to the transmit strategy as the *selfish solution.* Since the conditions (4.11) are not sufficient for strong Pareto optimality, the selfish solution generally is in the interior of the utility region, as will also become apparent in the numerical examples in Section 4.5.

Similar to the SISO case, we can formulate a game with the K users as players, the individual utility functions as payoff, and the choice of the own beamformer $\boldsymbol{v}_k$ as the strategy of player k; then, again, the selfish solution is the unique Nash equilibrium of the game, i. e., the set of strategies from which it is not possible for any player to improve its own payoff by unilaterally changing its strategy.

4.3.2 Altruistic Solutions

As in Section 3.4.2, we define an altruistic solution to be a K-tuple of beamformers $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K)$ for which each user k

1) cannot unilaterally improve the utility function $u_j(\gamma_j)$ of any user $j \neq k$ by changing $\boldsymbol{v}_k$ and
2) cannot improve its own utility $u_k(\gamma_k)$ by changing $\boldsymbol{v}_k$ without decreasing at least one other utility $u_j(\gamma_j)$ with $j \neq k$.

Since $u_j(\gamma_j)$ is strictly decreasing in $|\boldsymbol{h}_{jk}^{\mathrm{T}}\boldsymbol{v}_k|^2$ if and only if $|\boldsymbol{h}_{jj}^{\mathrm{T}}\boldsymbol{v}_j|^2$ is non-zero, the first condition implies that the interference power caused by transmitter k at receiver j must be zero for all unintended users $j \neq k$ for which $\gamma_j > 0$. If we partition the K users into two sets, the *active* users with $\gamma_k > 0$ and the *inactive* users with $\gamma_k = 0$, each user in the active set must solve the following optimization problem in order to be able to fulfill both conditions for an altruistic solution:

$$\begin{aligned} \boldsymbol{v}_k^{\mathrm{ZF}} = \arg\max_{\boldsymbol{v}_k} |\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2 \qquad \text{s.t.:} \quad & \boldsymbol{h}_{jk}^{\mathrm{T}}\boldsymbol{v}_k = 0 \quad \text{for all users } j \neq k \text{ in the active set} \\ & \text{and} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1. \end{aligned} \tag{4.19}$$

The super-script "ZF" for zero-forcing indicates that the beamformer is designed to cause zero interference to other active users.

As is shown in Appendix A5, the beamformer that solves above optimization problem is

$$\boldsymbol{v}_k^{\mathrm{ZF}} = \frac{1}{\|\boldsymbol{\Pi}_k \boldsymbol{h}_{kk}^*\|_2} \cdot \boldsymbol{\Pi}_k \boldsymbol{h}_{kk}^* \tag{4.20}$$

where the projection matrix $\boldsymbol{\Pi}_k$ is defined as

$$\boldsymbol{\Pi}_k = \mathbf{I} - \boldsymbol{H}_k^{+} \boldsymbol{H}_k \tag{4.21}$$

and the matrix $\boldsymbol{H}_k$ contains the stacked row vectors $\boldsymbol{h}_{jk}^{\mathrm{T}}$ for all $j \neq k$ in the active set. If $\boldsymbol{H}_k \boldsymbol{H}_k^{\mathrm{H}}$ is invertible, i. e., if the vectors $\boldsymbol{h}_{jk}^{\mathrm{T}}$ for all $j \neq k$ in the active set are linearly independent, the projection matrix $\boldsymbol{\Pi}_k$ can also be written as

$$\boldsymbol{\Pi}_k = \mathbf{I} - \boldsymbol{H}_k^{\mathrm{H}} \left(\boldsymbol{H}_k \boldsymbol{H}_k^{\mathrm{H}} \right)^{-1} \boldsymbol{H}_k. \tag{4.22}$$

If the matrix $\boldsymbol{H}_k$ has rank N_k, the projection matrix $\boldsymbol{\Pi}_k$ is the zero matrix. In this case, it is impossible for user k to fulfill both altruism conditions while remaining in the set of active users with $\gamma_k > 0$.

A K-tuple of beamformers is a candidate for an altruistic solution when all users k in the active set employ their respective beamformer $\boldsymbol{v}_k^{\mathrm{ZF}}$ and thus all users either receive zero interference or are inactive. For an altruistic solution, it must additionally be impossible for a user in the inactive set to become active without harming the users in the active set, i. e., $\boldsymbol{v}_k^{\mathrm{ZF}}$ must be the zero vector for all users in the inactive set, as otherwise the second condition would be violated. In general there may be multiple altruistic solutions with different sets of active and inactive users.

If all users can simultaneously perform zero-forcing, i. e., if the active set consists of all K users and $\boldsymbol{v}_k^{\mathrm{ZF}}$ is non-zero for all $k \in \{1, \ldots, K\}$, the altruistic solution is unique. Zero-forcing is possible for all users, e. g., when for all $k \in \{1, \ldots, K\}$ $N_k \geq K$ and the vectors $\boldsymbol{h}_{jk}$ for all $j \in \{1, \ldots, K\}$ are linearly independent.

All altruistic solutions also fulfill the necessary conditions for strong Pareto optimality (4.11): the corresponding values of the parameters are $\lambda_{kk} = 0$, $\lambda_{jk} \neq 0$ for all users j in the active set, and $\lambda_{jk} = 0$ for those in the inactive set. However, the altruistic strategies in general are not on the Pareto boundary, as will be seen in some numerical examples in Section 4.5.

4.3.3 Successive Zero-Forcing Strategies

Another set of interesting transmit strategies for MISO interference networks can be constructed by assigning the beamformers one after another in a certain order, where each user has the constraint that it may not decrease the previously allocated users' SINRs.

For ease of notation, let us assume without loss of generality that the order in which the beamformers are allocated coincides with the user indices, i. e., $\boldsymbol{v}_1$ is allocated first, $\boldsymbol{v}_2$ second, and so on. The first user must not fulfill any zero-forcing constraints and therefore employs the selfish beamformer $\boldsymbol{v}_1^{\mathrm{MF}}$. The second user chooses $\boldsymbol{v}_2$ to maximize $|\boldsymbol{h}_{22}^{\mathrm{T}} \boldsymbol{v}_2|^2$ subject to the constraint that the first user may not be harmed, i. e., $\boldsymbol{h}_{12}^{\mathrm{T}} \boldsymbol{v}_2 = 0$. Accordingly, the beamformer of user k solves the optimization problem

$$\max_{\boldsymbol{v}_k} |\boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k|^2 \qquad \text{s. t.:} \quad \boldsymbol{h}_{jk}^{\mathrm{T}} \boldsymbol{v}_k = 0 \quad \text{for all users } j < k \text{ with } \gamma_j > 0 \quad \text{and} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1. \tag{4.23}$$

The solution can be found analogously to (4.19) by means of a projection matrix that orthogonally projects the conjugate channel vector $\boldsymbol{h}_{kk}^{*}$ onto the subspace orthogonal to the interference channel vectors of the zero-forcing constraint. If the number of previously allocated active users is greater or equal to the number of antennas N_k or if the channel vectors are not linearly independent, the solution may be the zero beamformer.

The SINR or utility K-tuple resulting from this successive zero-forcing strategy is guaranteed to be on the strong Pareto boundary of the SINR or utility region, as no user's SINR can be improved without decreasing another user's SINR. To show this, let us assume that we would like to change the transmit strategy to improve the SINR of user k without decreasing any of the SINRs of the users j with $j < k$; we observe that

- $|\boldsymbol{h}_{11}^{\mathrm{T}}\boldsymbol{v}_1|^2$ may not be changed, as it is already at its maximum and any change would thus lead to a decrease in γ_1;
- $|\boldsymbol{h}_{22}^{\mathrm{T}}\boldsymbol{v}_2|^2$ may not be changed, as this would either lead to a decrease in γ_1 (if the ZF constraint is violated) or to a decrease in γ_2 (if the ZF constraint is not violated);
- in the same way, all direct power gains up to user $k-1$ may not be changed, as otherwise either the own SINR or that of a previously allocated user would decrease;
- also, the power gain $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2$ of user k, of which we would like to improve the SINR, may not be increased, as this would lead to a decrease in the SINR of a previously allocated user;
- and finally, changing any beamformer $\boldsymbol{v}_j$ with $j > k$ cannot improve γ_k, as those users cause no interference to receiver k.

Therefore, such a change in transmit strategy is not possible.

Pareto optimality of the successive ZF strategies holds regardless of the order in which the beamformers are allocated; consequently, up to $K!$ different points on the Pareto boundary can be constructed with this method. In a system with $K = 2$, there are only two such possibilities: user one employs the selfish beamformer $\boldsymbol{v}_1^{\mathrm{MF}}$ and user two employs the altruistic beamformer $\boldsymbol{v}_2^{\mathrm{ZF}}$, or vice versa, i. e., the pair of beamformers is $\left(\boldsymbol{v}_1^{\mathrm{ZF}}, \boldsymbol{v}_2^{\mathrm{MF}}\right)$.

4.3.4 The Feasibility and Balancing Problems

In the SISO case it was possible to check the feasibility of a given SINR K-tuple by solving a system of linear equations and examining the corresponding K-tuple of powers, cf. Section 3.4.3. As previously discussed, there is no one-to-one correspondence between SINR K-tuples and transmit strategies in the MISO case; in general many different K-tuples of beamformers may result in the same SINR K-tuple.

The difficulty of checking feasibility in the MISO case becomes apparent when we state the feasibility conditions for the SINR K-tuple $(\bar{\gamma}_1, \ldots, \bar{\gamma}_K)$ in a way similar to the SISO case in (A10)–(A12):

$$\frac{1}{\bar{\gamma}_1}|\boldsymbol{h}_{11}^{\mathrm{T}}\boldsymbol{v}_1|^2 - \sum_{j\neq 1}|\boldsymbol{h}_{1j}^{\mathrm{T}}\boldsymbol{v}_j|^2 \geq \sigma^2 \tag{4.24}$$

$$\vdots$$

$$\frac{1}{\bar{\gamma}_K}|\boldsymbol{h}_{KK}^{\mathrm{T}}\boldsymbol{v}_K|^2 - \sum_{j\neq K}|\boldsymbol{h}_{Kj}^{\mathrm{T}}\boldsymbol{v}_j|^2 \geq \sigma^2 \tag{4.25}$$

From this system of inequalities it is not evident whether one or more sets of feasible beamformers can fulfill the conditions, let alone how to determine a feasible transmit strategy. It was shown, however, in [60] (and previously for a similar system model in [61, 62]) that these conditions can be reformulated as *second-order cone* constraints (cf. [36, Section 4.4.2]) of the form

$$\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}\|_2 \leq \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} + d. \tag{4.26}$$

To this end, we first add $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2$ to both sides of the kth inequality for all $k \in \{1,\ldots,K\}$ to obtain

$$\left(1+\frac{1}{\bar{\gamma}_1}\right)|\boldsymbol{h}_{11}^{\mathrm{T}}\boldsymbol{v}_1|^2 \geq \sum_j |\boldsymbol{h}_{1j}^{\mathrm{T}}\boldsymbol{v}_j|^2 + \sigma^2 \tag{4.27}$$

$$\vdots$$

$$\left(1+\frac{1}{\bar{\gamma}_K}\right)|\boldsymbol{h}_{KK}^{\mathrm{T}}\boldsymbol{v}_K|^2 \geq \sum_j |\boldsymbol{h}_{Kj}^{\mathrm{T}}\boldsymbol{v}_j|^2 + \sigma^2. \tag{4.28}$$

Next, we define the matrices $\bar{\boldsymbol{H}}_k$ for all $k \in \{1,\ldots,K\}$ and the vector $\boldsymbol{v}$ in the following way:

$$\bar{\boldsymbol{H}}_k = \begin{bmatrix} \boldsymbol{h}_{k1}^{\mathrm{T}} & \cdots & \boldsymbol{0}^{\mathrm{T}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0}^{\mathrm{T}} & \cdots & \boldsymbol{h}_{kK}^{\mathrm{T}} \\ \boldsymbol{0}^{\mathrm{T}} & \cdots & \boldsymbol{0}^{\mathrm{T}} \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{v} = \begin{bmatrix} \boldsymbol{v}_1 \\ \vdots \\ \boldsymbol{v}_K \end{bmatrix}. \tag{4.29}$$

Also, the vector $\boldsymbol{\sigma} = [0,\ldots,0,\sigma]^{\mathrm{T}}$ contains the square root of the noise power with K preceding zeros, and the selection matrix $\boldsymbol{S}_k = [\boldsymbol{0},\mathbf{I},\boldsymbol{0}]$ is designed such that $\boldsymbol{S}_k\boldsymbol{v} = \boldsymbol{v}_k$ for all $k \in \{1,\ldots,K\}$. With these definitions feasibility is equivalent to

$$\|\bar{\boldsymbol{H}}_k\boldsymbol{v}+\boldsymbol{\sigma}\|_2^2 \leq \left(1+\frac{1}{\bar{\gamma}_k}\right)|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{S}_k\boldsymbol{v}|^2 \qquad \forall k \in \{1,\ldots,K\}. \tag{4.30}$$

Furthermore, since all SINRs γ_k are invariant to a complex phase shift of any beamforming vector $\boldsymbol{v}_j$, we can assume without loss of generality that $\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{S}_k\boldsymbol{v}$ is real-valued and non-negative for all $k \in \{1,\ldots,K\}$, so that we can take the square root of both sides of the above conditions:

$$\|\bar{\boldsymbol{H}}_k\boldsymbol{v}+\boldsymbol{\sigma}\|_2 \leq \sqrt{1+\frac{1}{\bar{\gamma}_k}}\cdot\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{S}_k\boldsymbol{v} \qquad \forall k \in \{1,\ldots,K\}. \tag{4.31}$$

Finally, the power constraints can be written as

$$\|\boldsymbol{S}_k\boldsymbol{v}\|_2 \leq 1 \qquad \forall k \in \{1,\ldots,K\} \tag{4.32}$$

and we have thus expressed all conditions that must be fulfilled for feasibility of the SINR K-tuple $(\bar{\gamma}_1,\ldots,\bar{\gamma}_K)$ as second-order cones in the variable $\boldsymbol{v}$.

A *second-order cone problem* (SOCP) is a convex optmization problem in which a linear objective is minimized subject to second-order cone constraints and linear equality constraints [36, Section 4.4.2]. SOCPs can be solved very efficiently with general purpose algorithms, relying, e. g., on interior point methods. An overview and evaluation of implementations of such algorithms can be found in [63]. By simply using zero as the objective function, we can use such an algorithm to perform our feasibility check. Alternatively, in [60], two methods were proposed to split the burden of running the optimization algorithm between the K users. If the SINR K-tuple is indeed feasible, the SOCP feasibility check also returns a corresponding set of beamformers, which is, however, not necessarily optimal in the sense of total transmit power.

In order to find the transmit strategy that achieves the SINR K-tuple $(\bar{\gamma}_1,\ldots,\bar{\gamma}_K)$ with the lowest total power, we introduce a slack variable p which upper-bounds the square root of the total power, cf. [62]. Now, the problem of minimizing p over the variables $\boldsymbol{v}$ and p subject to the above feasibility constraints and the additional conic constraint $\|\boldsymbol{v}\|_2 \leq p$ is an SOCP that solves the power minimization problem.

4.3.4.1 Solving Balancing Problems with Repeated Feasibility Checks

As discussed in Section 3.4.4, the SINR balancing problem with the target K-tuple $(\bar{\gamma}_1, \ldots, \bar{\gamma}_K)$ is

$$\max_{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_K,\beta} \beta \qquad \text{s.t.:} \quad \gamma_k \geq \beta\bar{\gamma}_k \quad \forall k \in \{1,\ldots,K\}. \tag{4.33}$$

The solution to this problem can be used as a feasibility check: the K-tuple $(\bar{\gamma}_1, \ldots, \bar{\gamma}_K)$ is feasible if the resulting scaling parameter β is greater or equal to one, otherwise the K-tuple is infeasible. Consequently, the balancing problem is inherently at least as difficult as a feasibility check.

A straightforward solution to the balancing problem in the MISO case is not known. Instead, it was proposed in [60] (and in [62] for a similar system model) to find the optimal β by performing a line search with bisection and repeated feasibility checks. To begin with, a surely feasible value $\check{\beta}$ is needed, e. g., $\check{\beta} = 0$, as well as a surely infeasibly value $\hat{\beta}$, which can be determined, e. g., by assuming that all users employ their respective selfish beamformers $\boldsymbol{v}_k^{\mathrm{MF}}$ and bounding the SINR γ_k from above by assuming that all interference power gains are zero. Next, the arithmetic mean $\bar{\beta} = \left(\check{\beta} + \hat{\beta}\right)/2$ is computed and the K-tuple $\left(\bar{\beta}\bar{\gamma}_1, \ldots, \bar{\beta}\bar{\gamma}_K\right)$ is checked for feasibility using an SOCP solver. If $\bar{\beta}$ leads to a feasible SINR K-tuple, $\check{\beta}$ is updated to the value of $\bar{\beta}$, otherwise $\hat{\beta}$ is updated to the value of $\bar{\beta}$; thus $\check{\beta}$ and $\hat{\beta}$ after the update are still a lower and an upper bound on the solution, one of which has been tightened by the feasibility check. This procedure of checking the feasibility of the arithmetic mean of $\check{\beta}$ and $\hat{\beta}$ and updating one of the two depending on the outcome of the feasibility check is continued until $\check{\beta}$ and $\hat{\beta}$ have converged to the same value up to a given precision. This value is the solution to the balancing problem.

The bisection approach can also be used to balance the utilities, i. e., to maximize β such that $(\beta\bar{u}_1, \ldots, \beta\bar{u}_K)$ is a feasible utility K-tuple. For each feasibility check, the SINR K-tuple corresponding to the utility K-tuple $\left(\bar{\beta}\bar{u}_1, \ldots, \bar{\beta}\bar{u}_K\right)$ must be computed, which, however, is a straightforward operation, as the utility functions are by definition invertible.

4.3.5 The Sum Utility Problem

The problem of finding the beamformers that maximize the sum utility is

$$\max_{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_K} \sum_{k=1}^{K} u_k(\gamma_k) \qquad \text{s.t.:} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1 \quad \forall k \in \{1,\ldots,K\}. \tag{4.34}$$

With the Lagrangian multipliers $\mu_1, \ldots, \mu_K$ we obtain the KKT conditions

$$\sum_{j=1}^{K} \frac{\partial u_j(\gamma_j)}{\partial \boldsymbol{v}_k^*} - \mu_k \boldsymbol{v}_k = \boldsymbol{0} \qquad \forall k \in \{1,\ldots,K\} \tag{4.35}$$

$$\mu_k \geq 0 \qquad \forall k \in \{1,\ldots,K\} \tag{4.36}$$

$$\mu_k \left(1 - \|\boldsymbol{v}_k\|_2^2\right) = 0 \qquad \forall k \in \{1,\ldots,K\} \tag{4.37}$$

$$\|\boldsymbol{v}_k\|_2^2 \leq 1 \qquad \forall k \in \{1,\ldots,K\} \tag{4.38}$$

which are necessary, but in general not sufficient, for optimality.[1] Note that we have fewer KKT conditions than in the SISO case, cf. (3.48)–(3.53). This can be explained by the fact that the

[1] When taking derivatives with respect to complex variables in the context of optimization, we use *Wirtinger calculus* (cf. [64, p. 64]), which can be understood as a shorthand for taking the derivatives with respect to the real and imaginary parts of the variable and, when searching for a stationary point, setting both to zero.

beamformer is not the multi-antenna generalization of the transmit power, but of the square root of the transmit power; the generalization of the transmit power is the transmit covariance matrix. If we were to optimize over the matrices $\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_K$ instead of the vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$, we would indeed have the additional constraint that all covariances must be positive semi-definite, corresponding to the constraint of non-negative powers in the SISO case. When optimizing over the beamformers, this constraint is fulfilled automatically.

With the chain rule, we can express the derivatives as

$$\frac{\partial u_k(\gamma_k)}{\partial \boldsymbol{v}_k^*} = u_k'(\gamma_k) \cdot \frac{\partial \gamma_k}{\partial \boldsymbol{v}_k^*} = \frac{u_k'(\gamma_k)}{\sum_{i \neq k} |\boldsymbol{h}_{ki}^{\mathrm{T}} \boldsymbol{v}_i|^2 + \sigma^2} \cdot \boldsymbol{h}_{kk}^* \boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k \tag{4.39}$$

and

$$\frac{\partial u_j(\gamma_j)}{\partial \boldsymbol{v}_k^*} = u_j'(\gamma_j) \cdot \frac{\partial \gamma_j}{\partial \boldsymbol{v}_k^*} = -\frac{u_j'(\gamma_j) |\boldsymbol{h}_{jj}^{\mathrm{T}} \boldsymbol{v}_j|^2}{\left(\sum_{i \neq j} |\boldsymbol{h}_{ji}^{\mathrm{T}} \boldsymbol{v}_i|^2 + \sigma^2 \right)^2} \cdot \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}} \boldsymbol{v}_k \tag{4.40}$$

for $j \neq k$. The first KKT condition (4.35) can now be stated as

$$\boldsymbol{A}_k \boldsymbol{v}_k = \mu_k \boldsymbol{v}_k \qquad \forall k \in \{1, \ldots, K\} \tag{4.41}$$

with the matrix

$$\boldsymbol{A}_k = \frac{u_k'(\gamma_k)}{\sum_{i \neq k} |\boldsymbol{h}_{ki}^{\mathrm{T}} \boldsymbol{v}_i|^2 + \sigma^2} \cdot \boldsymbol{h}_{kk}^* \boldsymbol{h}_{kk}^{\mathrm{T}} - \sum_{j \neq k} \frac{u_j'(\gamma_j) |\boldsymbol{h}_{jj}^{\mathrm{T}} \boldsymbol{v}_j|^2}{\left(\sum_{i \neq j} |\boldsymbol{h}_{ji}^{\mathrm{T}} \boldsymbol{v}_i|^2 + \sigma^2 \right)^2} \cdot \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}}. \tag{4.42}$$

Note that the structure of $\boldsymbol{A}_k$ is similar to that of $\boldsymbol{Z}_k$ in (4.12): it is the sum of the outer products of the channel vectors from transmitter k to all receivers j, weighted with a real-valued scalar with positive sign for the direct channel and non-positive sign for the cross channels. Also, similar to (4.11), the beamformer $\boldsymbol{v}_k$ that fulfills the first KKT condition is the eigenvector corresponding to a non-negative eigenvalue of $\boldsymbol{A}_k$, cf. (4.41). The difficulty in fulfilling the KKT conditions lies in the fact that each matrix $\boldsymbol{A}_k$ depends on all beamformers $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$.

The objective function in the optimization problem (4.34) is in general not concave in the beamformers. Furthermore, it is possible that multiple transmit strategies that result in different utility K-tuples fulfill the KKT conditions and that gradient searches converge to different local optima, depending on their initialization and step size. In fact, it is shown in [65] that the optimization problem (4.34) is NP-hard for the rate utility, the proportional fair rate utility, and the α-fair rate utility with $\alpha = 2$ if the transmitters have more than one antenna; this is remarkable because, as discussed in the previous chapter, in the SISO case both the proportional fair rate utility and the α-fair rate utility with $\alpha = 2$ guarantee a concave sum utility maximization problem, which can be solved efficiently.

An algorithm that finds the globally optimal solution to (4.34) with arbitrary precision was proposed for the case of $K = 2$ users in [66]; it is based on successively refining an approximation of the utility region by means of the polyblock algorithm, cf. [48]. Similar techniques can also be applied for more than two users, but their computational complexity grows extremely fast in the system dimensions and they do not scale well to larger networks. In this work, the focus therefore is on algorithms that are computationally simple and perform reasonably well, even though global sum utility optimality cannot be guaranteed.

4.3.5.1 Maximizing the Sum Rate

For the achievable rate utility function $u_k(\gamma_k) = R_k = \log(1+\gamma_k)$ the derivative w. r. t. the SINR γ_k is

$$u_k'(\gamma_k) = \frac{1}{1+\gamma_k} = \frac{\sum_{i\neq k}|\boldsymbol{h}_{ki}^{\mathrm{T}}\boldsymbol{v}_i|^2 + \sigma^2}{\sum_i |\boldsymbol{h}_{ki}^{\mathrm{T}}\boldsymbol{v}_i|^2 + \sigma^2} \qquad \forall k \in \{1,\ldots,K\}. \tag{4.43}$$

With (4.39) and (4.40) the derivatives of R_k and R_j w. r. t. the conjugate beamformer $\boldsymbol{v}_k^*$ are

$$\frac{\partial R_k}{\partial \boldsymbol{v}_k^*} = \frac{u_k'(\gamma_k)}{\sum_{i\neq k}|\boldsymbol{h}_{ki}^{\mathrm{T}}\boldsymbol{v}_i|^2 + \sigma^2} \cdot \boldsymbol{h}_{kk}^* \boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k = \frac{1}{\sum_i |\boldsymbol{h}_{ki}^{\mathrm{T}}\boldsymbol{v}_i|^2 + \sigma^2} \cdot \boldsymbol{h}_{kk}^* \boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k \tag{4.44}$$

and

$$\begin{aligned}
\frac{\partial R_j}{\partial \boldsymbol{v}_k^*} &= -\frac{u_j'(\gamma_j)|\boldsymbol{h}_{jj}^{\mathrm{T}}\boldsymbol{v}_j|^2}{\left(\sum_{i\neq j}|\boldsymbol{h}_{ji}^{\mathrm{T}}\boldsymbol{v}_i|^2 + \sigma^2\right)^2} \cdot \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}} \boldsymbol{v}_k \\
&= -\frac{|\boldsymbol{h}_{jj}^{\mathrm{T}}\boldsymbol{v}_j|^2}{\left(\sum_i |\boldsymbol{h}_{ji}^{\mathrm{T}}\boldsymbol{v}_i|^2 + \sigma^2\right)\left(\sum_{i\neq j}|\boldsymbol{h}_{ji}^{\mathrm{T}}\boldsymbol{v}_i|^2 + \sigma^2\right)} \cdot \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}} \boldsymbol{v}_k \\
&= -\left(\frac{1}{\sum_{i\neq j}|\boldsymbol{h}_{ji}^{\mathrm{T}}\boldsymbol{v}_i|^2 + \sigma^2} - \frac{1}{\sum_i |\boldsymbol{h}_{ji}^{\mathrm{T}}\boldsymbol{v}_i|^2 + \sigma^2}\right) \cdot \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}} \boldsymbol{v}_k
\end{aligned} \tag{4.45}$$

for $j \neq k$.

While the KKT conditions do not lead to a straightforward solution to the sum rate maximization problem, they do allow us to gain some insights into the optimal strategies at asymptotically low and high SNR. In the low-SNR regime, where $\sigma^2 \to \infty$, the noise power dominates the other terms in the denominators of (4.44) and (4.45). More precisely, we can use (4.44) and (4.45) to construct the matrix $\boldsymbol{A}_k$, cf. (4.42), and state that

$$\begin{aligned}
\lim_{\sigma^2\to\infty} \sigma^2 \boldsymbol{A}_k \boldsymbol{v}_k^{\mathrm{MF}} &= \boldsymbol{h}_{kk}^* \boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k^{\mathrm{MF}} - \lim_{\sigma^2\to\infty} \sum_{j\neq k} \frac{|\boldsymbol{h}_{jj}^{\mathrm{T}}\boldsymbol{v}_j|^2}{\sigma^2} \cdot \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}} \boldsymbol{v}_k^{\mathrm{MF}} \\
&= \boldsymbol{h}_{kk}^* \boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k^{\mathrm{MF}} = \boldsymbol{h}_{kk}^* \boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{h}_{kk}^* / \|\boldsymbol{h}_{kk}\|_2 = \|\boldsymbol{h}_{kk}\|_2^2 \boldsymbol{v}_k^{\mathrm{MF}}
\end{aligned} \tag{4.46}$$

i. e., the selfish beamformer $\boldsymbol{v}_k^{\mathrm{MF}}$ in the limit is an eigenvector of $\sigma^2\boldsymbol{A}_k$ corresponding to a positive eigenvalue. Thus, if every user employs its respective selfish beamformer, all KKT conditions are fulfilled. Furthermore, since $\sigma^2\boldsymbol{A}_k$ in the limit is constant and does not depend on the transmit strategy, the selfish solution is the only solution to the KKT conditions for which $\mu_k > 0$ for all $\{1,\ldots,K\}$. Additional candidates for optimality with $\mu_k = 0$ can be constructed, e. g., $\boldsymbol{v}_k = \boldsymbol{0}$; it is intuitive, though, that such strategies cannot be local maxima in the low-SNR regime, as a small increase in power can improve the own rate while decreasing the other users' rates by only a negligible amount.[2] Consequently, the selfish solution is the global optimum for sufficiently high noise power.

[2] To rigorously show that the zero beamformers cannot be optimal in the low-SNR regime, let us define the beamformer as $\boldsymbol{v}_k = \sqrt{p_k}\boldsymbol{v}_k'$, where $\boldsymbol{v}_k'$ is fixed with $\|\boldsymbol{v}_k'\|_2^2 = 1$. From (3.62) it can be seen that $\partial R_k/\partial p_k = \mathrm{O}(\sigma^{-2})$, whereas $-\partial R_j/\partial p_k = \mathrm{O}(\sigma^{-4})$ for $j \neq k$, as $\sigma^2 \to \infty$. Therefore, an increase in power will always lead to an increase in sum rate for sufficiently high σ^2.

The behavior of the rate R_k of user k in the high-SNR regime, i. e., for $\sigma^{-2} \to \infty$, depends on whether the received interference is exactly zero or not:

$$R_k = \begin{cases} \log(\sigma^{-2}) + \mathrm{O}(1) & \text{if} \quad \sum_{j \neq k} |\boldsymbol{h}_{kj}^{\mathrm{T}} \boldsymbol{v}_j|^2 = 0 \quad \text{and} \quad |\boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k|^2 > 0 \\ \mathrm{O}(1) & \text{otherwise.} \end{cases} \tag{4.47}$$

More precisely, we can state that

$$\lim_{\sigma^{-2} \to \infty} \frac{\sum_k R_k}{\log \sigma^{-2}} = D \tag{4.48}$$

where D is the number of users k for which $\sum_{j \neq k} |\boldsymbol{h}_{kj}^{\mathrm{T}} \boldsymbol{v}_j|^2 = 0$ and $|\boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k|^2 > 0$.[3] Therefore, to maximize the sum rate in the high-SNR regime, it is necessary to maximize the number of active users that do not experience any interference.

In the special case that zero-forcing is possible for all users, we can achieve $D = K$ and the unique altruistic strategy is globally optimal. In general, however, there may be multiple configurations of interference-free users that achieve the same D and are candidates for the high-SNR optimum. Also, the high-SNR optimum is not necessarily an altruistic solution.

4.4 Distributed Algorithms

Even when a centralized computer is available, determining the sum-utility optimal transmit strategy appears to be prohibitively complex for the utility functions of interest (unless the number of users K is small). With the additional constraints of our distributed computation model, i. e., computationally simple updates at each node and limited information exchange between the nodes, attempting to fully solve problem (4.34) seems to be impractical. Instead, the proposed distributed techniques are aimed at convergence towards local sum-utility optimality or based entirely on heuristics.

It is often assumed in the literature that the sum rate is the desired figure of merit. Some of the algorithms discussed in the following are designed to optimize other metrics, but are nonetheless attractive due to their simplicity. In this section we introduce the algorithms and discuss their properties, and then proceed to numerically evaluate their average performance in terms of the rate utility in the following section.

4.4.1 The Selfish Solution

The selfish solution

$$\boldsymbol{v}_k^{\mathrm{MF}} = \frac{1}{\|\boldsymbol{h}_{kk}^*\|_2} \cdot \boldsymbol{h}_{kk}^* \tag{4.49}$$

does not require iterative updates or information exchange between the users. Furthermore, user k must only know its own direct channel vector $\boldsymbol{h}_{kk}$; estimation of the cross channel vectors $\boldsymbol{h}_{kj}$ with $j \neq k$ is not necessary. As is intuitive and will become evident in the numerical evaluation, however, not taking into account the caused interference leads to severely suboptimal performance in many scenarios.

[3] The definition of D is very similar to the definition of the DoF in (2.85). The DoF, however, refer to the sum capacity of the channel, whereas here we use D to denote the number of non-interfering data streams of a given transmit strategy.

4.4.2 Maximization of the "Virtual SINR"

A more sophisticated non-iterative technique that takes into account the caused interference as well as the noise power was proposed in [67]: it is based on maximizing the *virtual SINR,* which is defined as

$$\xi_k = \frac{|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2}{\sum_{j\neq k}|\boldsymbol{h}_{jk}^{\mathrm{T}}\boldsymbol{v}_k|^2+\sigma^2}. \tag{4.50}$$

As opposed to the conventional SINR, the summands in the denominator of above expression are the interference power gains that transmitter k is causing, not the interference power gains that receiver k is experiencing. Consequently, the virtual SINR ξ_k does not depend on the beamformers $\boldsymbol{v}_j$ with $j \neq k$.

The beamformer $\boldsymbol{v}_k^{\mathrm{V}}$ that maximizes ξ_k can be found from the KKT conditions of the optimization problem

$$\boldsymbol{v}_k^{\mathrm{V}} = \arg\max_{\boldsymbol{v}_k} \ \xi_k \qquad \text{s. t.:} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1. \tag{4.51}$$

To begin with, it is clear that $\|\boldsymbol{v}_k^{\mathrm{V}}\|_2^2 = 1$ since multiplying the beamformer $\boldsymbol{v}_k$ with a scalar factor larger than one will lead to an increase in ξ_k. By taking the derivative of the objective function w. r. t. $\boldsymbol{v}_k^*$ using the quotient rule and omitting the denominator we obtain the first KKT condition

$$\left(\sum_{j\neq k}|\boldsymbol{h}_{jk}^{\mathrm{T}}\boldsymbol{v}_k|^2+\sigma^2\right)\boldsymbol{h}_{kk}^*\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k - |\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2\sum_{j\neq k}\boldsymbol{h}_{jk}^*\boldsymbol{h}_{jk}^{\mathrm{T}}\boldsymbol{v}_k = \mu\boldsymbol{v}_k \tag{4.52}$$

where the Lagrange multiplier μ must be non-negative. Dividing both sides by $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2$ yields

$$\frac{1}{\xi_k}\boldsymbol{h}_{kk}^*\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k - \sum_{j\neq k}\boldsymbol{h}_{jk}^*\boldsymbol{h}_{jk}^{\mathrm{T}}\boldsymbol{v}_k = \mu'\boldsymbol{v}_k \tag{4.53}$$

where again μ' must be non-negative. By multiplying from the left with $\boldsymbol{v}_k^{\mathrm{H}}$ we obtain

$$\frac{1}{\xi_k}|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2 = \sum_{j\neq k}|\boldsymbol{h}_{jk}^{\mathrm{T}}\boldsymbol{v}_k|^2+\mu' \tag{4.54}$$

and it is clear that $\mu' = \sigma^2$. Consequently, the optimal beamformer fulfills

$$\frac{1}{\xi_k}\boldsymbol{h}_{kk}^*\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k = \left(\sum_{j\neq k}\boldsymbol{h}_{jk}^*\boldsymbol{h}_{jk}^{\mathrm{T}}+\sigma^2\mathbf{I}\right)\boldsymbol{v}_k \tag{4.55}$$

and is therefore the unit-norm eigenvector corresponding to the only positive eigenvalue of the rank-one matrix

$$\boldsymbol{X}_k = \left(\sum_{j\neq k}\boldsymbol{h}_{jk}^*\boldsymbol{h}_{jk}^{\mathrm{T}}+\sigma^2\mathbf{I}\right)^{-1}\boldsymbol{h}_{kk}^*\boldsymbol{h}_{kk}^{\mathrm{T}} \tag{4.56}$$

i. e.,

$$\boldsymbol{v}_k^{\mathrm{V}} = \frac{\left(\sum_{j\neq k}\boldsymbol{h}_{jk}^*\boldsymbol{h}_{jk}^{\mathrm{T}}+\sigma^2\mathbf{I}\right)^{-1}\boldsymbol{h}_{kk}^*}{\left\|\left(\sum_{j\neq k}\boldsymbol{h}_{jk}^*\boldsymbol{h}_{jk}^{\mathrm{T}}+\sigma^2\mathbf{I}\right)^{-1}\boldsymbol{h}_{kk}^*\right\|_2}. \tag{4.57}$$

Just as the selfish solution, the virtual-SINR optimal solution does not require iterative updates. It is, however, necessary to exchange the cross channel information among the users, so that each user k knows $\boldsymbol{h}_{jk}$ for all $j \in \{1, \ldots, K\}$.

Remarkably, the virtual SINR maximizing beamformer $\boldsymbol{v}_k^{\mathrm{V}}$ converges towards the selfish beamformer $\boldsymbol{v}_k^{\mathrm{MF}}$ in the low-SNR regime, i. e., when the noise power σ^2 is asymptotically high and the matrix inverse approaches a scaled identity. Also, when the channel vectors permit zero-forcing for user k, $\boldsymbol{v}_k^{\mathrm{V}}$ converges towards the zero-forcing beamformer in the high-SNR regime. To show this, we apply the matrix-inversion lemma (e. g., [37, Section 0.7.4]) to the inverse in (4.57):

$$\left(\sum_{j\neq k} \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}} + \sigma^2 \mathbf{I}\right)^{-1} = \left(\boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{H}_k + \sigma^2 \mathbf{I}\right)^{-1} \tag{4.58}$$

$$= \sigma^{-2}\left(\mathbf{I} - \boldsymbol{H}_k^{\mathrm{H}}\left(\boldsymbol{H}_k \boldsymbol{H}_k^{\mathrm{H}} + \sigma^2 \mathbf{I}\right)^{-1} \boldsymbol{H}_k\right) \tag{4.59}$$

which approaches a scaled version of the projection matrix $\boldsymbol{\Pi}_k$ from (4.22) as $\sigma^{-2} \to \infty$. Therefore, the strategy of maximizing the virtual SINR is sum-rate optimal at low SNR and, in scenarios where zero-forcing is possible for all users, also at high SNR.

For $K = 2$ the resulting strategy is in fact Pareto optimal regardless of the noise power σ^2: since $\gamma_1 \cdot \gamma_2 = \xi_1 \cdot \xi_2$ and consequently $\log \gamma_1 + \log \gamma_2 = \log \xi_1 + \log \xi_2$, it is clear that the strategy of maximal individual virtual SINRs is also sum-utility optimal in the logarithmic utility and therefore on the Pareto boundary of the SINR region.

The disadvantage of this method comes into play in scenarios in which not all users can perform zero-forcing simultaneously: the virtual SINR maximization always results in full power for all users, whereas the sum-rate optimal strategy at high SNR involves deactivating some users, so that the active users can enjoy interference-free transmission.

4.4.3 Maximization of the "Global SINR"

Another SINR-related objective function was proposed for MIMO systems in [68], but can be applied to single-antenna receivers as well. The goal is to maximize the ratio of the sum of all desired power gains to the sum of all interference power gains and noise powers, i. e., the sum of all numerators divided by the sum of all denominators of the individual SINRs:

$$\xi_{\mathrm{sum}} = \frac{\sum_k |\boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k|^2}{\sum_k \sum_{j\neq k} |\boldsymbol{h}_{jk}^{\mathrm{T}} \boldsymbol{v}_k|^2 + K\sigma^2}. \tag{4.60}$$

It was shown in [68] that

$$\sum_k \log(1 + \gamma_k) \geq \log(1 + \xi_{\mathrm{sum}}) \tag{4.61}$$

i. e., that by maximizing ξ_{sum} a lower bound on the sum rate is maximized.

Unlike the two previously discussed strategies, the problem of maximizing ξ_{sum} is not separable into K independent sub-problems, each of which depends only on one beamformer $\boldsymbol{v}_k$. Therefore, the value of ξ_{sum} is increased iteratively by repeatedly updating all beamformers $\boldsymbol{v}_k$, either sequentially or in parallel.

In iteration ℓ, the beamformer of user k is updated to fulfill

$$\boldsymbol{v}_k^{(\ell+1)} = \arg\max_{\boldsymbol{v}_k} \; \xi_{\mathrm{sum}}\Big|_{\boldsymbol{v}_j = \boldsymbol{v}_j^{(\ell)} \; \forall j \neq k} \qquad \text{s. t.:} \quad \|\boldsymbol{v}_k\|_2^2 = 1. \tag{4.62}$$

Note that the power constraint must be fulfilled with equality; it would also be possible to formulate the update using an inequality power constraint, but the resulting algorithm turns out to have some undesirable properties, as discussed below. With the definitions

$$a_k^{(\ell)} = \sum_{j \neq k} \left| \boldsymbol{h}_{jj}^{\mathrm{T}} \boldsymbol{v}_j^{(\ell)} \right|^2 \tag{4.63}$$

$$b_k^{(\ell)} = \sum_{j \neq k} \sum_{i \neq j} \left| \boldsymbol{h}_{ij}^{\mathrm{T}} \boldsymbol{v}_j^{(\ell)} \right|^2 + K\sigma^2 \tag{4.64}$$

we can express the objective function of the optimization problem as

$$\xi_{\mathrm{sum}} \Big|_{\boldsymbol{v}_j = \boldsymbol{v}_j^{(\ell)} \; \forall j \neq k} = \frac{\boldsymbol{v}_k^{\mathrm{H}} \boldsymbol{h}_{kk}^* \boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k + a_k^{(\ell)}}{\boldsymbol{v}_k^{\mathrm{H}} \cdot \sum_{j \neq k} \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}} \cdot \boldsymbol{v}_k + b_k^{(\ell)}} = \frac{\boldsymbol{v}_k^{\mathrm{H}} \left(\boldsymbol{h}_{kk}^* \boldsymbol{h}_{kk}^{\mathrm{T}} + a_k^{(\ell)} \mathbf{I} \right) \boldsymbol{v}_k}{\boldsymbol{v}_k^{\mathrm{H}} \left(\sum_{j \neq k} \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}} + b_k^{(\ell)} \mathbf{I} \right) \boldsymbol{v}_k}. \tag{4.65}$$

Since the objective function is a *generalized Rayleigh quotient* (cf. [37, Section 4.2]), the solution $\boldsymbol{v}_k^{(\ell+1)}$ can be shown (cf. [68]) to be the principal eigenvector of the matrix

$$\boldsymbol{X}_k^{(\ell)} = \left(\sum_{j \neq k} \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}} + b_k^{(\ell)} \mathbf{I} \right)^{-1} \left(\boldsymbol{h}_{kk}^* \boldsymbol{h}_{kk}^{\mathrm{T}} + a_k^{(\ell)} \mathbf{I} \right). \tag{4.66}$$

In order to compute the update, user k must know the channel vectors $\boldsymbol{h}_{jk}$ for all $j \in \{1, \ldots, K\}$, but also the current values of $a_k^{(\ell)}$ and $b_k^{(\ell)}$. To this end, it is necessary that the users broadcast both their received signal power and their received total interference power to all other users after each update, i. e., user k must announce both the numerator $\left| \boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k^{(\ell)} \right|^2$ and the denominator $\sum_{j \neq k} \left| \boldsymbol{h}_{kj}^{\mathrm{T}} \boldsymbol{v}_j^{(\ell)} \right|^2 + \sigma^2$ of $\gamma_k^{(\ell)}$ before iteration $\ell + 1$.

When the updates are performed sequentially, it is clear that a beamformer update cannot lead to a decrease ξ_{sum}; therefore, the sequence of ξ_{sum} over the iterations is non-decreasing and must converge. For parallel updates, on the other hand, monotonicity of the objective ξ_{sum} cannot be guaranteed, as each user attempts to increase ξ_{sum} under the false assumption that the other users' beamformers remain constant. While neither convergence towards the globally optimal value of ξ_{sum} nor convergence in terms of the beamforming vectors has been proven, the numerical evaluations in Section 4.5 show reliable convergence after a small number of iterations.

Similar to the virtual-SINR optimal strategy, this method is by design not capable of deactivating individual users if it is in the common interest. Therefore, it cannot be sum-rate optimal at high SNR in settings where zero-forcing is not possible. The capability to deactivate users for the common good can be added by replacing the equality power constraint by an inequality power constraint in (4.62); however, the interference terms $|\boldsymbol{h}_{kj} \boldsymbol{v}_j|^2$ for an inactive user $j \neq k$ are still part of the denominator of ξ_{sum}, and consequently the remaining active users would attempt to minimize the interference caused to both the inactive and the active users when computing their updated beamformers. This effect in fact leads to significantly lower sum-rate performance than with an equality power constraint in many cases.

4.4.4 Minimization of the Sum Mean Squared Error

In [69, 68, 70], the use of the sum MSE cost function for MIMO interference networks was explored. In the following we apply the resulting distributed algorithm to the special case of single-antenna receivers.

As discussed in Section 3.3.5, the MSE of user k is the expected power of the error between the transmitted data symbol s_k and its estimate at receiver k, denoted by $\hat{s}_k$:

$$\varepsilon_k = \mathrm{E}\left[|\hat{s}_k - s_k|^2\right]. \tag{4.67}$$

The estimated symbol $\hat{s}_k$ is obtained by multiplying the received signal y_k with a scalar factor $g_k \in \mathbb{C}$ that compensates the gain resulting from the combination of beamformer and channel:

$$\hat{s}_k = g_k y_k = g_k \left(\sum_j \boldsymbol{h}_{kj}^{\mathrm{T}} \boldsymbol{v}_j s_j + n_k \right). \tag{4.68}$$

Since the transmit symbols have mean zero and are uncorrelated, the MSE of user k is

$$\varepsilon_k = \sum_j |g_k \boldsymbol{h}_{kj}^{\mathrm{T}} \boldsymbol{v}_j|^2 - 2\,\mathrm{Re}\{g_k \boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k\} + 1 + |g_k|^2 \sigma^2 \tag{4.69}$$

and the sum MSE can be expressed as

$$\sum_k \varepsilon_k = \sum_k \sum_j |g_k \boldsymbol{h}_{kj}^{\mathrm{T}} \boldsymbol{v}_j|^2 - \sum_k 2\,\mathrm{Re}\{g_k \boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k\} + K + \sum_k |g_k|^2 \sigma^2. \tag{4.70}$$

The sum MSE depends on both the receiver gains $g_1, \ldots, g_K$ and the beamformers $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$. The proposed iterative algorithm therefore requires both updates of the receiver gains and the beamformers. We begin with the receiver gain update:

$$g_k^{(\ell+1)} = \arg\min_{g_k} \; \varepsilon_k \Big|_{\boldsymbol{v}_j = \boldsymbol{v}_j^{(\ell)} \; \forall j.} \tag{4.71}$$

Note that the other users' MSEs ε_j with $j \neq k$ do not depend on g_k, so that the above optimization also minimizes the sum MSE. The solution, which can be found by setting the derivative to zero, is

$$g_k^{(\ell+1)} = \frac{{\boldsymbol{v}_k^{(\ell)}}^{\mathrm{H}} \boldsymbol{h}_{kk}^*}{\sum_j \left| \boldsymbol{h}_{kj}^{\mathrm{T}} \boldsymbol{v}_j^{(\ell)} \right|^2 + \sigma^2} \tag{4.72}$$

where, to avoid confusion, it should be noted that the denominator contains not only the interference power gains, but also the desired power gain.

The problem that must be solved for the beamformer update is

$$\boldsymbol{v}_k^{(\ell+1)} = \arg\min_{\boldsymbol{v}_k} \; \sum_j \varepsilon_j \Big|_{g_i = g_i^{(\ell)} \; \forall i} \qquad \text{s.t.:} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1. \tag{4.73}$$

Since there are no summands in the sum MSE (4.70) that contain both $\boldsymbol{v}_k$ and another beamformer $\boldsymbol{v}_j$ with $j \neq k$, the solution does not depend on the beamformers of other users. With the Lagrangian multiplier λ, the optimality conditions are

$$\left(\sum_j \left| g_j^{(\ell)} \right|^2 \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}} + \lambda \mathbf{I} \right) \boldsymbol{v}_k^{(\ell+1)} = {g_k^{(\ell)}}^* \boldsymbol{h}_{kk}^* \tag{4.74}$$

as well as $\lambda \geq 0$, $\left\|\boldsymbol{v}_k^{(\ell+1)}\right\|_2^2 \leq 1$, and $\lambda\left(1 - \left\|\boldsymbol{v}_k^{(\ell+1)}\right\|_2^2\right) = 0$. We must therefore examine two possibilities:

1) $\lambda = 0$ and $\left\|\boldsymbol{v}_k^{(\ell+1)}\right\|_2^2 \leq 1$.
First of all, it can be observed that even if the matrix on the left-hand side of (4.74) is not invertible, there is always a solution, as the vector on the right-hand side lies in the span of the matrix on the left-hand side. Furthermore, by stacking the row vectors $g_j^{(\ell)}\boldsymbol{h}_{jk}^{\mathrm{T}}$ for all $j \in \{1,\ldots,K\}$ in the matrix $\boldsymbol{G}_k^{(\ell)}$ and examining the singular value decomposition of $\boldsymbol{G}_k^{(\ell)}$, it can be shown that even if there are multiple solutions to (4.74), they are all equivalent in terms of sum MSE and the one with the lowest power is

$$\hat{\boldsymbol{v}}_k^{(\ell+1)} = g_k^{(\ell)*}\left(\sum_j \left|g_j^{(\ell)}\right|^2 \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}}\right)^{+} \boldsymbol{h}_{kk}^*. \tag{4.75}$$

2) $\lambda > 0$ and $\left\|\boldsymbol{v}_k^{(\ell+1)}\right\|_2^2 = 1$.
In this case the matrix on the left-hand side of (4.74) is invertible and the beamformer can be expressed as

$$\check{\boldsymbol{v}}_k^{(\ell+1)} = g_k^{(\ell)*}\left(\sum_j \left|g_j^{(\ell)}\right|^2 \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}} + \lambda\mathbf{I}\right)^{-1} \boldsymbol{h}_{kk}^*. \tag{4.76}$$

Determining the value of λ for which $\check{\boldsymbol{v}}_k^{(\ell+1)}$ has unit power is not possible in closed form. From

$$\left\|\check{\boldsymbol{v}}_k^{(\ell+1)}\right\|_2^2 = \left|g_k^{(\ell)}\right|^2 \boldsymbol{h}_{kk}^{\mathrm{T}}\left(\sum_j \left|g_j^{(\ell)}\right|^2 \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}} + \lambda\mathbf{I}\right)^{-2} \boldsymbol{h}_{kk}^* \tag{4.77}$$

it is, however, clear that the power of the beamformer is decreasing in λ and tends towards zero for $\lambda \to \infty$. Furthermore, it can again be seen with the singular value decomposition of $\boldsymbol{G}_k^{(\ell)}$ that, for $\lambda \to 0$, $\check{\boldsymbol{v}}_k^{(\ell+1)}$ approaches $\hat{\boldsymbol{v}}_k^{(\ell+1)}$. Consequently, if the power of $\hat{\boldsymbol{v}}_k^{(\ell+1)}$ is strictly larger than one, there is a unique value of $\lambda > 0$ that fulfills the optimality conditions. The behavior of the beamformer power over λ is illustrated in Figure 4.2.

With these observations, the optimal updated beamformer $\boldsymbol{v}_k^{(\ell+1)}$ can be determined as follows: first, $\hat{\boldsymbol{v}}_k^{(\ell+1)}$ is computed; if the norm of $\hat{\boldsymbol{v}}_k^{(\ell+1)}$ is less than or equal to one, the updated beamformer is $\boldsymbol{v}_k^{(\ell+1)} = \hat{\boldsymbol{v}}_k^{(\ell+1)}$. If the norm is larger than one, λ must be strictly greater than zero and must be determined with a line search, such that the norm of $\check{\boldsymbol{v}}_k^{(\ell+1)}$ is equal to one. The updated beamformer then is $\boldsymbol{v}_k^{(\ell+1)} = \check{\boldsymbol{v}}_k^{(\ell+1)}$.

In order to determine λ efficiently, we propose to use Newton's method. We define

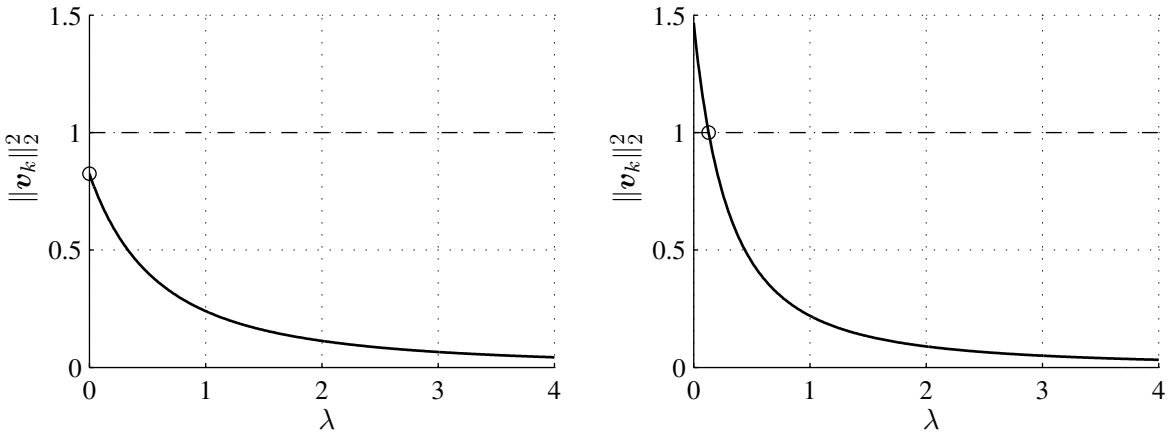
$$f(\lambda) = \left\|\check{\boldsymbol{v}}_k^{(\ell+1)}\right\|_2^2 - 1 \tag{4.78}$$

and search for the root of $f(\lambda)$ by iteratively updating λ such that

$$\lambda_{\mathrm{new}} = \lambda_{\mathrm{old}} - \frac{f\left(\lambda_{\mathrm{old}}\right)}{f'\left(\lambda_{\mathrm{old}}\right)} \tag{4.79}$$

where $f'(\lambda) = \mathrm{d}f(\lambda)/\,\mathrm{d}\lambda$. The Newton update in our case can be explicitly stated as

$$\lambda_{\mathrm{new}} = \lambda_{\mathrm{old}} + \frac{|g_k|^2 \boldsymbol{h}_{kk}^{\mathrm{T}}\left(\sum_j |g_j|^2 \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}} + \lambda_{\mathrm{old}}\mathbf{I}\right)^{-2} \boldsymbol{h}_{kk}^* - 1}{2|g_k|^2 \boldsymbol{h}_{kk}^{\mathrm{T}}\left(\sum_j |g_j|^2 \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}} + \lambda_{\mathrm{old}}\mathbf{I}\right)^{-3} \boldsymbol{h}_{kk}^*} \tag{4.80}$$



(a) $\lambda = 0$ and $\left\|\boldsymbol{v}_k^{(\ell+1)}\right\|_2^2 \leq 1$

(b) $\lambda > 0$ and $\left\|\boldsymbol{v}_k^{(\ell+1)}\right\|_2^2 = 1$

Figure 4.2: Two possible cases for the behavior of the beamformer power over the regularization parameter λ. The circle marks the solution to the KKT conditions.

where we omitted the iteration indices for brevity. As an initial value, $\lambda = 0$ appears to be a suitable choice.[4] The Newton method converges reliably in a small number of iterations.

The sum MSE minimization algorithm consists of updates of the receiver coefficients $g_1, \ldots, g_K$ and beamformers $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$. An update of a receiver coefficient is only necessary if at least one of the beamformers has changed since the last update; similarly, an update of a beamformer is only necessary if at least one of the receiver coefficients has changed. Therefore, the algorithm must alternate between receiver and transmitter updates. For sequential operation of the algorithm, the updates are performed in the following order: $g_1, \boldsymbol{v}_1, \ldots, g_K, \boldsymbol{v}_K$; for parallel updates, the order is: $g_1, \ldots, g_K, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$.

In conclusion, we remark on some properties of the sum MSE minimization algorithm:

- If $\boldsymbol{v}_k^{(\ell)} = \boldsymbol{0}$, then $g_k^{(\ell+1)} = 0$, and vice versa. Therefore, if a beamformer is initially zero, it will not be activated in the course of the algorithm.
- Likewise, a beamformer that is non-zero cannot be set to zero from one iteration to the next. The power of a beamformer can, however, be gradually reduced and approach zero.
- Regardless of whether the updates are performed sequentially or in parallel, the sum MSE is never increased by an update. Furthermore, the sum MSE is bounded from below by zero. Therefore, the sum MSE must converge over the iterations.
- Furthermore, if the algorithm reaches a stationary K-tuple of beamformers $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K)$, it is straightforward to show that the KKT conditions of the sum MSE minimization problem

$$\min_{\substack{g_1,\ldots,g_K \\ \boldsymbol{v}_1,\ldots,\boldsymbol{v}_K}} \sum_k \varepsilon_k \qquad \text{s.t.:} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1 \quad \forall k \in \{1, \ldots, K\} \tag{4.81}$$

  are fulfilled.
- For the update of g_k, user k must know the total received signal power at receiver k as well as the combined gain of beamformer and direct channel $\boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k^{(\ell)}$; we assume that these two quantities can be straightforwardly estimated by receiver k. For the update of $\boldsymbol{v}_k$, user k must

[4] For $\lambda_{\mathrm{old}} = 0$ it can be necessary to replace the inverse with the pseudo-inverse in (4.80).

know the channel vectors $\boldsymbol{h}_{jk}$ for all $j \in \{1,\ldots,K\}$ as well as the receiver coefficients $g_j^{(\ell)}$ for all $j \in \{1,\ldots,K\}$. While the channel information can be exchanged during the initialization of the algorithm, the receiver coeffecients must be exchanged among the users at each iteration.

4.4.4.1 Weighted Sum MSE Minimization with Adaptive Weights

The sum MSE minimization algorithm can be easily extended to allow for assigning different priorities to the users by means of real non-negative weights $\alpha_1,\ldots,\alpha_K$. The objective is to minimize the weighted sum of MSEs, i. e.,

$$\min_{\substack{g_1,\ldots,g_K\\ \boldsymbol{v}_1,\ldots,\boldsymbol{v}_K}} \sum_k \alpha_k \varepsilon_k \qquad \text{s. t.:} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1 \quad \forall k \in \{1,\ldots,K\}. \tag{4.82}$$

A higher weight α_k of user k therefore has the effect that minimizing the MSE ε_k at receiver k has a higher priority. By following the same steps as for the previously discussed unweighted sum MSE objective, it is straightforward to show

- that the receiver coefficient update is identical to the unweighted case (4.72) and
- that, assuming $\alpha_k > 0$, the updated beamformer must fulfill

$$\left(\sum_j \frac{\alpha_j}{\alpha_k} \left|g_j^{(\ell)}\right|^2 \boldsymbol{h}_{jk}^* \boldsymbol{h}_{jk}^{\mathrm{T}} + \lambda \mathbf{I}\right) \boldsymbol{v}_k^{(\ell+1)} = g_k^{(\ell)*} \boldsymbol{h}_{kk}^* \tag{4.83}$$

instead of (4.74) in the unweighted case. The subsequent discussion on determining $\boldsymbol{v}_k^{(\ell+1)}$ and λ, cf. (4.75)–(4.80), can be applied to the weighted sum MSE in the same way, where the channel vectors $\boldsymbol{h}_{jk}$ are additionally weighted with a factor $\sqrt{\alpha_j/\alpha_k}$.

In [69], we proposed a method of adaptively choosing the weights $\alpha_1,\ldots,\alpha_K$ to mimic the behavior of arbitrary utility functions close to the current operating point (also cf. [71] for a different system model). To derive this method, we begin by pointing out that the MSE of user k with the optimal receiver coefficient g_k fulfilling (4.72) is

$$\begin{aligned}\varepsilon_k &= \frac{|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2 \sum_j |\boldsymbol{h}_{kj}^{\mathrm{T}}\boldsymbol{v}_j|^2}{\left(\sum_j |\boldsymbol{h}_{kj}^{\mathrm{T}}\boldsymbol{v}_j|^2 + \sigma^2\right)^2} - \frac{2|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2}{\sum_j |\boldsymbol{h}_{kj}^{\mathrm{T}}\boldsymbol{v}_j|^2 + \sigma^2} + 1 + \frac{|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2\sigma^2}{\left(\sum_j |\boldsymbol{h}_{kj}^{\mathrm{T}}\boldsymbol{v}_j|^2 + \sigma^2\right)^2} \\ &= \frac{\sum_{j\neq k} |\boldsymbol{h}_{kj}^{\mathrm{T}}\boldsymbol{v}_j|^2 + \sigma^2}{\sum_j |\boldsymbol{h}_{kj}^{\mathrm{T}}\boldsymbol{v}_j|^2 + \sigma^2} = \frac{1}{1+\gamma_k}\end{aligned} \tag{4.84}$$

and note that this relationship holds after each update of receiver k. In Section 3.3.5, the same is stated for the SISO case. Now we can express the SINR (after an update of receiver k) in terms of the MSE

$$\gamma_k = \frac{1}{\varepsilon_k} - 1 \tag{4.85}$$

and the utility of user k can also be seen as a function of the MSE ε_k. We proceed to formulate a first-order Taylor expansion of the sum utility around an arbitrary operating point that is described by the SINR K-tuple $(\bar{\gamma}_1,\ldots,\bar{\gamma}_K)$ or, equivalently, the MSE K-tuple $(\bar{\varepsilon}_1,\ldots,\bar{\varepsilon}_K)$:

$$\sum_k u_k(\gamma_k) = \sum_k u_k(\bar{\gamma}_k) + \sum_k \left.\frac{\partial u_k}{\partial \varepsilon_k}\right|_{\varepsilon_k=\bar{\varepsilon}_k} \cdot (\varepsilon_k - \bar{\varepsilon}_k) + r_1(\varepsilon_1,\ldots,\varepsilon_K). \tag{4.86}$$

Thus, by choosing the weights

$$\alpha_k = -\frac{\partial u_k}{\partial \varepsilon_k}\bigg|_{\varepsilon_k = \bar{\varepsilon}_k} \tag{4.87}$$

minimizing the weighted sum MSE is equivalent to maximizing the sum utility minus the residual term $r_1(\varepsilon_1, \ldots, \varepsilon_K)$. If the MSE K-tuple $(\bar{\varepsilon}_1, \ldots, \bar{\varepsilon}_K)$ results from the current beamformer K-tuple $\left(\boldsymbol{v}_1^{(\ell)}, \ldots, \boldsymbol{v}_K^{(\ell)}\right)$, the residual term $r_1(\varepsilon_1, \ldots, \varepsilon_K)$ is zero at the operating point and can be considered negligible for small changes in the transmit strategy.

Different "schedules" for updating the weights come into question: for instance, $\alpha_1, \ldots, \alpha_K$ could be computed after every transmitter/receiver update, after a fixed number of iterations, or even after the weighted sum MSE updates have converged. Each time the weights are updated, however, the new weights must be communicated among the users, as for the update of beamformer $\boldsymbol{v}_k$ all α_j must be known.

At a stationary transmit strategy of the adaptively weighted sum MSE minimization algorithm, the receiver coefficients $g_1, \ldots, g_K$, the beamformers $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$, and the weights $\alpha_1, \ldots, \alpha_K$ do not change from one iteration to the next. We can then state that the beamformers fulfill the KKT conditions of the problem

$$\max_{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K} -\sum_k \alpha_k \varepsilon_k \qquad \text{s. t.:} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1 \quad \forall k \in \{1, \ldots, K\} \tag{4.88}$$

where we implicitly assumed that the optimal receivers $g_1, \ldots, g_K$ are used. Also, we can express the derivative of the objective function of (4.88) w. r. t. a beamformer $\boldsymbol{v}_k^*$ as

$$-\sum_j \alpha_j \frac{\partial \varepsilon_j}{\partial \boldsymbol{v}_k^*} = \sum_j \frac{\partial u_j}{\partial \varepsilon_j}\bigg|_{\varepsilon_j = \bar{\varepsilon}_j} \cdot \frac{\partial \varepsilon_j}{\partial \boldsymbol{v}_k^*}. \tag{4.89}$$

Since for a stationary transmit strategy the MSE K-tuple $(\bar{\varepsilon}_1, \ldots, \bar{\varepsilon}_K)$ results from the beamformer K-tuple $\left(\boldsymbol{v}_1^{(\ell)}, \ldots, \boldsymbol{v}_K^{(\ell)}\right)$, it follows that

$$-\sum_j \alpha_j \frac{\partial \varepsilon_j}{\partial \boldsymbol{v}_k^*}\bigg|_{\boldsymbol{v}_i = \boldsymbol{v}_i^{(\ell)}\ \forall i} = \sum_j \frac{\partial u_j}{\partial \boldsymbol{v}_k^*}\bigg|_{\boldsymbol{v}_i = \boldsymbol{v}_i^{(\ell)}\ \forall i.} \tag{4.90}$$

As a consequence, at a stationary transmit strategy the KKT conditions of (4.88) are equivalent to the KKT conditions (4.35)–(4.38) of the sum utility problem; a stationary transmit strategy of the adaptively weighted sum MSE minimization algorithm therefore is a candidate for local sum utility optimality.

Furthermore, for utility functions that are convex in ε_k, it is known that the residual term $r_1(\varepsilon_1, \ldots, \varepsilon_K)$ is non-negative, i. e., the linear approximation of the sum utility by means of the negative weighted sum MSE is an under-estimation of the true sum utility that is tight in the operating point. Thus, decreasing the weighted sum MSE by means of one or more beamformer updates results in an increase of sum utility; consequently, the sum utility is non-decreasing from one weight update to the next and thus converges. By computing the second derivative of the utility function w. r. t. the MSE it can be shown that a utility function $u_k(\gamma_k)$ is convex in ε_k if and only if $c_k(\gamma_k) \leq 2\gamma_k/(1+\gamma_k)$, i. e., the coefficient $c_k(\gamma_k)$ of the utility function may not exceed the coefficient of the negative MSE utility function given in (3.34). As can be seen from Figure 3.3, of the utility functions discussed in Section 3.3 only the rate utility fulfills this condition.

For the achievable rate utility $R_k = \log(1+\gamma_k) = -\log \varepsilon_k$ the weights are simply $\alpha_k = 1/\bar{\varepsilon}_k$; also, R_k is clearly convex in ε_k, and the sum rate therefore converges with the adaptively weighted sum MSE algorithm.

4.4.5 Interference Pricing

We recall from the SISO case that the interference price of user k is the derivative of $u_k(\gamma_k)$ w. r. t. the interference power evaluated at the current operating point, cf. (3.88). The objective of the transmit strategy update of user k in the interference pricing algorithm is to maximize the own utility $u_k(\gamma_k)$ minus the "cost" of causing interference to the other users, which can be interpreted as a linearization of the other users' utility functions by means of the interference prices, cf. (3.90). The same concept can be applied to the MISO case; the resulting update problem for user k is

$$\boldsymbol{v}_k^{(\ell+1)} = \arg\max_{\boldsymbol{v}_k} \; u_k(\gamma_k)\Big|_{\boldsymbol{v}_j=\boldsymbol{v}_j^{(\ell)}\;\forall j\neq k} - \sum_{j\neq k} \pi_j^{(\ell)} |\boldsymbol{h}_{jk}^{\mathrm{T}}\boldsymbol{v}_k|^2 \qquad \text{s. t.:} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1 \tag{4.91}$$

where

$$\pi_j^{(\ell)} = -\frac{\partial u_j(\gamma_j)}{\partial |\boldsymbol{h}_{jk}^{\mathrm{T}}\boldsymbol{v}_k|^2}\bigg|_{\boldsymbol{v}_i=\boldsymbol{v}_i^{(\ell)}\;\forall i} \qquad \text{for any } k \neq j. \tag{4.92}$$

If $u_k(\gamma_k)$ is concave in γ_k (which corresponds to $c_k(\gamma_k) \geq 0$), it follows that $u_k(\gamma_k)$ is also concave in $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2 = \boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{Q}_k\boldsymbol{h}_{kk}^*$, and thus also in $\boldsymbol{Q}_k$. Since furthermore the set of all positive semi-definite matrices $\boldsymbol{Q}_k$ that fulfill $\mathrm{tr}(\boldsymbol{Q}_k) \leq 1$ is a convex set, the pricing update is a concave maximization problem when formulated in terms of the covariance matrix $\boldsymbol{Q}_k = \boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{H}}$; this observation is consistent with the SISO case, where the update problem is a concave maximization problem in the transmit power p_k for $c_k(\gamma_k) \geq 0$.

While the objective function is not necessarily concave in the beamformer $\boldsymbol{v}_k$, formulating the problem in terms of $\boldsymbol{v}_k$ yields a solution procedure that is fairly simple and involves at most a line search. In the following we describe the resulting update of the beamformer $\boldsymbol{v}_k$ and refer to Appendix A6 for the derivation. As in the SISO case, we require the utility function to fulfill $-u_k''(\gamma_k) \geq C > 0$ for all feasible SINRs γ_k.

For ease of notation, we abbreviate the power gain of the desired signal as

$$\zeta = |\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2 \tag{4.93}$$

and the derivative of $u_k(\gamma_k)$ w. r. t. ζ as

$$\rho(\zeta) = u_k'(\gamma_k)\Big|_{\substack{\boldsymbol{v}_j=\boldsymbol{v}_j^{(\ell)}\;\forall j\neq k \\ |\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2=\zeta}} \cdot \frac{1}{\sum_{j\neq k}\left|\boldsymbol{h}_{kj}^{\mathrm{T}}\boldsymbol{v}_j^{(\ell)}\right|^2 + \sigma^2} = u_k'(\gamma_k)\Big|_{\substack{\boldsymbol{v}_j=\boldsymbol{v}_j^{(\ell)}\;\forall j\neq k \\ |\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2=\zeta}} \cdot \frac{\gamma_k^{(\ell)}}{\left|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k^{(\ell)}\right|^2}. \tag{4.94}$$

We furthermore define the matrix $\boldsymbol{B}^{\mathrm{H}}$ as the stacked row vectors $\sqrt{\pi_j^{(\ell)}}\boldsymbol{h}_{jk}^{\mathrm{T}}$ for all $j \neq k$ and the matrix

$$\boldsymbol{A}(\rho) = \rho \cdot \boldsymbol{h}_{kk}^*\boldsymbol{h}_{kk}^{\mathrm{T}} - \sum_{j\neq k} \pi_j^{(\ell)} \boldsymbol{h}_{jk}^*\boldsymbol{h}_{jk}^{\mathrm{T}} = \rho \cdot \boldsymbol{h}_{kk}^*\boldsymbol{h}_{kk}^{\mathrm{T}} - \boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}. \tag{4.95}$$

We assume that $\boldsymbol{w}$ is the unit-norm eigenvector of $\boldsymbol{A}$ corresponding to the largest eigenvalue λ; if λ has multiplicity higher than one, $\boldsymbol{w}$ is the eigenvector that maximizes $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2$. The reduced singular value decomposition of $\boldsymbol{B}$ is $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}$, where $\boldsymbol{\Sigma}$ is invertible, $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U} = \mathbf{I}$, and $\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V} = \mathbf{I}$.

In the following we distinguish between the case where zero-forcing is possible, i. e., where $\boldsymbol{h}_{kk}^*$ is linearly independent of the columns of $\boldsymbol{B}$, and the case where zero-forcing is not possible, i. e., where $\boldsymbol{h}_{kk}^*$ lies in the span of the columns of $\boldsymbol{B}$. The condition can be checked by evaluating $(\mathbf{I} - \boldsymbol{U}\boldsymbol{U}^{\mathrm{H}})\boldsymbol{h}_{kk}^*$: if the result is the zero vector, zero-forcing is not possible.

If zero-forcing is possible, the updated beamformer has full power and can be found by performing a line search over $\zeta \in [0, \|\boldsymbol{h}_{kk}\|_2^2]$ to find the unique value ζ_1, such that the resulting eigenvector $\boldsymbol{w}$ fulfills $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2 = \zeta_1$. The line search can, for example, be performed with bisection: from a lower bound $\check{\zeta}$ and an upper bound $\hat{\zeta}$ on ζ_1 a candidate $\bar{\zeta} = (\check{\zeta} + \hat{\zeta})/2$ is computed and the resulting eigenvector $\boldsymbol{w}$ is determined; if $\bar{\zeta} > |\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2$, the upper bound is updated, i. e., $\hat{\zeta} = \bar{\zeta}$, if $\bar{\zeta} < |\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2$, the lower bound is updated, i. e., $\check{\zeta} = \bar{\zeta}$. Once ζ_1 is found with the desired precision, the updated beamformer $\boldsymbol{v}_k^{(\ell+1)}$ is simply the eigenvector $\boldsymbol{w}$.

If zero-forcing is not possible, the updated beamformer does not necessarily have full power. In order to determine the pricing update, the following steps are taken:

1) Compute $\rho_2 = 1/\|\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{h}_{kk}^*\|_2^2$ and the corresponding value of ζ_2, such that $\rho(\zeta_2) = \rho_2$. If $\zeta_2 \leq 0$, the update is $\boldsymbol{v}_k^{(\ell+1)} = \boldsymbol{0}$.
2) If $0 < \zeta_2 \leq \left(\rho_2^2 \cdot \|\boldsymbol{\Sigma}^{-2}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{h}_{kk}^*\|_2^2\right)^{-1}$, the update is
$$\boldsymbol{v}_k^{(\ell+1)} = \rho_2\sqrt{\zeta_2} \cdot \boldsymbol{U}\boldsymbol{\Sigma}^{-2}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{h}_{kk}^*. \tag{4.96}$$
3) If $\zeta_2 > \left(\rho_2^2 \cdot \|\boldsymbol{\Sigma}^{-2}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{h}_{kk}^*\|_2^2\right)^{-1}$, the value ζ_1 for which $\boldsymbol{w}$ fulfills $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2 = \zeta_1$ is determined via line search over the interval $\left[0, \min\left\{\zeta_2, \|\boldsymbol{h}_{kk}\|_2^2\right\}\right]$. The bisection method from the case where ZF is possible can be used here as well. The updated beamformer $\boldsymbol{v}_k^{(\ell+1)}$ is the resulting eigenvector $\boldsymbol{w}$.

In order to perform the pricing update, user k must know the channel vectors $\boldsymbol{h}_{jk}$ for all $j \in \{1, \ldots, K\}$. In addition, the current SINR $\gamma_k^{(\ell)}$, the current beamformer $\boldsymbol{v}_k^{(\ell)}$, and the interference prices $\pi_j^{(\ell)}$ for all $j \in \{1, \ldots, K\}$ are required. Since $\gamma_k^{(\ell)}$ and $\boldsymbol{v}_k^{(\ell)}$ can be assumed to be known at user k, the only information that must be exchanged between the users in every iteration are the scalar interference prices.

For the MISO interference pricing algorithm, a convergence argument similar to that in Section 3.5.5 can be given, cf. also [55]. In addition to the concavity of the utility functions $u_k(\gamma_k)$ in $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2$ for all $k \in \{1, \ldots, K\}$, it is necessary that $u_k(\gamma_k)$ is convex in $|\boldsymbol{h}_{kj}^{\mathrm{T}}\boldsymbol{v}_j|^2$ for all $j \neq k$, which again results in the condition $c_k(\gamma_k) \in [0, 2]$ for all $k \in \{1, \ldots, K\}$. With sequential updates and up-to-date knowledge of the interference prices it can be argued that in each iteration a lower bound for the sum utility is maximized, which is tight at the beamformer K-tuple $\left(\boldsymbol{v}_1^{(\ell)}, \ldots, \boldsymbol{v}_K^{(\ell)}\right)$. The sequence of sum utilities over the iterations is therefore non-decreasing and thus convergent.

Furthermore, it is straightforward to show that a stationary transmit strategy fulfills the KKT conditions (4.35)–(4.38). Therefore, if the beamformers have converged, the transmit strategy is a candidate for local optimality in the sum utility problem.

For the rate utility
$$\rho(\zeta) = \frac{1}{\zeta + \sum_{j\neq k}\left|\boldsymbol{h}_{kj}^{\mathrm{T}}\boldsymbol{v}_j^{(\ell)}\right|^2 + \sigma^2} = \left(\frac{\left|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k^{(\ell)}\right|^2}{\gamma_k^{(\ell)}} + \zeta\right)^{-1} \tag{4.97}$$
and
$$\zeta_2 = \frac{1}{\rho_2} - \sum_{j\neq k}\left|\boldsymbol{h}_{kj}^{\mathrm{T}}\boldsymbol{v}_j^{(\ell)}\right|^2 - \sigma^2 = \frac{1}{\rho_2} - \frac{\left|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k^{(\ell)}\right|^2}{\gamma_k^{(\ell)}}. \tag{4.98}$$
The interference price of user k is
$$\pi_k^{(\ell)} = \frac{1}{\sum_{j\neq k}\left|\boldsymbol{h}_{kj}^{\mathrm{T}}\boldsymbol{v}_j^{(\ell)}\right|^2 + \sigma^2} - \frac{1}{\sum_{j}\left|\boldsymbol{h}_{kj}^{\mathrm{T}}\boldsymbol{v}_j^{(\ell)}\right|^2 + \sigma^2} = \frac{{\gamma_k^{(\ell)}}^2}{\left(\gamma_k^{(\ell)} + 1\right)\left|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k^{(\ell)}\right|^2}. \tag{4.99}$$

4.4.5.1 Simplified Interference Pricing

In order to avoid the line search involved in the interference pricing updates, we proposed a simplification in [3]: if we assume $\zeta = \left|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k^{(\ell)}\right|^2$ to be constant at the value of the previous iteration, the matrix $\boldsymbol{A}$ does not depend on the beamformer $\boldsymbol{v}_k$. The simplified pricing update can then be performed by determining the highest eigenvalue λ of the matrix

$$\boldsymbol{A} = \rho\Big(\left|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k^{(\ell)}\right|^2\Big)\boldsymbol{h}_{kk}^{*}\boldsymbol{h}_{kk}^{\mathrm{T}} - \sum_{j\neq k}\pi_j^{(\ell)}\boldsymbol{h}_{jk}^{*}\boldsymbol{h}_{jk}^{\mathrm{T}}. \tag{4.100}$$

If $\lambda > 0$, the updated beamformer $\boldsymbol{v}_k^{(\ell+1)}$ is the unit-norm eigenvector corresponding to λ, otherwise it is the zero vector. It is not possible to obtain a beamformer with a power strictly between zero and one.

Since at a stationary beamformer K-tuple it does not make a difference whether ζ is kept constant from the previous iteration or not, stationarity in the simplified pricing algorithm also implies fulfillment of the KKT conditions of the sum utility problem. However, the sum utility is not necessarily increased monotonically over the iterations, and oscillations of the algorithm can indeed be observed in numerical simulations. A numerical evaluation of the convergence behavior and performance of the simplified pricing algorithm for an i. i. d. Gaussian channel model is given in Section 4.5.

4.4.6 Cyclic Coordinate Descent

In [65], a distributed projected gradient algorithm for MISO interference networks was proposed. In contrast to the SISO projected gradient algorithm discussed in Section 3.5.4, the step size is not constant across users and iterations, but is adaptively determined in each iteration by each user separately. For sequential updates, it is proven that the algorithm converges to a transmit strategy fulfilling the KKT conditions without any restrictions on the curvature of the utility functions. As the convergence proof relies on the users updating their transmit strategies one after another and thereby in each step decreasing the total cost function, i. e., the negative sum utility, the algorithm is said to perform a "cyclic coordinate descent". In the following we describe the update procedure of the algorithm and refer to [65] for the derivation and convergence proof.

We define the gradient vector for user k in iteration ℓ as

$$\boldsymbol{g}_k^{(\ell)} = \sum_j \frac{\partial u_j}{\partial \boldsymbol{v}_k^{*}}\bigg|_{\boldsymbol{v}_i=\boldsymbol{v}_i^{(\ell)}\ \forall i.} \tag{4.101}$$

A step size for the gradient projection step is computed as

$$\kappa_k^{(\ell)} = \frac{\left\|\boldsymbol{v}_k^{(\ell)} - \boldsymbol{v}_k^{(\ell-1)}\right\|_2^2}{2\left|\mathrm{Re}\left\{\left(\boldsymbol{v}_k^{(\ell)} - \boldsymbol{v}_k^{(\ell-1)}\right)^{\mathrm{H}}\left(\boldsymbol{g}_k^{(\ell)} - \boldsymbol{g}_k^{(\ell-1)}\right)\right\}\right|}. \tag{4.102}$$

This rule is a heuristic based on the Barzilai-Borwein method [72], which is known to have good numerical properties for quadratic cost functions. As the step size $\kappa_k^{(\ell)}$ depends on the beamformer and gradient of iteration $\ell - 1$, it is necessary to define $\kappa_k^{(0)}$, e. g., as one.

An update direction is determined by taking a step into the direction of the gradient and projecting the result such that the power constraint is fulfilled:

$$\boldsymbol{d}_k^{(\ell)} = \Pi\left(\boldsymbol{v}_k^{(\ell)} + \kappa_k^{(\ell)}\boldsymbol{g}_k^{(\ell)}\right) - \boldsymbol{v}_k^{(\ell)} \tag{4.103}$$

where

$$\Pi(\boldsymbol{v}) = \begin{cases} \boldsymbol{v} & \text{if} \quad \|\boldsymbol{v}\|_2^2 \leq 1 \\ \boldsymbol{v}/\|\boldsymbol{v}\|_2 & \text{otherwise.} \end{cases} \tag{4.104}$$

In order to ensure convergence it is not sufficient to simply use $\boldsymbol{v}_k^{(\ell)} + \boldsymbol{d}_k^{(\ell)}$ as the updated beamformer. Instead, a line search between $\boldsymbol{v}_k^{(\ell)}$ and $\boldsymbol{v}_k^{(\ell)} + \boldsymbol{d}_k^{(\ell)}$ is performed with the goal of guaranteeing that the sum utility function is "sufficiently improved". Specifically, the step size $0 < \delta_k^{(\ell)} \leq 1$ that is applied to find the updated beamformer with

$$\boldsymbol{v}_k^{(\ell+1)} = \boldsymbol{v}_k^{(\ell)} + \delta_k^{(\ell)} \boldsymbol{d}_k^{(\ell)} \tag{4.105}$$

must fulfill the *Armijo condition*

$$u_{\text{sum}}^{(\ell+1)} \geq u_{\text{sum}}^{(\ell)} + 2\eta \,\text{Re}\left\{ \boldsymbol{g}_k^{(\ell)\text{H}} \left(\boldsymbol{v}_k^{(\ell+1)} - \boldsymbol{v}_k^{(\ell)} \right) \right\} = u_{\text{sum}}^{(\ell)} + 2\eta\delta_k^{(\ell)} \,\text{Re}\left\{ \boldsymbol{g}_k^{(\ell)\text{H}} \boldsymbol{d}_k^{(\ell)} \right\} \tag{4.106}$$

for a given parameter $0 < \eta < 1$ and with

$$u_{\text{sum}}^{(\ell)} = \sum_k u_k(\gamma_k) \Big|_{\boldsymbol{v}_j = \boldsymbol{v}_j^{(\ell)} \; \forall j.} \tag{4.107}$$

A suitable step size $\delta_k^{(\ell)}$ can be found by means of a backtracking procedure: starting with $\delta_k^{(\ell)} = 1$, $\delta_k^{(\ell)}$ is decreased by repeated multiplication with a factor $0 < \beta < 1$ until the above condition is fulfilled. The intuition behind the Armijo condition is that it ensures that the sum utility is increased with each update and that a small increase in sum utility is not accompanied by a large change in the beamforming vector, thereby eliminating the possibility of an oscillation. Example parameters used in our numerical simulations are $\beta = 0.5$ and $\eta = 0.1$. For a more detailed discussion of the Armijo condition and backtracking line search, cf. [73, Chapter 3].

It should be noted that repeatedly checking whether the Armijo condition is fulfilled requires user k to be able to compute the sum utility for different values of $\delta_k^{(\ell)}$. Therefore, user k must be able to compute all users' SINRs γ_j depending on $\boldsymbol{v}_k$ and subsequently be able to compute $u_j(\gamma_j)$ for all $j \in \{1, \ldots, K\}$. Assuming that every user has a-priori knowledge of the utility functions of all users, and that the channel vectors have been exchanged so that user k has knowledge of $\boldsymbol{h}_{jk}$ for all $j \in \{1, \ldots, K\}$, it is sufficient for the users to announce the numerator and the denominator of $\gamma_k^{(\ell)}$ after each iteration: with the knowledge of $\left|\boldsymbol{h}_{jj}^{\text{T}} \boldsymbol{v}_j^{(\ell)}\right|^2$ and $\sum_{i \neq j} \left|\boldsymbol{h}_{ji}^{\text{T}} \boldsymbol{v}_i^{(\ell)}\right|^2 + \sigma^2$ user k can evaluate γ_j for any candidate beamformer $\boldsymbol{v}_k$ by subtracting $\left|\boldsymbol{h}_{jk}^{\text{T}} \boldsymbol{v}_k^{(\ell)}\right|^2$ from the denominator and adding $|\boldsymbol{h}_{jk}^{\text{T}} \boldsymbol{v}_k|^2$. Also, as can be seen from (4.40), the numerator and denominator of the current SINR $\gamma_j^{(\ell)}$ enable user k to compute the gradient $\partial u_j / \partial \boldsymbol{v}_k^*$.

While the convergence proof relies on sequential updates, the cyclic coordinate descent algorithm performs similarly with parallel updates. A detailed numerical evaluation of the convergence behavior is given in Section 4.5.

4.4.7 Other Proposed Algorithms

The preceding list of distributed techniques is not exhaustive; the algorithms were selected for the simplicity of their updates and their limited information exchange requirements. Some further algorithms are intentionally not included in our comparison. For example, in [57] a distributed

Algorithm	Iterative	Power control	Sum rate	Convergence
Selfish solution	No	Always full power	No	—
Virtual SINR	No	Always full power	No	—
Global SINR	Yes	Always full power	No	Seq. updates
Unweighted MMSE	Yes	Yes	No	Yes
Adaptively weighted MMSE	Yes	Yes	Yes	Yes
Interference pricing	Yes	Yes	Yes	Seq. updates
Simplified pricing	Yes	Full or zero power	Yes	No
Cyclic coordinate descent	Yes	Yes	Yes	Seq. updates

Table 4.1: Overview of the discussed algorithms for MISO interference networks

Algorithm	Line search	EVD/SVD	Matrix inverse
Selfish solution	No	No	No
Virtual SINR	No	No	Yes
Global SINR	No	Yes	Yes
Unweighted MMSE	Newton	No	Yes
Adaptively weighted MMSE	Newton	No	Yes
Interference pricing	Bisection	Yes	No
Simplified pricing	No	Yes	No
Cyclic coordinate descent	Backtracking	No	No

Table 4.2: Complexity of the updates for the discussed MISO algorithms

algorithm was proposed that is based on the concept of "interference temperatures", i. e., updating beamformers with constraints on the allowed interference power gains. The algorithm updates the beamformers of a pair of users in each iteration with the goal of increasing both SINRs without decreasing the SINRs of all other users, thereby moving in the direction of the Pareto boundary. Each update, however, requires the solution of a convex optimization problem in K real variables and is therefore significantly more complex than the updates of the previously discussed schemes, which require at most a line search. Furthermore, the algorithm cannot be used for the maximization of a sum utility.

We also did not further examine the distributed feasibility-check and balancing techniques proposed in [60], which we briefly mentioned in Section 4.3.4: an iteration of these algorithms requires each user to solve an SOCP in $\sum_k N_k$ variables. Furthermore, it is questionable whether a performance comparison between a balancing algorithm and sum-utility oriented algorithms is meaningful.

4.4.8 Algorithm Comparison and Information Exchange Analysis

Table 4.1 gives an overview over some of the properties of the distributed algorithms discussed in this section. The third column lists whether the norm of the resulting beamformers can be between zero and one; the fourth column states whether the algorithm can be used to search for locally optimal solutions to the sum rate maximization problem; in the fifth column, finally, we indicate whether convergence in some metric, e. g., the sum utility, can be guaranteed.

Algorithm	Rx $k \to$ Tx k once	Tx $k \to$ Tx j once	Rx $k \to$ Tx k per iteration	Tx $k \to$ Tx j per iteration
Selfish solution	$\boldsymbol{h}_{kk}$, γ_k	—	—	—
Virtual SINR	$\boldsymbol{h}_{kj}$ $\forall j$, γ_k	$\boldsymbol{h}_{kj}$	—	—
Global SINR	$\boldsymbol{h}_{kj}$ $\forall j$	$\boldsymbol{h}_{kj}$	$\gamma_k^{(\ell)}$	$n_k^{(\ell)}$, $d_k^{(\ell)}$
Unweighted MMSE	$\boldsymbol{h}_{kj}$ $\forall j$	$\boldsymbol{h}_{kj}$	$\gamma_k^{(\ell)}$	$\left\lvert g_k^{(\ell)}\right\rvert^2$
Adaptively weighted MMSE	$\boldsymbol{h}_{kj}$ $\forall j$	$\boldsymbol{h}_{kj}$	$\gamma_k^{(\ell)}$	$\left\lvert g_k^{(\ell)}\right\rvert^2$, $\alpha_k^{(\ell)}$
Interference pricing	$\boldsymbol{h}_{kj}$ $\forall j$	$\boldsymbol{h}_{kj}$	$\gamma_k^{(\ell)}$	$\pi_k^{(\ell)}$
Simplified pricing	$\boldsymbol{h}_{kj}$ $\forall j$	$\boldsymbol{h}_{kj}$	$\gamma_k^{(\ell)}$	$\pi_k^{(\ell)}$
Cyclic coordinate descent	$\boldsymbol{h}_{kj}$ $\forall j$	$\boldsymbol{h}_{kj}$	$\gamma_k^{(\ell)}$	$n_k^{(\ell)}$, $d_k^{(\ell)}$

Table 4.3: Information exchange requirements of the MISO algorithms assuming that the transmitters compute the updates and are connected via signaling links, cf. Figure 3.6(a).

Even though it is difficult to compare the computational complexity of the updates without explicitly counting the number of arithmetic operations in an efficient implementation, we show some qualitative differences in computational effort between the algorithms by listing in Table 4.2 whether an update requires a line search, the computation of eigenvalues, eigenvectors, or singular values, or the computation of the inverse or pseudo-inverse of an $N_k \times N_k$ matrix. It should be noted, however, that the perhaps most important factor in the overall computational effort required by an algorithm is the number of iterations needed until convergence, which we will experimentally examine in the following section.

For the information exchange analysis in Table 4.3, we assume that all transmitters are connected among each other via signaling links and the receivers are connected to their respective transmitters via a feedback link, cf. Section 3.5; again, the information exchange requirements for the case of connected receivers are very similar. We furthermore assume that σ^2 and the utility functions are known a-priori to each transmitter and that receiver k is able to perfectly estimate the channel vectors $\boldsymbol{h}_{kj}$ for all $j \in \{1, \ldots, K\}$. Also, after each iteration, receiver k is able to measure the current SINR $\gamma_k^{(\ell)}$ and the complex gain of the desired signal $\boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k^{(\ell)}$.

We denote the numerator and denominator of $\gamma_k^{(\ell)}$ by $n_k^{(\ell)}$ and $d_k^{(\ell)}$, respectively, and note that transmitter k can determine $n_k^{(\ell)}$ and $d_k^{(\ell)}$ from $\gamma_k^{(\ell)}$ since it has knowledge of $\boldsymbol{h}_{kk}$ and $\boldsymbol{v}_k^{(\ell)}$. Also, for the MMSE-based algorithms it is of significance that transmitter k can determine the receiver gain $g_k^{(\ell)}$ from the reported SINR $\gamma_k^{(\ell)}$ and knowledge of the direct channel and beamformer, thus eliminating the need for receiver k to feed back both $g_k^{(\ell)}$ and $\gamma_k^{(\ell)}$ after each iteration. The current SINR $\gamma_k^{(\ell)}$ must always be fed back to the transmitter, so that an appropriate channel coding scheme can be chosen for the transmission of data.

For the adaptively weighted MMSE algorithm, we assumed that the weights are updated in each iteration; if the weights are updated less frequently, the $\alpha_k^{(\ell)}$ can accordingly be exchanged less frequently as well.

4.5 Numerical Evaluation of the Distributed Algorithms

In the following we investigate the performance of the previously discussed distributed algorithms in an i. i. d. Gaussian channel model. Our performance metrics are the sum rate utility and the

number of iterations necessary for convergence. We also briefly address the question of how to initialize the algorithms and examine the performance difference between parallel and sequential updates. To begin with, we show some typical SINR and rate regions for the two-user case.

4.5.1 Example SINR and Rate Regions for $K = 2$

The regions depicted in Figures 4.3–4.8 result from a scenario with $N_1 = N_2 = 2$ antennas at each transmitter and the following channel coefficients:

$$\begin{aligned} \boldsymbol{h}_{11}^{\mathrm{T}} &= [-0.75 + 0.25\,\mathrm{j}\ ,\ 0.66 - 0.02\,\mathrm{j}] & \boldsymbol{h}_{12}^{\mathrm{T}} &= [0.16 - 0.26\,\mathrm{j}\ ,\ 0.30 - 0.17\,\mathrm{j}] \\ \boldsymbol{h}_{21}^{\mathrm{T}} &= [0.13 + 0.06\,\mathrm{j}\ ,\ -1.10 + 1.13\,\mathrm{j}] & \boldsymbol{h}_{22}^{\mathrm{T}} &= [0.12 - 0.87\,\mathrm{j}\ ,\ 0.03 - 0.02\,\mathrm{j}]\,. \end{aligned} \tag{4.108}$$

Figures 4.3–4.5 show the SINR region for the noise powers $\sigma^2 = 1$, $\sigma^2 = 0.1$, and $\sigma^2 = 0.01$, respectively. Contrary to the SISO case, the SINR region can be either convex (Figure 4.3) or non-convex (Figures 4.4 and 4.5); the non-convexity generally becomes more pronounced with decreasing noise power σ^2. As was shown in Section 4.4.2, the strategy of maximizing the virtual SINR is always Pareto optimal for $K = 2$. Also, the two successive ZF strategies, where one user employs the selfish beamformer while the other user performs ZF (cf. Section 4.3.3), are always Pareto optimal. From each of the two successive ZF points, the boundary of the SINR region runs orthogonally to the axis in a straight line: these sections of the boundary can be achieved with strategies where the prioritized user achieves its maximal SINR while the other user performs ZF with less than unit power. The selfish and the ZF strategy (which is the unique altruistic solution in this case) are in general not Pareto optimal; however, for higher σ^2 the selfish solution moves closer to the Pareto boundary, whereas for lower σ^2 the ZF solution moves closer to the boundary.

The three corresponding rate regions are shown in Figures 4.6–4.8. A distinct non-convexity can be seen for $\sigma^2 = 0.01$. The trajectory of the interference pricing algorithm (with sequential updates, initialized with the selfish solution) is also included. In all three examples, the pricing algorithm converges to a point close to the virtual SINR maximizer. The strategy of maximizing the virtual SINR in fact seems to be close to sum-rate optimal in these examples. In the following section this will be confirmed to be a general trend for systems in which ZF is possible for all users.

4.5.2 Average Algorithm Performance in a Gaussian Channel Model with $N_k \geq K$

For the numerical performance comparison of the distributed algorithms we again use a simple i. i. d. Gaussian channel model: all entries of all channel vectors are drawn independently from a complex Gaussian distribution. We investigate the case where both the direct channel coefficients and the cross channel coefficients have unit variance, but also present results for a model in which the entries of the cross channel coefficient vectors $\boldsymbol{h}_{kj}$ with $k \neq j$ have variance 0.01, while the direct channel coefficients have variance one.

Previously, e. g., in the discussion of Pareto optimality and high-SNR sum-rate optimality as well as in the derivation of the interference pricing algorithm, it was often necessary to distinguish between the case in which ZF is possible (i. e., where $\boldsymbol{h}_{kk}$ is linearly independent of the space spanned by all $\boldsymbol{h}_{jk}$ with $j \neq k$) and the case in which ZF is not possible. This is an indication that the properties of the system and thus the behavior of the algorithms can be fundamentally different depending on whether ZF is possible for all, for some, or for no users; the numerical results indeed show a qualitative difference, motivating us to investigate two different settings: we first examine a system with four users in which each transmitter has four antennas. Clearly, in our i. i. d. Gaussian

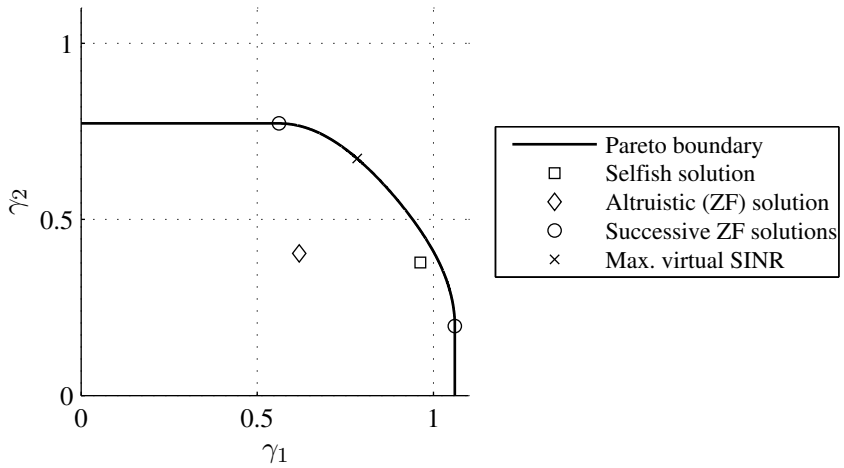


Figure 4.3: SINR region for the example channel (4.108) with $\sigma^2 = 1$.

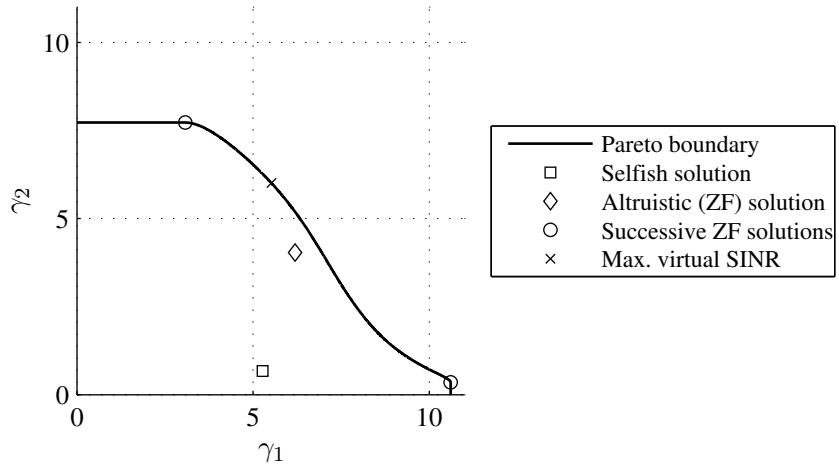


Figure 4.4: Scenario as in Figure 4.3 with $\sigma^2 = 0.1$.

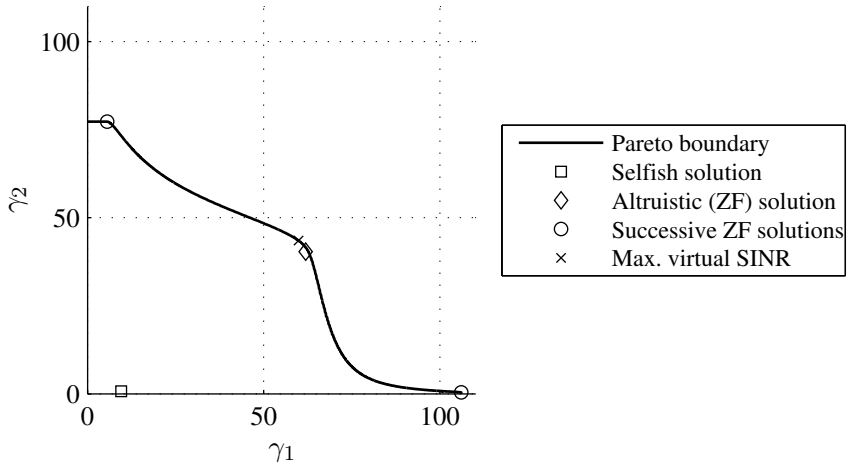


Figure 4.5: Scenario as in Figure 4.3 with $\sigma^2 = 0.01$.

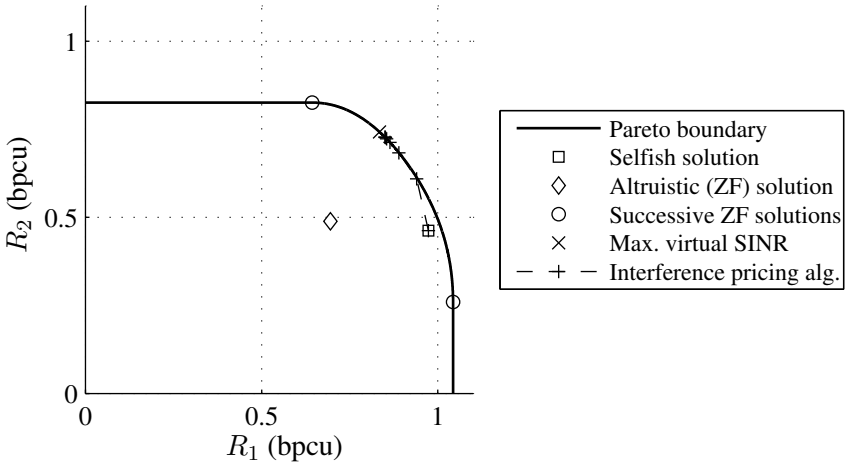


Figure 4.6: Achievable rate region for the example channel (4.108) with $\sigma^2 = 1$. The pricing algorithm is initialized with the selfish strategy.

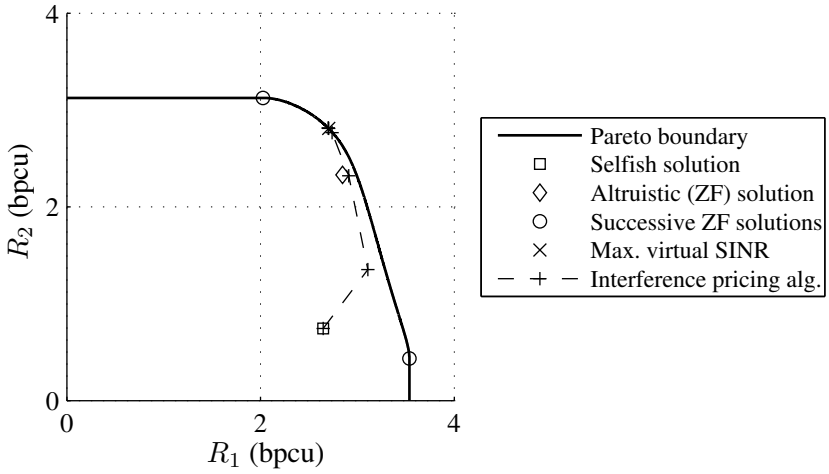


Figure 4.7: Scenario as in Figure 4.6 with $\sigma^2 = 0.1$.

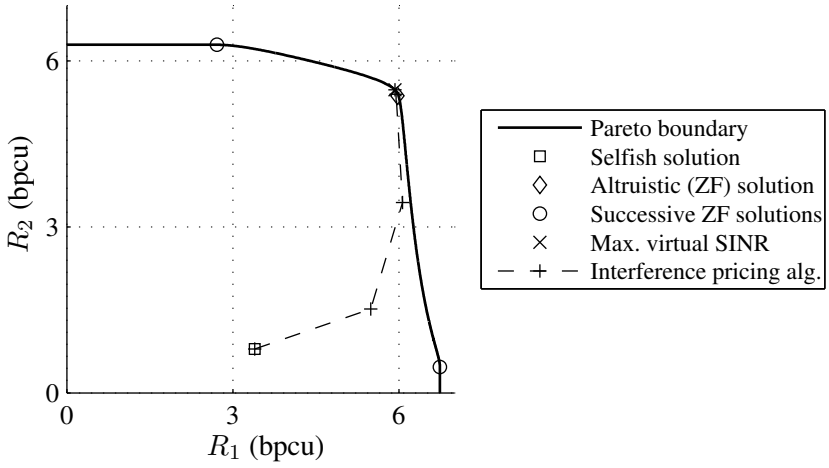


Figure 4.8: Scenario as in Figure 4.6 with $\sigma^2 = 0.01$.

Update mode	Sequential
Max. iteration number	1000 / 10 000
Convergence threshold	10^{-4} / 10^{-6}
Newton steps	10
Bisection steps	20
Backtracking parameters	$\beta = 0.5$, $\eta = 0.1$

Table 4.4: Parameters used for the numerical evaluation of the algorithms: both MMSE-based algorithms and the cyclic coordinate descent algorithm are allowed 10 000 iterations, the other algorithms only 1000 iterations. The stricter convergence threshold 10^{-6} is used only for the adaptively weighted MMSE algorithm and the cyclic coordinate descent algorithm.

channel model, ZF will be possible for every user almost surely. In the next section, we assume eight users with four antennas each, so that ZF will be impossible for every user almost surely.

All algorithms are operated with sequential updates; we note, however, that parallel updates yield very similar sum-rate performance and iteration counts, as will be discussed in Section 4.5.4.5. The weights in the adaptively weighted sum MSE minimization algorithm are updated each time all K beamformers have been updated. The algorithms are run until either a convergence criterion is fulfilled or a maximum number of iterations is reached, where one iteration in this context consists of one beamformer update of every user, i. e., altogether K updates. The convergence criterion is fulfilled when a measure for the combined change of beamformers from one iteration to the next is below a certain threshold; specifically, we use the sum of the Euclidean norms of the difference between the updated and old beamforming vectors, i. e., $\sum_k \left\| \boldsymbol{v}_k^{\mathrm{new}} - \boldsymbol{v}_k^{\mathrm{old}} \right\|_2$. Some further algorithm parameters are listed in Table 4.4; they were chosen by trial and error to yield the best possible performance in terms of sum rate.

We note that it turns out to be necessary to use a very strict convergence criterion and higher maximum iteration number for the cyclic coordinate descent and adaptively weighted MMSE algorithms in order to avoid a significant performance loss at high SNR. Also, the unweighted MMSE algorithm can require very many iterations to converge and is therefore also allowed a higher maximum iteration number. The performance of the remaining iterative algorithms, on the other hand, is relatively robust against the choice of the convergence parameters, so that we choose to use a less strict convergence threshold. We nonetheless believe this to be a fair comparison, as the basic premise is that each algorithm must be operated with a set of parameters that does not incur a performance loss in any of the examined scenarios.

In the case in which ZF is possible, it turns out to be beneficial to initialize the algorithms with the ZF solution, which is also globally optimal at high SNR. If ZF is not possible, we must resort to other initializations, such as selfish or random beamformers. This issue is discussed in Section 4.5.4.6.

Figure 4.9 shows the average performance for unit-variance cross channels, for the results in Figure 4.10 the cross channels are attenuated to have variance 0.01. The median number of iterations required for convergence as well as the percentage of channel realizations for which convergence fails for these two scenarios is given in Table 4.5.

The figures show that the pricing algorithms, the adaptively weighted MMSE algorithm, and the cyclic coordinate descent algorithm achieve virtually identical sum-rate performance. The strategy of maximizing the virtual SINR is mostly very close to optimal, followed by the maximization of the global SINR. The unweighted MMSE algorithm, on the other hand, suffers a small perfor-

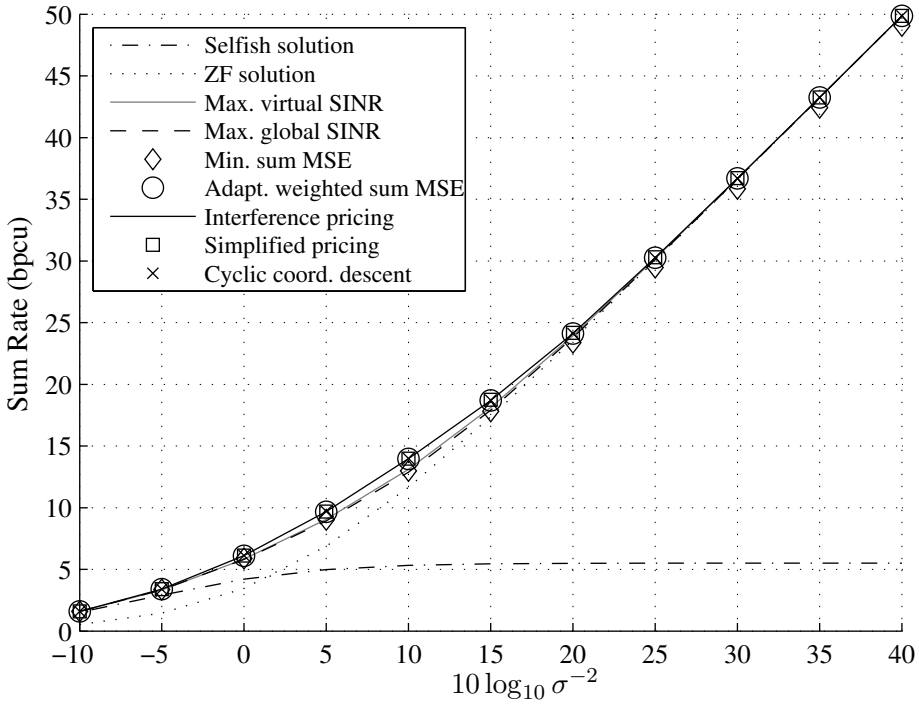


Figure 4.9: $K = 4$ users, $N_k = 4$ antennas for all $k \in \{1, \ldots, 4\}$, sum rate averaged over 1000 channel realizations. The iterative algorithms are initialized with the ZF solution.

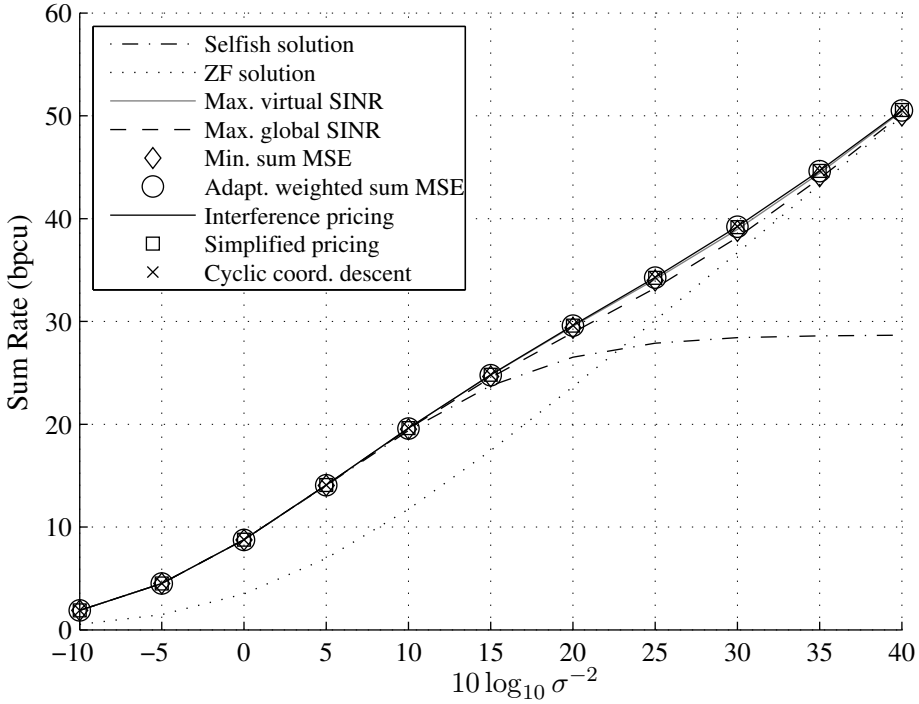


Figure 4.10: $K = 4$ users, $N_k = 4$ antennas for all $k \in \{1, \ldots, 4\}$, sum rate averaged over 1000 channel realizations of a channel model where the variance of the cross channels is 0.01 and the variance of the direct channels is 1. The iterative algorithms are initialized with the ZF solution.

Algorithm	–10 dB	15 dB	40 dB	–10 dB	15 dB	40 dB
Max. global SINR	4 (0)	3 (0)	2 (0)	3 (0)	4 (0)	3 (0)
Min. sum MSE	8 (0)	26 (0)	51 (0)	4 (0)	110 (0)	545.5 (0)
Adapt. weighted sum MSE	13 (0)	55 (0.1)	24 (0)	5 (0)	190 (0)	2106 (6.9)
Interference pricing	6 (0)	7 (0)	2 (0)	3 (0)	7 (0)	4 (0)
Simplified pricing	6 (0)	7 (1.3)	2 (0.1)	3 (0)	7 (0)	5 (1.8)
Cyclic coord. descent	17 (0)	71.5 (0)	25 (0)	10 (0)	21 (0)	65 (0)

Table 4.5: Median number of iterations until convergence for Figure 4.9 (left) and Figure 4.10 (right). In parentheses is the percentage of channel realizations for which the algorithm did not converge before reaching the maximum iteration number.

mance loss at high SNR. As is intuitive, the selfish solution is optimal at low SNR, but saturates when the noise power term is small compared to the interference power terms in the denominator of the SINR; the effect of the saturation is more negative when the cross channels are stronger. The ZF solution is globally optimal at high SNR, but clearly suboptimal when the noise power is large compared to the interference power terms in the denominator of the SINR; the suboptimality is more pronounced when the cross channels are weaker.

The global SINR algorithm and the pricing algorithms generally converge within very few iterations in these scenarios; the simplified pricing algorithm does fail to converge for a few channel realizations, however. The MSE-based algorithms and the cyclic coordinate descent algorithm, on the other hand, converge significantly more slowly when the SNR is high. The adaptively weighted MMSE algorithm, in particular, occasionally even fails to converge within 10 000 iterations in the scenario with attenuated cross channels. The performance and convergence behavior of the algorithms is discussed in more detail in Section 4.5.4.

4.5.3 Algorithm Performance in Systems with $N_k < K$

For the results in Figures 4.11 and 4.12 as well as Table 4.6 we assumed $K = 8$ users with $N_k = 4$ antennas each. The algorithm parameters are the same as in the four-user case. In this setting it is, however, not possible to initialize the algorithms with the ZF solution; we therefore determine the initialization beamformers as follows: first, the elements of the beamformer are drawn independently from a complex Gaussian distribution; next, the beamformer is normalized to have unit norm (yielding an isotropically distributed unit-norm vector); and finally, the beamformer is scaled so that the power (or squared norm) is uniformly distributed between zero and one. As is discussed in Section 4.5.4.6, the performance with this random initialization is marginally superior to the performance with the selfish initialization.

In this scenario it is necessary that some users are deactivated at high SNR for optimal performance. Therefore, the schemes for which the beamformers always have full power, i. e., the selfish solution and the virtual SINR and global SINR maximizers, saturate at high SNR. The unweighted MMSE algorithm does not appear to saturate, but does not reach the full slope either. The pricing algorithms, the adaptively weighted MMSE algorithm, and the cyclic coordinate descent algorithm perform approximately equally well for the scenario with strong cross channels and achieve the full slope. In the scenario with weak cross channels, the simplified pricing algorithm performs poorly at high SNR. As can be seen from Table 4.6, it fails to converge in every single channel realization. Similarly, the adaptively weighted MMSE algorithm in most instances fails to converge within 10 000 iterations.

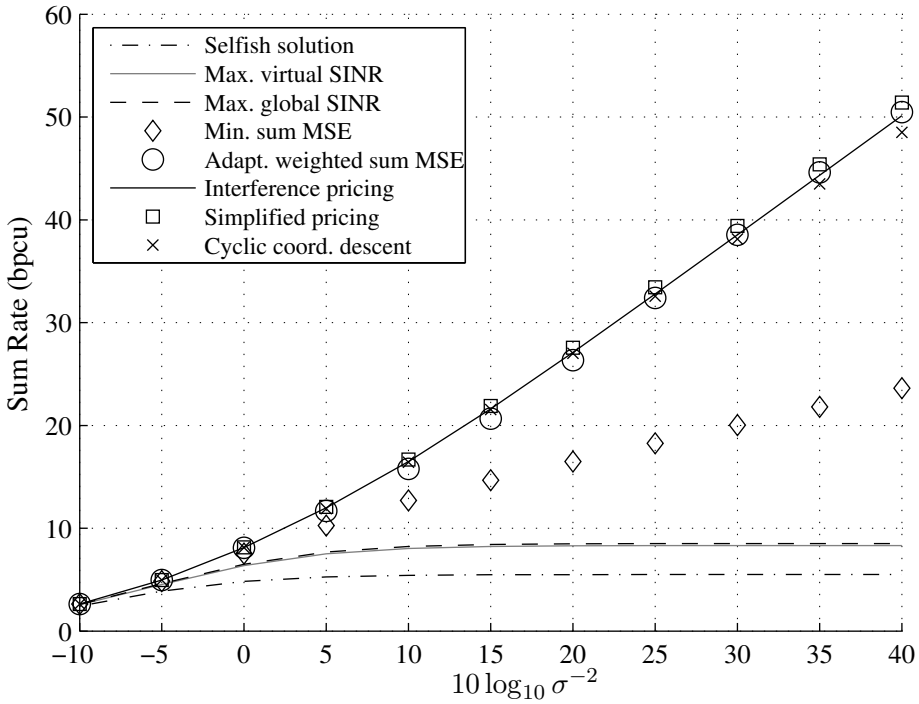


Figure 4.11: $K = 8$ users, $N_k = 4$ antennas for all $k \in \{1, \ldots, 8\}$, sum rate averaged over 1000 channel realizations. The iterative algorithms are initialized with random beamformers.

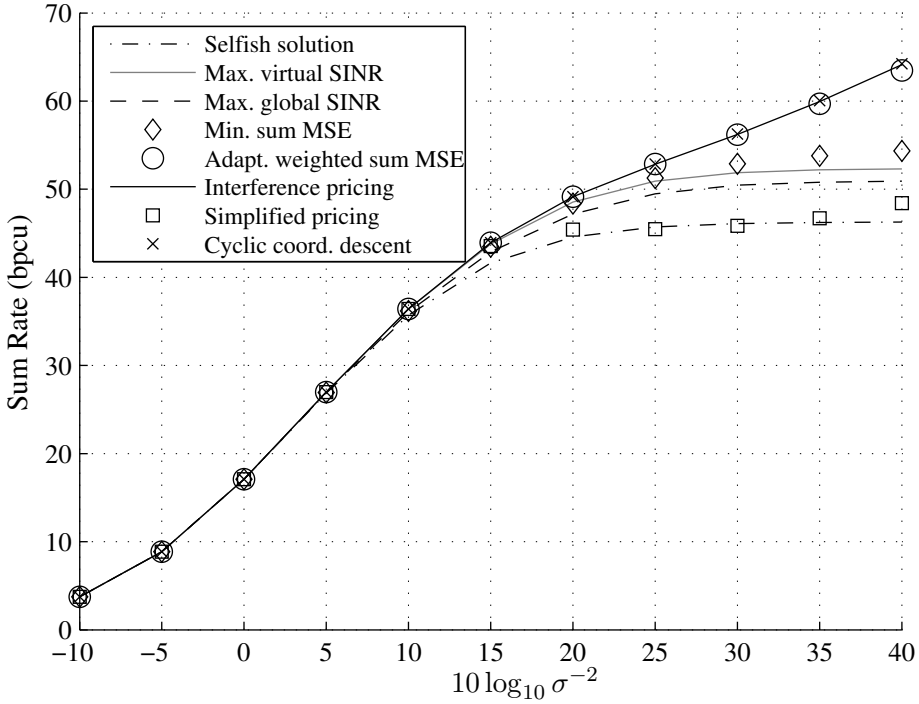


Figure 4.12: $K = 8$ users, $N_k = 4$ antennas for all $k \in \{1, \ldots, 8\}$, sum rate averaged over 1000 channel realizations of a channel model where the variance of the cross channels is 0.01 and the variance of the direct channels is 1. The iterative algorithms are initialized with random beamformers.

Algorithm	–10 dB	15 dB	40 dB	–10 dB	15 dB	40 dB
Max. global SINR	4 (0)	4 (0)	4 (0)	3 (0)	4 (0)	4 (0)
Min. sum MSE	12 (0)	71 (0)	299 (0)	4 (0)	118 (0)	2267 (0)
Adapt. weighted sum MSE	21 (0)	73 (0)	5673 (29.9)	5 (0)	202.5 (0)	10 000 (97.4)
Interference pricing	9 (0)	13 (0)	11 (0)	3 (0)	10 (0)	22 (0)
Simplified pricing	9 (0)	11 (2.4)	9 (0)	3 (0)	12 (12.4)	1000 (99.7)
Cyclic coord. descent	30 (0)	557.5 (0)	4614.5 (2.4)	12 (0)	28 (0)	751 (0)

Table 4.6: Median iteration number for Figure 4.11 (left) and Figure 4.12 (right). In parentheses is the percentage of channel realizations for which the algorithm did not converge before reaching the maximum iteration number.

4.5.4 Discussion of the Numerical Properties of the Algorithms

4.5.4.1 Non-Iterative Methods and Maximization of the Global SINR

The selfish solution in general is clearly not a good strategy in terms of sum rate, unless it is known that the noise power σ^2 is significantly higher than the interference terms in the denominators of the SINR expressions. The strategy of maximizing the virtual SINR, which also does not require iterative updates, achieves close to optimal performance, on the other hand, as long as the scenario allows for ZF, and in those cases provides an excellent compromise between simplicity and performance. The iterative technique of maximizing the global SINR does not yield any substantial improvement over maximizing the virtual SINR.

If, however, ZF is not possible, i. e., the number of users exceeds the number of antennas per user in our channel model, the full potential of the system cannot be achieved without iterative techniques.[5]

4.5.4.2 MMSE-Based Methods

While the adaptively weighted MMSE algorithm shows excellent performance in terms of sum rate, the necessary number of iterations can be very high, especially if σ^2 is very low. This problem is also present, but not as severe, for the unweighted MMSE algorithm, which, however, does not reach a satisfactory sum rate in scenarios that require reducing the number of active users. Altogether, we conclude that the convergence properties of the MMSE-based approaches are not ideal for general MISO interference networks.

4.5.4.3 Pricing Methods

The interference pricing algorithm offers very reliable and rapid convergence as well as excellent sum-rate performance in all examined scenarios. The simulation results indicate clearly that interference pricing is the strategy of choice for general MISO interference networks.

The simplified version of the pricing algorithm often performs very well, but has convergence issues in certain scenarios. Specifically, it seems that if power control is necessary, i. e., when ZF is not possible and the SNR is neither asymptotically high nor asymptotically low, the probability

[5] Alternatively, a preliminary user selection step could reduce the number of users in such a system, so that ZF is possible again. We note, however, that performing such a user selection without a centralized computer is not trivial; the iterative algorithms, on the other hand, perform this user selection "automatically" when it is needed.

Algorithm	–10 dB	15 dB	40 dB
Max. global SINR	5 (0)	6 (0)	6 (0)
Min. sum MSE	11 (0)	69 (0)	310 (0)
Adapt. weighted sum MSE	19 (0)	83 (0)	4659 (29.4)
Interference pricing	11 (0)	16 (0)	14 (0)
Simplified pricing	11 (0)	15 (3.7)	12 (0.3)
Cyclic coord. descent	32 (0)	533 (0)	4325 (2.3)

Table 4.7: Median iteration number and failed convergence percentage for the scenario in Figure 4.11 (cf. Table 4.6, left side) but with parallel instead of sequential updates.

of running into an oscillation can be high. Simplified interference pricing might be a good low-complexity alternative if measures can be taken to prevent or detect oscillations, which we do not further pursue in this work.

4.5.4.4 Cyclic Coordinate Descent

The performance of the cyclic coordinate descent algorithm is very good as well. The number of iterations required for convergence is very high at high SNR, however, and a low convergence threshold is required, as otherwise small changes is $\boldsymbol{v}_k$ from one iteration to the next are falsely interpreted as convergence of the algorithm, when in fact they are a consequence of a small step size $\kappa_k^{(\ell)}$, cf. (4.102). This raises the question of whether the heuristic for the step size $\kappa_k^{(\ell)}$ can be improved; in any case, finding a step size heuristic and convergence parameters that are suitable for different scenarios, channel models, and a wide range of noise powers certainly presents a challenge.

4.5.4.5 Sequential vs. Parallel Updates

The convergence proofs of some of the discussed algorithms rely on sequential updates. Also, it is intuitive that sequential updates are beneficial for rapid convergence as we can think of parallel updates as sequential updates with outdated information on the other users' conditions. Parallel updates, on the other hand, have the advantage of requiring less information exchange: the users must communicate over their signaling links only once per K beamformer updates.

In order to quantitatively investigate the difference in performance, we simulated the scenario with $K = 8$ users, $N_k = 4$ antennas for all users $k \in \{1, \ldots, 4\}$, and equal direct and cross channel gains with parallel updates. We omit the sum-rate performance plot, as it is virtually identical to Figure 4.11; a slight difference can be, however, observed in the number of iterations necessary for convergence, as is shown in Table 4.7: the global SINR and pricing algorithms require several additional iterations for convergence, for the remaining algorithms such an effect is not evident. Altogether the difference between sequential and parallel updates does not appear to be very large.

4.5.4.6 Choice of the Initialization

In Figure 4.13 and Table 4.8, the sum rate and convergence performance is shown for the interference pricing algorithm with sequential updates and different initialization strategies in a scenario with $K = 4$ users, $N_k = 4$ antennas for all users $k \in \{1, \ldots, 4\}$, and equal direct and cross channel gains. The algorithm performs best with zero-forcing initializations, while the other three

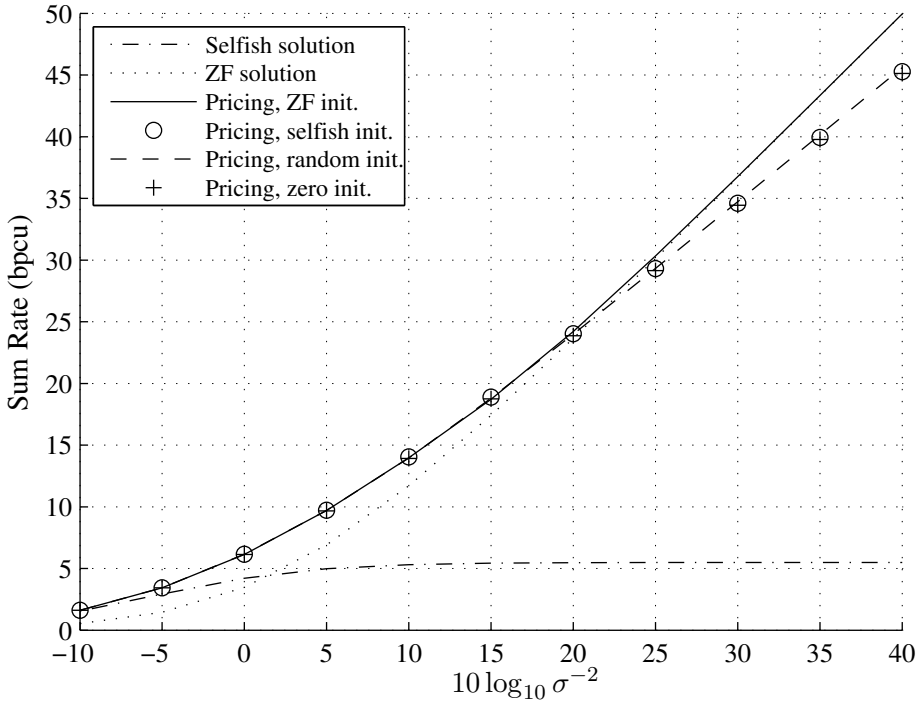


Figure 4.13: $K = 4$ users, $N_k = 4$ antennas for all $k \in \{1, \ldots, 4\}$, sum rate of the interference pricing algorithm starting from different initializations averaged over 1000 channel realizations.

Initialization	–10 dB	15 dB	40 dB
Zero-forcing solution	6 (0)	7 (0)	2 (0.1)
Selfish solution	6 (0)	9 (0)	8 (0)
Random beamformers	6 (0)	9 (0)	8 (0)
Zero beamformers	7 (0)	10 (0)	9 (0)

Table 4.8: Median iteration number and failed convergence percentage for Figure 4.13.

strategies lead to similar results on average, with a marginal advantage for the random initialization. The fact that the slope of the graphs for non-ZF initializations is slightly lower than for the ZF initialization indicates that the interference pricing algorithm in some cases converges towards a suboptimal strategy for which not all users have close to zero interference.

We note that further numerical experiments show that the behavior is similar for other scenarios and algorithms, with the limitation that the zero initialization is not suitable for the MMSE-based and cyclic coordinate descent methods, as for those algorithms the zero beamformers are a stationary transmit strategy. We therefore conclude that a sensible strategy is to initialize the beamformers with the zero-forcing solution whenever it exists and with a random beamformer otherwise.

5. Single-Stream MIMO Interference Networks

In a general multi-antenna network, both the transmitters and the receivers may have more than one antenna. We refer to such scenarios as multiple-input multiple-output (MIMO) systems. In contrast to the MISO case, single-stream beamforming is not necessarily optimal in MIMO systems; in fact, the potential of MIMO systems lies in the capability to transmit several spatially multiplexed data streams, which can greatly increase the capacity [74]. In this work, however, we investigate only cases where at most one stream is transmitted per user. These special cases already provide insights into some of the fundamental properties of MIMO interference networks without being burdened by issues such as intra-user interference, choosing the optimal stream configuration, and allocating the power among the different streams of one user. Also, since the single-stream MIMO case is closely related to the MISO case, we can generalize many of the previously discussed algorithms in a straightforward way.

By combining the signals from their different antennas, multi-antenna receivers are—to a certain extent—able to distinguish the desired signal from the interference. The goal of the transmitters is therefore no longer to form the beam such that the interference is canceled at the receivers, but such that the interference is distinguishable from the desired signal. As will be discussed in the following, the key idea is that the interference from the different transmitters is *aligned* in a lower-dimensional subspace at each receiver.

Even for the single-stream case, the problem of finding optimal strategies in MIMO interference networks is highly complex. Analytical results are only available for the rate utility in the asymptotic regimes of low and high SNR. In this chapter we discuss these asymptotic results and proceed to compare the known distributed algorithms, which can for the most part be derived from the MISO algorithms presented in the previous chapter. We numerically evaluate the performance of the algorithms in an i. i. d. Gaussian channel model.

5.1 System Model

As in the previous chapter, transmitter k has N_k antennas; we now additionally allow multiple antennas at the receivers and denote the number of antennas at receiver k as M_k. The complex channel gains between the N_j antennas of transmitter j and the M_k antennas of receiver k form the matrix $\boldsymbol{H}_{kj} \in \mathbb{C}^{M_k \times N_j}$. We assume the direct channels to be non-zero, i. e., $\boldsymbol{H}_{kk} \neq \mathbf{0}$ for all $k \in \{1, \ldots, K\}$.

The vector $\boldsymbol{y}_k \in \mathbb{C}^{M_k}$ containing the received symbols at the M_k antennas of receiver k is

$$\boldsymbol{y}_k = \boldsymbol{H}_{kk}\boldsymbol{x}_k + \sum_{j \neq k} \boldsymbol{H}_{kj}\boldsymbol{x}_j + \boldsymbol{n}_k \qquad \forall k \in \{1, \ldots, K\} \tag{5.1}$$

where $\boldsymbol{x}_j \in \mathbb{C}^{N_j}$ is the vector of symbols transmitted from the N_j antennas of transmitter j and $\boldsymbol{n}_k \in \mathbb{C}^{M_k}$ is the vector of additive noise experienced at the M_k antennas of receiver k. The mean of the noise is zero and the noise covariance matrix is

$$\mathrm{E}\left[\boldsymbol{n}_k \boldsymbol{n}_k^{\mathrm{H}}\right] = \sigma^2 \mathbf{I} \qquad \forall k \in \{1, \ldots, K\} \tag{5.2}$$

i. e., the noise power is $\sigma^2 > 0$ at every antenna of every receiver and is not correlated between the antennas of one receiver. We also assume the transmit symbol vectors $\boldsymbol{x}_k$ to have mean zero and, since the transmitters cannot cooperatively encode their signals, to be uncorrelated between users, i. e., $\mathrm{E}\left[\boldsymbol{x}_k \boldsymbol{x}_j^{\mathrm{H}}\right] = \mathbf{0}$ for $k \neq j$. The transmit covariance matrix of user k

$$\boldsymbol{Q}_k = \mathrm{E}\left[\boldsymbol{x}_k \boldsymbol{x}_k^{\mathrm{H}}\right] \qquad \forall k \in \{1, \ldots, K\} \tag{5.3}$$

is by definition positive semi-definite and is, as in the MISO case, subject to the power constraint

$$\mathrm{tr}\left(\boldsymbol{Q}_k\right) \leq 1 \qquad \forall k \in \{1, \ldots, K\}. \tag{5.4}$$

While it is possible to define the SINR of user k using the power of the desired signal, interference, and noise terms summed up across all antennas of receiver k, it is not clear whether this is a relevant metric. In particular, the achievable rate (when the transmit symbols and noise vectors are Gaussian and interference is treated as noise) is in general not a function of a so-defined SINR. Instead, we avoid the discussion of a generalization of the SINR to multi-antenna receivers and from this point on restrict our attention to the case of single-stream beamforming, where defining the SINR is straightforward.

For single-stream transmission, user k multiplies the unit-variance data symbol $s_k \in \mathbb{C}$ with the beamformer $\boldsymbol{v}_k \in \mathbb{C}^{N_k}$ to form the transmit vector

$$\boldsymbol{x}_k = \boldsymbol{v}_k \cdot s_k \qquad \forall k \in \{1, \ldots, K\} \tag{5.5}$$

where, to fulfill the unit power constraint, the beamformers must satisfy

$$\|\boldsymbol{v}_k\|_2^2 \leq 1 \qquad \forall k \in \{1, \ldots, K\}. \tag{5.6}$$

Receiver k combines the received symbols from the M_k antennas using the receive filter vector $\boldsymbol{g}_k \in \mathbb{C}^{M_k}$ to form the estimated symbol

$$\hat{s}_k = \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{y}_k \qquad \forall k \in \{1, \ldots, K\} \tag{5.7}$$

where we require that $\boldsymbol{g}_k \neq \mathbf{0}$. The SINR of user k after the receive filter can be stated as

$$\gamma_k = \frac{|\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k|^2}{\sum\limits_{j \neq k} |\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kj}^{\mathrm{T}} \boldsymbol{v}_j|^2 + \|\boldsymbol{g}_k\|_2^2 \sigma^2} \qquad \forall k \in \{1, \ldots, K\}. \tag{5.8}$$

We note that γ_k does not change when $\boldsymbol{g}_k$ is multiplied with a non-zero scalar.

As noted in the introduction to this chapter, single-stream transmission is in general not optimal. As a simple counter-example, we can consider a point-to-point MIMO link, i. e., a system with $K = 1$ user, and the rate utility; it is well known that it is optimal to diagonalize the channel matrix and distribute the power over several orthogonal modes, thus transmitting up to $\min\{N_1, M_1\}$ streams [74]. We note, however, that there are many MIMO scenarios, such as the *fully loaded* systems discussed in Section 5.3.4, where single-stream beamforming is known to be sum-rate optimal in the high-SNR regime.

5.1.1 General Power Constraints and Noise Covariances

For notational convenience, we assume in our system model that all users have a unit power constraint and that the noise is uncorrelated and has the same variance at every receive antenna. This does not, however, present a loss of generality; systems with different power constraints and colored noise can be transformed into an equivalent system fulfilling our assumptions, as we will show in the following.

Let us assume a more general system with received and transmitted symbol vectors $\boldsymbol{y}'_k$ and $\boldsymbol{x}'_k$, respectively, as well as noise vectors $\boldsymbol{n}'_k$ and channel matrices $\boldsymbol{H}'_{kj}$, which is governed by the equations

$$\boldsymbol{y}'_k = \sum_j \boldsymbol{H}'_{kj}\boldsymbol{x}'_j + \boldsymbol{n}'_k \qquad \forall k \in \{1, \ldots, K\}. \tag{5.9}$$

The power constraint for user k is $\mathrm{E}[\|\boldsymbol{x}'_k\|_2^2] = \mathrm{tr}(\boldsymbol{Q}'_k) \leq P_k$ and the noise covariance matrix at receiver k is $\mathrm{E}\left[\boldsymbol{n}'_k \boldsymbol{n}'^{\mathrm{H}}_k\right] = \boldsymbol{R}_k$, where $\boldsymbol{R}_k$ has full rank.[1]

For our equivalent system, we define the channel matrices $\boldsymbol{H}_{kj} = \sigma\sqrt{P_j}\boldsymbol{R}_k^{-\frac{1}{2}}\boldsymbol{H}'_{kj}$, the transmit symbol vectors $\boldsymbol{x}_k = \boldsymbol{x}'_k/\sqrt{P_k}$, and the noise vectors $\boldsymbol{n}_k = \sigma\boldsymbol{R}_k^{-\frac{1}{2}}\boldsymbol{n}'_k$. Furthermore, we let each receiver k perform a linear operation on its received symbol vector yielding the "whitened" receive symbol vector $\boldsymbol{y}_k = \sigma\boldsymbol{R}_k^{-\frac{1}{2}}\boldsymbol{y}'_k$, so that we obtain

$$\boldsymbol{y}_k = \sigma\boldsymbol{R}_k^{-\frac{1}{2}}\boldsymbol{y}'_k = \sum_j \sigma\boldsymbol{R}_k^{-\frac{1}{2}}\boldsymbol{H}'_{kj}\boldsymbol{x}'_j + \sigma\boldsymbol{R}_k^{-\frac{1}{2}}\boldsymbol{n}'_k = \sum_j \boldsymbol{H}_{kj}\boldsymbol{x}_j + \boldsymbol{n}_k. \tag{5.10}$$

Due to the definitions of $\boldsymbol{n}_k$ and $\boldsymbol{x}_k$, $\mathrm{E}\left[\boldsymbol{n}_k\boldsymbol{n}_k^{\mathrm{H}}\right] = \sigma^2\mathbf{I}$ and the power constraint translates to $\mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{x}_k\boldsymbol{x}_k^{\mathrm{H}}\right]\right) \leq 1$. Thus, by adding a "noise whitening" filter $\sigma\boldsymbol{R}_k^{-\frac{1}{2}}$ as a preliminary receive processing step and renormalizing the channel matrices, we are able transform any system with general power constraints and noise covariances into an equivalent system that fits our framework.

It remains to be shown that the two systems have identical SINR regions for single-stream transmission. To this end, let us assume that the SINR K-tuple $(\gamma_1, \ldots, \gamma_K)$ is achievable in the original system by the receive filter K-tuple $(\boldsymbol{g}'_1, \ldots, \boldsymbol{g}'_K)$ and the beamformer K-tuple $(\boldsymbol{v}'_1, \ldots, \boldsymbol{v}'_K)$. It is straightforward that in the equivalent system the receive filters $\boldsymbol{g}_k^{\mathrm{T}} = \boldsymbol{g}'^{\mathrm{T}}_k\boldsymbol{R}_k^{\frac{1}{2}}/\sigma$ and beamformers $\boldsymbol{v}_k = \boldsymbol{v}'_k/\sqrt{P_k}$ for all $k \in \{1, \ldots, K\}$ achieve the same SINR K-tuple. Conversely, for an SINR K-tuple that is achievable in the equivalent system, $\boldsymbol{g}'_k$ and $\boldsymbol{v}'_k$ can in the same way be obtained from $\boldsymbol{g}_k$ and $\boldsymbol{v}_k$ such that the SINRs in the original system are identical. Therefore, any SINR K-tuple that is achievable in the original system is also achievable in the equivalent system and vice versa, and the SINR regions are identical.

5.2 Optimal Receive Filters

The SINR γ_k depends only on $\boldsymbol{g}_k$ and not on any $\boldsymbol{g}_j$ with $j \neq k$. Therefore, we can choose the receive filter $\boldsymbol{g}_k$ to maximize the SINR γ_k without affecting the other users' performance by solving the optimization problem

$$\boldsymbol{g}_k^{\mathrm{opt}} = \arg\max_{\boldsymbol{g}_k} \ \gamma_k \qquad \text{s. t.:} \quad \|\boldsymbol{g}_k\|_2^2 = 1. \tag{5.11}$$

[1] If the noise covariance matrix $\boldsymbol{R}_k$ does not have full rank, there exists a receive filter vector $\boldsymbol{g}_k$ such that $\mathrm{E}[|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{n}'_k|^2] = 0$, i. e., user k can transmit a data stream without noise and the achievable rate is infinite.

Note that we impose a unit-norm equality constraint on the vector $\boldsymbol{g}_k$: since the SINR γ_k is invariant to a scaling of $\boldsymbol{g}_k$, this restriction does not cause a loss of optimality; assuming unit-norm receive filters, however, provides some notational advantages later on.

We observe that the SINR can be stated as a generalized Rayleigh quotient

$$\gamma_k = \frac{\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k \boldsymbol{v}_k^{\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^*}{\boldsymbol{g}_k^{\mathrm{T}} \left(\sum_{j \neq k} \boldsymbol{H}_{kj} \boldsymbol{v}_j \boldsymbol{v}_j^{\mathrm{H}} \boldsymbol{H}_{kj}^{\mathrm{H}} + \sigma^2 \mathbf{I} \right) \boldsymbol{g}_k^*} \tag{5.12}$$

and therefore (5.11) can be solved in the same way as the similar optimization problems in Sections 4.4.2 and 4.4.3 (also cf. [37, Section 4.2]): assuming that $\boldsymbol{v}_k \neq \mathbf{0}$, the complex conjugate optimal receive filter is

$$\boldsymbol{g}_k^{\mathrm{opt}*} = \frac{1}{\left\| \left(\sum_{j \neq k} \boldsymbol{H}_{kj} \boldsymbol{v}_j \boldsymbol{v}_j^{\mathrm{H}} \boldsymbol{H}_{kj}^{\mathrm{H}} + \sigma^2 \mathbf{I} \right)^{-1} \boldsymbol{H}_{kk} \boldsymbol{v}_k \right\|_2} \left(\sum_{j \neq k} \boldsymbol{H}_{kj} \boldsymbol{v}_j \boldsymbol{v}_j^{\mathrm{H}} \boldsymbol{H}_{kj}^{\mathrm{H}} + \sigma^2 \mathbf{I} \right)^{-1} \boldsymbol{H}_{kk} \boldsymbol{v}_k \tag{5.13}$$

and, if $\boldsymbol{v}_k = \mathbf{0}$, any unit-norm receive filter is optimal, as the SINR is zero regardless of the receive filter. Inserting the optimal receive filter into the expression for the SINR yields

$$\gamma_k^{\mathrm{opt}} = \gamma_k \Big|_{\boldsymbol{g}_k = \boldsymbol{g}_k^{\mathrm{opt}}} = \boldsymbol{v}_k^{\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} \left(\sum_{j \neq k} \boldsymbol{H}_{kj} \boldsymbol{v}_j \boldsymbol{v}_j^{\mathrm{H}} \boldsymbol{H}_{kj}^{\mathrm{H}} + \sigma^2 \mathbf{I} \right)^{-1} \boldsymbol{H}_{kk} \boldsymbol{v}_k. \tag{5.14}$$

Instead of maximizing the SINR, we can also choose the receive filter that minimizes the MSE:

$$\boldsymbol{g}_k^{\mathrm{MMSE}} = \arg\min_{\boldsymbol{g}_k} \ \varepsilon_k \tag{5.15}$$

where

$$\begin{aligned} \varepsilon_k &= \mathrm{E}\left[|\hat{s}_k - s_k|^2 \right] = \mathrm{E}\left[\left| (\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k - 1) s_k + \sum_{j \neq k} \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kj} \boldsymbol{v}_j s_j + \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{n}_k \right|^2 \right] \\ &= \sum_j |\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kj} \boldsymbol{v}_j|^2 - 2 \,\mathrm{Re}\{ \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k \} + 1 + \|\boldsymbol{g}_k\|_2^2 \sigma^2. \end{aligned} \tag{5.16}$$

Since ε_k is not invariant to a scaling of $\boldsymbol{g}_k$, we do not impose a constraint on the norm of $\boldsymbol{g}_k$. Setting the derivative of the MSE to zero results in the necessary condition for optimality

$$\frac{\partial \varepsilon_k}{\partial \boldsymbol{g}_k} = \sum_j \boldsymbol{H}_{kj} \boldsymbol{v}_j \boldsymbol{v}_j^{\mathrm{H}} \boldsymbol{H}_{kj}^{\mathrm{H}} \boldsymbol{g}_k^* - \boldsymbol{H}_{kk} \boldsymbol{v}_k + \boldsymbol{g}_k^* \sigma^2 = \mathbf{0}. \tag{5.17}$$

Consequently,

$$\boldsymbol{g}_k^{\mathrm{MMSE}*} = \left(\sum_j \boldsymbol{H}_{kj} \boldsymbol{v}_j \boldsymbol{v}_j^{\mathrm{H}} \boldsymbol{H}_{kj}^{\mathrm{H}} + \sigma^2 \mathbf{I} \right)^{-1} \boldsymbol{H}_{kk} \boldsymbol{v}_k. \tag{5.18}$$

Note that for $\boldsymbol{v}_k = \mathbf{0}$ the MSE-optimal receive filter is also the null vector, which is problematic since the SINR is not defined for this case. It is intuitive, though, that this is a technicality that can be resolved by defining the SINR to be zero if both $\boldsymbol{v}_k = \mathbf{0}$ and $\boldsymbol{g}_k = \mathbf{0}$.

While (5.13) and (5.18) appear to be similar, the inverse in (5.18) contains the sum over all indices j including k. Nonetheless, $\boldsymbol{g}_k^{\mathrm{MMSE}}$ and $\boldsymbol{g}_k^{\mathrm{opt}}$ are collinear, as we will show in the following. With the abbreviation

$$\boldsymbol{X}_k = \sum_{j \neq k} \boldsymbol{H}_{kj} \boldsymbol{v}_j \boldsymbol{v}_j^{\mathrm{H}} \boldsymbol{H}_{kj}^{\mathrm{H}} + \sigma^2 \mathbf{I} \tag{5.19}$$

we can express the MSE-optimal receive filter as

$$\boldsymbol{g}_k^{\mathrm{MMSE}*} = \left(\boldsymbol{H}_{kk} \boldsymbol{v}_k \boldsymbol{v}_k^{\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} + \boldsymbol{X}_k \right)^{-1} \boldsymbol{H}_{kk} \boldsymbol{v}_k. \tag{5.20}$$

By applying the matrix-inversion lemma we obtain

$$\begin{aligned} \boldsymbol{g}_k^{\mathrm{MMSE}*} &= \left(\boldsymbol{X}_k^{-1} - \frac{1}{1 + \boldsymbol{v}_k^{\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{X}_k^{-1} \boldsymbol{H}_{kk} \boldsymbol{v}_k} \boldsymbol{X}_k^{-1} \boldsymbol{H}_{kk} \boldsymbol{v}_k \boldsymbol{v}_k^{\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{X}_k^{-1} \right) \boldsymbol{H}_{kk} \boldsymbol{v}_k \\ &= \frac{1}{1 + \boldsymbol{v}_k^{\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{X}_k^{-1} \boldsymbol{H}_{kk} \boldsymbol{v}_k} \boldsymbol{X}_k^{-1} \boldsymbol{H}_{kk} \boldsymbol{v}_k. \end{aligned} \tag{5.21}$$

We recall that

$$\boldsymbol{g}_k^{\mathrm{opt}*} = \frac{1}{\left\| \boldsymbol{X}_k^{-1} \boldsymbol{H}_{kk} \boldsymbol{v}_k \right\|_2} \boldsymbol{X}_k^{-1} \boldsymbol{H}_{kk} \boldsymbol{v}_k \tag{5.22}$$

and it is evident that the two optimal receive filters are identical up to a scalar factor. Since γ_k is invariant to the scaling, we can also state that

$$\gamma_k \Big|_{\boldsymbol{g}_k = \boldsymbol{g}_k^{\mathrm{MMSE}}} = \gamma_k \Big|_{\boldsymbol{g}_k = \boldsymbol{g}_k^{\mathrm{opt}}} = \gamma_k^{\mathrm{opt}}. \tag{5.23}$$

5.2.1 Altruism in MIMO Interference Networks

We recall from Section 4.3.2 that the strategy of user k is altruistic if
1) user k cannot change its strategy to increase any γ_j with $j \neq k$ and
2) user k cannot change its strategy to increase γ_k without decreasing at least one γ_j with $j \neq k$.
In the MIMO case, the strategy of user k consists of the beamformer $\boldsymbol{v}_k$ and the receive filter $\boldsymbol{g}_k$. The second of the two altruism conditions implies that user k must use the optimal receive filter $\boldsymbol{g}_k^{\mathrm{opt}}$; otherwise γ_k can be increased with no effect on the other users' SINRs by changing $\boldsymbol{g}_k$. The first condition implies that if all users behave altruistically, it cannot be possible to increase γ_k^{opt} by changing any $\boldsymbol{v}_j$ with $j \neq k$. This is clearly the case when $\boldsymbol{H}_{kj} \boldsymbol{v}_j = \mathbf{0}$, but can also be achieved in a different way, as can be seen by applying the matrix-inversion lemma to the optimal SINR:

$$\begin{aligned} \gamma_k^{\mathrm{opt}} &= \boldsymbol{v}_k^{\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{X}_k^{-1} \boldsymbol{H}_{kk} \boldsymbol{v}_k = \boldsymbol{v}_k^{\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} \left(\boldsymbol{Y}_k \boldsymbol{Y}_k^{\mathrm{H}} + \sigma^2 \mathbf{I} \right)^{-1} \boldsymbol{H}_{kk} \boldsymbol{v}_k \\ &= \sigma^{-2} \boldsymbol{v}_k^{\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{H}_{kk} \boldsymbol{v}_k - \sigma^{-2} \boldsymbol{v}_k^{\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{Y}_k \left(\sigma^2 \mathbf{I} + \boldsymbol{Y}_k^{\mathrm{H}} \boldsymbol{Y}_k \right)^{-1} \boldsymbol{Y}_k^{\mathrm{H}} \boldsymbol{H}_{kk} \boldsymbol{v}_k \end{aligned} \tag{5.24}$$

where $\boldsymbol{Y}_k$ contains the vectors $\boldsymbol{H}_{kj} \boldsymbol{v}_j$ with $j \neq k$ as columns. It is straightforward to show that the optimal SINR γ_k^{opt} cannot be further improved by changing any $\boldsymbol{v}_j$ with $j \neq k$ only when $\boldsymbol{Y}_k^{\mathrm{H}} \boldsymbol{H}_{kk} \boldsymbol{v}_k = \mathbf{0}$, i. e., when $\boldsymbol{v}_j^{\mathrm{H}} \boldsymbol{H}_{kj}^{\mathrm{H}} \boldsymbol{H}_{kk} \boldsymbol{v}_k = 0$ for all $j \neq k$.

Altruism in the MIMO case therefore requires that interference and desired signal are orthogonal at every receiver. It is, however, not trivial to compute the altruistic solutions. Furthermore, the altruistic solutions do not appear to have any special relevance in terms of, e. g., asymptotic optimality. As we discuss in the following section, the desirable property in MIMO systems is that the interference and desired signal are *linearly independent* rather than orthogonal at the receivers. We therefore do not further pursue the notion of altruism here.

5.3 Low- and High-SNR Asymptotics for the Rate Utility

Little is known about the region of achievable SINR K-tuples $(\gamma_1, \ldots, \gamma_K)$ with single-stream beamforming. It is evident that the SINR region is comprehensive: for a fixed set of receive filters, the system is equivalent to a MISO system in which the effective channel vectors are the combination of the channel matrices and receive filters; the MIMO SINR region then is the union of the effective MISO regions resulting from all possible sets of receive filters, each of which is a comprehensive region. Further results, such as convexity conditions or useful parametrizations of the Pareto boundary, however, remain elusive. Instead, the analysis in the literature focuses on the characterization of high-SNR optimality results for the achievable rate utility. In the following we discuss these results and begin by stating the low-SNR optimum.

5.3.1 The Low-SNR Optimal Strategy

Instead of rigorously deriving the low-SNR optimum from the KKT conditions of the sum rate maximization problem as we did in the previous chapters for the SISO and MISO case, we follow a more intuitive approach here. Unless stated otherwise, we assume unit-norm receive filters from here on.

For $\sigma^2 \to \infty$ the denominator of the SINR γ_k is dominated by the noise term and the interference terms can be neglected. Thus, we drop the interference terms in (5.8) and assume that $\gamma_k = |\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k|^2/\sigma^2$ so that the low-SNR optimal strategy is

$$\left(\boldsymbol{g}_k^{\mathrm{ME}}, \boldsymbol{v}_k^{\mathrm{ME}}\right) = \underset{(\boldsymbol{g}_k, \boldsymbol{v}_k)}{\arg\max}\, |\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k|^2 \qquad \text{s. t.:} \quad \|\boldsymbol{g}_k\|_2^2 = 1 \quad \text{and} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1. \tag{5.25}$$

For a fixed receive filter $\boldsymbol{g}_k$, the transmitter effectively sees a MISO channel with the channel vector $\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk}$. From (4.17) we know that the selfish beamformer for this channel is $\boldsymbol{v}_k = \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^*/\|\boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^*\|_2$. Consequently, the receive filter $\boldsymbol{g}_k^{\mathrm{ME}}$ must maximize $|\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^*|^2/\|\boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^*\|_2^2 = \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^*$.

It is known, e. g., from the properties of matrix norms [37, Section 5.6], that the vector $\boldsymbol{g}_k^{\mathrm{ME}*}$ that maximizes this expression is the principal eigenvector of the matrix $\boldsymbol{H}_{kk} \boldsymbol{H}_{kk}^{\mathrm{H}}$, or, if we examine the SVD $\boldsymbol{H}_{kk} = \boldsymbol{U}_k \boldsymbol{\Sigma}_k \boldsymbol{V}_k^{\mathrm{H}}$, where the diagonal entries of $\boldsymbol{\Sigma}_k$ are ordered from highest to lowest, $\boldsymbol{g}_k^{\mathrm{ME}*} = \boldsymbol{U}_k \boldsymbol{e}_1$. From $\boldsymbol{v}_k = \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^*/\|\boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^*\|_2$ it follows that $\boldsymbol{v}_k^{\mathrm{ME}} = \boldsymbol{V}_k \boldsymbol{e}_1$, i. e., $\boldsymbol{v}_k^{\mathrm{ME}}$ is the principal eigenvector of the matrix $\boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{H}_{kk}$.

The SINR (neglecting the interference) resulting from the low-SNR optimal receive filter $\boldsymbol{g}_k^{\mathrm{ME}}$ and beamformer $\boldsymbol{v}_k^{\mathrm{ME}}$ is $\gamma_k = \lambda_k/\sigma^2$, where λ_k is the maximum eigenvalue of both the matrices $\boldsymbol{H}_{kk} \boldsymbol{H}_{kk}^{\mathrm{H}}$ and $\boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{H}_{kk}$; hence the super-script "ME", which stands for "maximum eigenvalue/eigenvector".

5.3.2 High-SNR Slope and Offset

Let us assume a fixed K-tuple of unit-norm receive filters $(\boldsymbol{g}_1, \ldots, \boldsymbol{g}_K)$ and a fixed K-tuple of beamformers $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K)$, which we refer to as the *strategy*. The rate utility of user k is

$$R_k = \log(1 + \gamma_k) = \log\left(1 + \frac{|\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k|^2}{\sum_{j \neq k} |\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kj} \boldsymbol{v}_j|^2 + \sigma^2}\right) \tag{5.26}$$

which is zero if $|\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k|^2 = 0$. To analyze the rate of an active user k with $|\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k|^2 > 0$ in the high-SNR regime, i. e., for $\sigma^{-2} \to \infty$, we distinguish between two cases:

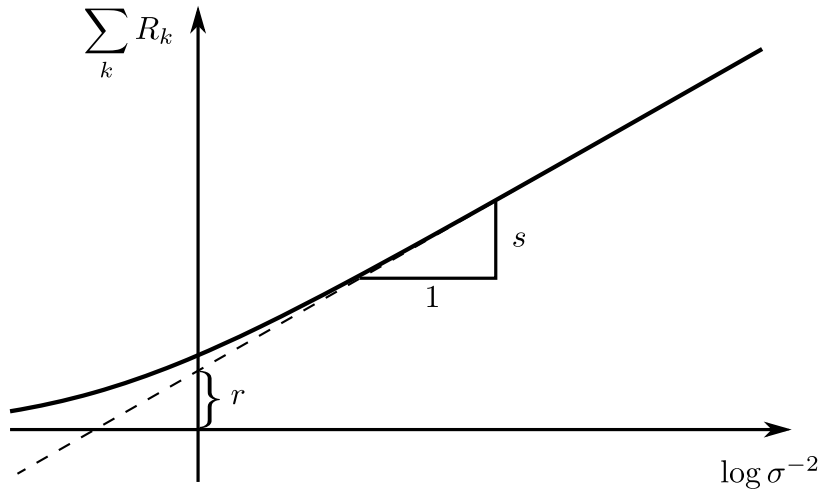


Figure 5.1: Slope s and offset r of the high-SNR asymptote of the sum rate plotted over the logarithmic SNR

1) $\sum_{j\neq k}|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_{kj}\boldsymbol{v}_j|^2 = 0$:
By stating the rate of user k as

$$R_k = \log\left(1+\frac{|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_{kk}\boldsymbol{v}_k|^2}{\sigma^2}\right) = \log\sigma^{-2}+\log|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_{kk}\boldsymbol{v}_k|^2+\log\left(1+\frac{\sigma^2}{|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_{kk}\boldsymbol{v}_k|^2}\right) \quad (5.27)$$

and noting that the last summand approaches zero as $\sigma^{-2}\to\infty$, it clear that $R_k\to\infty$ as $\sigma^{-2}\to\infty$.

2) $\sum_{j\neq k}|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_{kj}\boldsymbol{v}_j|^2 > 0$:

$$\lim_{\sigma^{-2}\to\infty} R_k = \log\left(1+\frac{|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_{kk}\boldsymbol{v}_k|^2}{\sum_{j\neq k}|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_{kj}\boldsymbol{v}_j|^2}\right) \quad (5.28)$$

i. e., R_k approaches a constant as $\sigma^{-2}\to\infty$.

Therefore, as $\sigma^{-2}\to\infty$, the sum rate is

$$\sum_k R_k = s\cdot\log\sigma^{-2}+r+\mathrm{o}(1) \quad (5.29)$$

where s is the number of users k for which $\sum_{j\neq k}|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_{kj}\boldsymbol{v}_j|^2 = 0$ and $|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_{kk}\boldsymbol{v}_k|^2 > 0$.

At high SNR, the sum rate achieved by a given strategy is thus fully described by the noise power σ^2 and the two parameters s and r. In particular, when the sum rate is plotted over $\log\sigma^{-2}$, it approaches a linear asymptote at high SNR, cf. Figure 5.1; the slope of the asymptote is s and the y-axis intercept is r. We therefore refer to r as the sum rate *offset;* note that the sum rate offset, as opposed to the sum rate, can be negative. For the special case in which all users are active and experience no interference, i. e., all users fall under the first case, the slope is $s = K$ and the sum rate offset is

$$r = \sum_k \log|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_{kk}\boldsymbol{v}_k|^2. \quad (5.30)$$

As an asymptote with a higher slope lies above an asymptote with a lower slope for sufficiently high SNR, it is necessary for the high-SNR sum-rate optimal strategy to achieve the maximum value of s. If more than one strategy achieves the maximum slope, the candidate with the highest sum rate offset r is the high-SNR optimum.

5.3.3 Example: $K = 3$ Users, $N_k = M_k = 2$ Antennas

To illustrate the properties of the high-SNR optimum, we examine a scenario with $K = 3$ users and $N_k = M_k = 2$ antennas at every transmitter and receiver k. We assume that all channel matrices $\boldsymbol{H}_{kj} \in \mathbb{C}^{2\times 2}$ are invertible. In the following we investigate whether it is possible to achieve an asymptotic slope of $s = 3$ in this setting.

Since for $s = K = 3$ all three users must be active and experience zero interference after the receive filter, we can state the following necessary and sufficient conditions:

$$\boldsymbol{g}_1^{\mathrm{T}}\boldsymbol{H}_{11}\boldsymbol{v}_1 \neq 0 \qquad \boldsymbol{g}_1^{\mathrm{T}}\boldsymbol{H}_{12}\boldsymbol{v}_2 = 0 \qquad \boldsymbol{g}_1^{\mathrm{T}}\boldsymbol{H}_{13}\boldsymbol{v}_3 = 0 \tag{5.31}$$

$$\boldsymbol{g}_2^{\mathrm{T}}\boldsymbol{H}_{21}\boldsymbol{v}_1 = 0 \qquad \boldsymbol{g}_2^{\mathrm{T}}\boldsymbol{H}_{22}\boldsymbol{v}_2 \neq 0 \qquad \boldsymbol{g}_2^{\mathrm{T}}\boldsymbol{H}_{23}\boldsymbol{v}_3 = 0 \tag{5.32}$$

$$\boldsymbol{g}_3^{\mathrm{T}}\boldsymbol{H}_{31}\boldsymbol{v}_1 = 0 \qquad \boldsymbol{g}_3^{\mathrm{T}}\boldsymbol{H}_{32}\boldsymbol{v}_2 = 0 \qquad \boldsymbol{g}_3^{\mathrm{T}}\boldsymbol{H}_{33}\boldsymbol{v}_3 \neq 0. \tag{5.33}$$

From the three conditions in (5.31) it follows that $\boldsymbol{g}_1 \neq \boldsymbol{0}$ and that $\boldsymbol{g}_1^{\mathrm{T}}(\boldsymbol{H}_{12}\boldsymbol{v}_2\boldsymbol{v}_2^{\mathrm{H}}\boldsymbol{H}_{12}^{\mathrm{H}} + \boldsymbol{H}_{13}\boldsymbol{v}_3\boldsymbol{v}_3^{\mathrm{H}}\boldsymbol{H}_{13}^{\mathrm{H}})\boldsymbol{g}_1^{*} = 0$. Therefore, the matrix $\boldsymbol{H}_{12}\boldsymbol{v}_2\boldsymbol{v}_2^{\mathrm{H}}\boldsymbol{H}_{12}^{\mathrm{H}} + \boldsymbol{H}_{13}\boldsymbol{v}_3\boldsymbol{v}_3^{\mathrm{H}}\boldsymbol{H}_{13}^{\mathrm{H}} \in \mathbb{C}^{2\times 2}$ may not have full rank and consequently $\boldsymbol{H}_{12}\boldsymbol{v}_2 \parallel \boldsymbol{H}_{13}\boldsymbol{v}_3$. Similar conditions follow from (5.32) and (5.33), and we obtain the necessary conditions

$$\boldsymbol{H}_{12}\boldsymbol{v}_2 = \lambda_1\boldsymbol{H}_{13}\boldsymbol{v}_3 \qquad \boldsymbol{H}_{23}\boldsymbol{v}_3 = \lambda_2\boldsymbol{H}_{21}\boldsymbol{v}_1 \qquad \boldsymbol{H}_{31}\boldsymbol{v}_1 = \lambda_3\boldsymbol{H}_{32}\boldsymbol{v}_2 \tag{5.34}$$

where $\lambda_k \in \mathbb{C}$ and $\lambda_k \neq 0$ for all $k \in \{1, 2, 3\}$. By combining these three conditions to eliminate $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$, we observe that $\boldsymbol{v}_1$ must fulfill

$$\boldsymbol{v}_1 = \lambda_3\boldsymbol{H}_{31}^{-1}\boldsymbol{H}_{32}\boldsymbol{v}_2 = \lambda_1\lambda_3\boldsymbol{H}_{31}^{-1}\boldsymbol{H}_{32}\boldsymbol{H}_{12}^{-1}\boldsymbol{H}_{13}\boldsymbol{v}_3 = \lambda_1\lambda_2\lambda_3\boldsymbol{H}_{31}^{-1}\boldsymbol{H}_{32}\boldsymbol{H}_{12}^{-1}\boldsymbol{H}_{13}\boldsymbol{H}_{23}^{-1}\boldsymbol{H}_{21}\boldsymbol{v}_1 \tag{5.35}$$

i. e., the beamformer $\boldsymbol{v}_1$ must be an eigenvector of the matrix $\boldsymbol{H}_{31}^{-1}\boldsymbol{H}_{32}\boldsymbol{H}_{12}^{-1}\boldsymbol{H}_{13}\boldsymbol{H}_{23}^{-1}\boldsymbol{H}_{21}$. For a given beamformer $\boldsymbol{v}_1$, the beamformers $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ can in turn be determined with (5.34), where λ_2 and λ_3 are chosen such that the beamformers have unit norm. Since the so-found 3-tuple of beamformers fulfills (5.34), a 3-tuple of unit-norm receive filters $(\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3)$ can be constructed such that all six zero-interference conditions of (5.31)–(5.33) are fulfilled. The number of distinct strategies for which the interference is zero after the receive filters simply is the number of linearly independent solutions to the above eigenvector problem. In most cases, a full-rank 2×2 matrix will have two linearly independent eigenvectors, but it is also possible to construct a set of channel matrices where there is only one solution or where any vector is a solution.

The key property of these strategies is that the interference is *aligned* at the receivers: the interference components at the first receiver, $\boldsymbol{H}_{12}\boldsymbol{v}_2$ and $\boldsymbol{H}_{13}\boldsymbol{v}_3$, coincide so that they span only a one-dimensional subspace of $\mathbb{C}^2$; the same holds for the interference components at the second and third receiver. Therefore, the receive filters can project the received signal into a subspace that is orthogonal to the interference and thereby fully eliminate the interference.

We have, however, not yet considered the three necessary conditions $\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_{kk}\boldsymbol{v}_k \neq 0$ for all $k \in \{1, 2, 3\}$. Clearly, when applied to the first receiver, this means that $\boldsymbol{H}_{11}\boldsymbol{v}_1 \nparallel \boldsymbol{H}_{12}\boldsymbol{v}_2$. With (5.34) it follows that $\boldsymbol{v}_1$ may not be an eigenvector of the matrix $\boldsymbol{H}_{31}^{-1}\boldsymbol{H}_{32}\boldsymbol{H}_{12}^{-1}\boldsymbol{H}_{11}$. Similar conditions can be formulated for the second and third user. Therefore, if, e. g., $\boldsymbol{H}_{11} = \boldsymbol{H}_{13}\boldsymbol{H}_{23}^{-1}\boldsymbol{H}_{21}$, the nine conditions of (5.31)–(5.33) cannot be simultaneously fulfilled, i. e., $s = 3$ is not achievable. It is intuitive, though, that these are in some way degenerate cases, and that "in most cases" all three users can cancel the interference without canceling the desired signal. This notion will be stated more precisely in the following section.

When the channel matrices do not have full rank, matters are further complicated. In particular, it is fairly easy to construct channel matrices such that further users can be added to the system and, e. g., $K = 4$ users can cancel out all interference, thus achieving a slope of $s = 4$. Altogether we observe that determining the maximal slope s for a given set of channel realizations is not trivial: the rank of the channel matrices certainly plays a role, but even if all matrices have full rank, degenerate conditions can occur.

In this example setting we were able to explicitly compute the beamformers that achieve the maximal slope s. This is in general not possible: for larger scenarios it turns out that strategies with zero interference can only be found by means of iterative algorithms and that reliably finding all such solutions appears to be computationally infeasible.

5.3.4 Achievable Slope for Random Channels: Spatial Interference Alignment

It is shown in [75] that for a given channel realization the problem of verifying whether a certain slope s is achievable is NP-hard when $N_k \geq 3$ and $M_k \geq 3$ for all $k \in \{1, \ldots, K\}$. Therefore, determining the maximal slope is also NP-hard. However, in [34] a very useful and comparatively simple achievability check was proposed that is valid for *almost all* channel realizations with a certain antenna configuration; we discuss this feasibility test in the following.

We assume a system with K users with a given antenna configuration $N_1, \ldots, N_K$ and $M_1, \ldots, M_K$. Furthermore, all users are active, i. e., $\boldsymbol{v}_k \neq \boldsymbol{0}$ for all $k \in \{1, \ldots, K\}$. The question that we investigate is whether the asymptotic slope $s = K$ can be achieved. In order to determine the maximal slope s of a system in which $s = K$ is not feasible, it is thus necessary to remove users from the system until a feasible configuration of active users is reached.

For the slope $s = K$ to be achieved, the interference after the receive filters of all K users must be zero. Therefore, it is necessary that the following system of equations is fulfilled:

$$\begin{array}{cccc} * & \boldsymbol{g}_1^{\mathrm{T}} \boldsymbol{H}_{12} \boldsymbol{v}_2 = 0 & \ldots & \boldsymbol{g}_1^{\mathrm{T}} \boldsymbol{H}_{1K} \boldsymbol{v}_K = 0 \quad (5.36) \\ \boldsymbol{g}_2^{\mathrm{T}} \boldsymbol{H}_{21} \boldsymbol{v}_1 = 0 & * & \ldots & \boldsymbol{g}_2^{\mathrm{T}} \boldsymbol{H}_{2K} \boldsymbol{v}_K = 0 \quad (5.37) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{g}_K^{\mathrm{T}} \boldsymbol{H}_{K1} \boldsymbol{v}_1 = 0 & \boldsymbol{g}_K^{\mathrm{T}} \boldsymbol{H}_{K2} \boldsymbol{v}_2 = 0 & \ldots & * \quad (5.38) \end{array}$$

Since for each user $K - 1$ scalar interference terms must be zero, altogether $K(K - 1)$ equations must be fulfilled. Furthermore, note that the equations are bilinear in the variables $\boldsymbol{g}_k$ and $\boldsymbol{v}_k$; the zero-interference conditions therefore form a system of multivariate complex polynomial equations. The key idea in checking for feasibility is the notion from algebraic geometry that a system of polynomials with "generic" coefficients can be solved if and only if there are no more equations than variables.

The variables in our system are the entries of the vectors $\boldsymbol{g}_k$ and $\boldsymbol{v}_k$. It is clear, though, that there is some "redundance" in the variables in the sense that any vector $\boldsymbol{g}_k$ or $\boldsymbol{v}_k$ can be multiplied with a non-zero complex scalar without changing the validity of the equations. We can therefore arbitrarily set the first entry of all vectors $\boldsymbol{g}_k$ and $\boldsymbol{v}_k$ to one and retain the properties of our system of equations.[2] Now, the number of "free" complex variables in the vector $\boldsymbol{v}_k$ is $N_k - 1$ and in the vector $\boldsymbol{g}_k$ is $M_k - 1$.

[2]Note that we are not concerned about the unit norm constraints on beamformers or receive filters at this point, as any solution to the zero-interference conditions can be rescaled to fulfill the constraints.

Definition 5.1 (cf. [34])**.** A system is *proper* if and only if, for any subset of the above equations, the number of variables involved is at least as large as the number of equations.[3]

While it can be tedious to determine whether a system is proper as the number of subsets of the $K(K-1)$ equations can be very large, we can formulate a simple necessary condition by comparing the total number of equations and the total number of variables:

$$\sum_k N_k + \sum_k M_k - 2K \geq K(K-1). \tag{5.39}$$

In symmetric settings, where $N_1 = \ldots = N_K = N$ and $M_1 = \ldots = M_K = M$, it is shown in [34] that comparing the total number of equations and variables is in fact sufficient for checking whether a system is proper; in this case the condition simplifies to

$$N + M - 1 \geq K \tag{5.40}$$

and is both necessary and sufficient. As an example, a symmetric MISO system with N antennas at each transmitter is proper if and only if $N \geq K$. Similarly, a MIMO system in which all channel matrices are square, i. e., where every transmitter and every receiver has N antennas, is proper if and only if $2N - 1 \geq K$.

Theorem 5.1 (cf. [34])**.** *If the entries of the channel matrices $\boldsymbol{H}_{kj}$ for all $(k, j) \in \{1, \ldots, K\}^2$ are random, statistically independent, and have a continuous distribution, then for a proper system the slope $s = K$ is achievable with probability one, and for an improper system it is achievable with probability zero.*

The proof relies on Bernstein's theorem, a fundamental result in algebraic geometry, which characterizes the solvability of sparse multivariate polynomial systems with generic coefficients, cf. [76, Chapter 7]. In the proof in [34] the connection is made between independent random channel coefficients and the genericity of the polynomial coefficients. Furthermore, it is intuitive that the remaining inequality conditions $\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k \neq 0$ for all $k \in \{1, \ldots, K\}$ are fulfilled with high probability for any beamformers and receive filters that solve the system of polynomial equations, since the direct channel matrices $\boldsymbol{H}_{kk}$ do not appear in the system of equations and are statistically independent of the cross channels.

As in the example setting discussed in the previous section, we can draw some conclusions on the subspaces spanned by the interference components at the K receivers. Let us consider, e. g., a symmetric setting where each transmitter and receiver has N antennas; the received signal vectors $\boldsymbol{y}_k$ in such a system are in $\mathbb{C}^N$. If at receiver k all interference components $\boldsymbol{H}_{kj}\boldsymbol{v}_j$ with $j \neq k$ span an $N-1$-dimensional subspace of $\mathbb{C}^N$, there is a one-dimensional subspace that is orthogonal to all received interference, on to which the receive filter $\boldsymbol{g}_k$ can project the received signal, thereby removing all interference. While it is clear that if there are exactly $N-1$ interferers, i. e., if $K = N$, the subspace spanned by the interference components is at most $N-1$-dimensional, it is in fact possible to accomodate more users: according to the above condition, the system is proper for up to $K = 2N - 1$ users. Therefore, up to $K - 1 = 2N - 2$ interferers can "fit into" an $N-1$-dimensional subspace of $\mathbb{C}^N$ at each receiver. Since the dimensionality of the subspace is lower than the number of vectors spanning it, we say that the interference is *aligned.* From here on we therefore refer to a $2K$-tuple of receive filters and beamformers that achieves the slope $s = K$ as an *aligned solution.*

[3] The term "proper" in this context has no connection to its use in Section 2.1 to describe the "circular symmetry" property of complex random variables.

Furthermore, we refer to a system as *fully loaded* when the system is proper and condition (5.39) is fulfilled with equality, i. e., the total number of equations is equal to the total number of variables. Fully loaded systems are significant in that they have a finite number of aligned solutions, as is shown in [34]; for example, in the three-user two-antenna setting discussed in the previous section, we found that there are generally two distinct solutions. If the system is proper, but not fully loaded, on the other hand, the aligned solutions in general form a manifold.[4]

We recall that among all aligned solutions (of which all achieve the maximal slope $s = K$), the one with the highest sum rate offset r is the high-SNR optimal strategy. The high-SNR optimum in a fully loaded system can therefore be determined by computing all aligned solutions (which depend only on the cross channels) and comparing the resulting offset r (which depends only on the direct channels, cf. (5.30)). In order to obtain an intuition regarding the complexity of such a procedure, we examine the number of aligned solutions for fully loaded systems.

The number of solutions to a sparse polynomial system with an equal number of variables and equations is again characterized by Bernstein's theorem and can be computed by means of mixed volumes of Newton polytopes (we refer to [34, 76] for details). The computation can be done by hand only for very small settings and does not scale well with the system dimensions. There are, however, specialized software libraries for this purpose: in Table 5.1 we give the results of applying such software[5] to our system of zero-interference conditions for different symmetric antenna configurations. We note that on state-of-the-art computer hardware the software required approximately one hour to obtain the result for the setting with $K = 5$ users and $N = M = 3$ antennas and was not able to compute the number of solutions for larger system dimensions in reasonable time.

In [78], it was observed that the number of aligned solutions for an i. i. d. channel model is identical to the number of aligned solutions in a different channel model, in which all channel matrices have rank one. For the rank-one model, computation of the number of solutions is significantly simpler and can be performed within reasonable time for some larger antenna configurations. We reproduce the results given in [78] in Table 5.2.

While the exact number of solutions is difficult to compute for even larger systems, a simple upper bound follows from Bézout's theorem (again, cf. [34]). The bound is obtained by taking the product of the degrees of all equations. As the degree of a bilinear equation is two, the bound in our case is $2^{K(K-1)}$ if $N_k > 1$ and $M_k > 1$ for all $k \in \{1, \ldots, K\}$; when a node has only one antenna, the number of variables associated with that node is zero and the degree of the respective equations is one instead of two. This bound on the number of solutions is tight when the polynomial system is dense, i. e., if all possible monomials with a degree less than or equal to the total degree of the equation have non-zero coefficients; as our system of bilinear equations is sparse, the bound can be very loose, as also becomes apparent in Tables 5.1 and 5.2.

Another, tighter, upper bound on the number of solutions can be computed by means of the *multi-homogeneous Bézout number* (e. g., [79]). Computing the multi-homogeneous Bézout number, however, is a combinatorial problem that also does not scale well with the system dimensions

[4] Another notable property of fully loaded systems is that single-stream transmission is sum-rate optimal at high SNR. This is a direct consequence of the fact that removing a user and adding a stream to one of the remaining users reduces the number of free variables by a higher margin than the number of equations. For the details of counting variables and equations in multi-stream systems we refer to [34].

[5] The software used is the "Mixed Volume Library" by I. Z. Emiris, which is based on the algorithm presented in [77] and can be downloaded at http://www-sop.inria.fr/galaad/logiciels/emiris/soft_geo.html.

N	M	K	Solutions	Upper bound (Bézout)
N	1	N	1	1
2	2	3	2	64
3	2	4	9	4096
4	2	5	44	1 048 576
3	3	5	216	1 048 576

Table 5.1: Number of solutions for different fully loaded symmetric scenarios. The results were computed with the software "Mixed Volume Library" by I. Z. Emiris, cf. [77]; for larger scenarios, computation on a state-of-the-art computer proved to be infeasible. Note that the values of M and N can be exchanged without changing the number of solutions or the upper bound. The results for $K = 3$ and $K = 4$ were previously presented in [34].

N	M	K	Solutions	Upper bound (Bézout)
5	2	6	265	$1.07 \cdot 10^9$
4	3	6	7570	$1.07 \cdot 10^9$
6	2	7	1854	$4.40 \cdot 10^{12}$
5	3	7	357 435	$4.40 \cdot 10^{12}$
4	4	7	1 975 560	$4.40 \cdot 10^{12}$
7	2	8	14 833	$7.21 \cdot 10^{16}$
6	3	8	22 040 361	$7.21 \cdot 10^{16}$
5	4	8	749 649 145	$7.21 \cdot 10^{16}$

Table 5.2: Number of solutions for larger fully loaded symmetric scenarios as presented in [78]. In order to determine these results, a rank-one channel model was assumed and the number of aligned solutions was shown to be the same as in a full-rank channel model.

and in particular does not help us estimate how the number of solutions in our case behaves for large K, N, and M. We therefore do not further pursue this bound here.

The solution numbers for some smaller fully loaded systems shown here indicate that the number of solutions grows very rapidly in the system dimensions. It seems likely that determining all aligned solutions for a fully loaded system is highly impractical for a MIMO interference network with many users; therefore, reliably finding the high-SNR optimal strategy with acceptable computational effort appears to be possible only in small systems.

5.3.5 Statistics of the Offset for a Gaussian Channel Model

In this section we examine the statistics of the sum rate offset r of the aligned solutions in a fully loaded system for an i. i. d. Gaussian channel model. This analysis was presented in [80].

We begin by examining the statistics of $r = \sum_k \log|\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k|^2$ for a fixed strategy, i. e., a $2K$-tuple of unit-norm vectors $\boldsymbol{g}_k \in \mathbb{C}^{M_k}$ and $\boldsymbol{v}_k \in \mathbb{C}^{N_k}$ for all $k \in \{1, \ldots, K\}$. In our i. i. d. Gaussian channel model, the direct channels $\boldsymbol{H}_{kk}$ are random matrices where each element has a complex Gaussian distribution with mean zero and variance one and the elements are statistically independent. We make the following observations:

- The random variable $h_m = \boldsymbol{e}_m^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k$ is zero-mean complex Gaussian for any $m \in \{1, \ldots, M_k\}$, as it is the weighted sum of zero-mean complex Gaussian variables. The vari-

ance of h_m is $\mathrm{E}[|h_m|^2] = \boldsymbol{v}_k^{\mathrm{H}} \mathrm{E}[\boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{e}_m \boldsymbol{e}_m^{\mathrm{T}} \boldsymbol{H}_{kk}] \boldsymbol{v}_k = \boldsymbol{v}_k^{\mathrm{H}} \mathbf{I} \boldsymbol{v}_k = 1$. Furthermore, h_m and h_n are statistically independent for $n \neq m$.

- The random variable $\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k = \boldsymbol{g}_k^{\mathrm{T}} \sum_m \boldsymbol{e}_m \boldsymbol{e}_m^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k = \sum_m \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{e}_m h_m$ is zero-mean complex Gaussian, as it is the weighted sum of zero-mean complex Gaussian variables. The variance is $\mathrm{E}[|\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k|^2] = \boldsymbol{g}_k^{\mathrm{T}} \mathrm{E}\left[\sum_m \boldsymbol{e}_m \boldsymbol{e}_m^{\mathrm{T}} |h_m|^2\right] \boldsymbol{g}_k^* = 1$.
- The power of the desired signal of user k can be written as $|\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k|^2 = a_{\mathrm{r}}^2 + a_{\mathrm{i}}^2$, where a_{r} and a_{i} are the real and imaginary parts of $\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k$, respectively; a_{r} and a_{i} are independent real Gaussian random variables with mean zero and variance $1/2$. Consequently, $2|\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k|^2 = \left(\sqrt{2} a_1\right)^2 + \left(\sqrt{2} a_2\right)^2$ has a $\chi^2(2)$ distribution, cf. [81, Chapter 7]; the PDF of $z_k = |\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k|^2$ is

$$f_{z_k}(z_k) = \mathrm{e}^{-z_k} . \tag{5.41}$$

- The distribution of $r_k = \log|\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k|^2 = \log z_k$ can be determined with a transformation of the distribution of the random variable z_k (cf. [81, Chapter 5]):

$$f_{r_k}(r_k) = \left. \frac{f_{z_k}(z_k)}{\left| \frac{\partial \log z_k}{\partial z_k} \right|} \right|_{z_k = \mathrm{e}^{r_k}} = \mathrm{e}^{r_k} \, \mathrm{e}^{-\mathrm{e}^{r_k}} . \tag{5.42}$$

  Consequently, the random variable $-r_k$ has a standard Gumbel distribution, cf. [82, Section 10.5]; the mean is $\mathrm{E}[r_k] = -\gamma$ and the variance is $\mathrm{E}\left[(r_k + \gamma)^2\right] = \pi^2/6$, where $\gamma \approx 0.5772$ is the Euler-Mascheroni constant.
- Since the random variables $r_1, \ldots, r_K$ are statistically independent, the mean and variance of the sum rate offset $r = \sum_k r_k$ are

$$\mathrm{E}[r] = -K\gamma \qquad \text{and} \qquad \mathrm{E}\left[(r + K\gamma)^2\right] = \frac{K\pi^2}{6}. \tag{5.43}$$

The statistics of r for a fixed $2K$-tuple of unit-norm receive filters and beamformers do not depend on any $\boldsymbol{g}_k$ or $\boldsymbol{v}_k$. Therefore, if we choose a different $2K$-tuple $(\boldsymbol{g}_1, \ldots, \boldsymbol{g}_K, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_K)$ for each channel realization, the statistics of r are the same, as long as the choice of the strategy is independent of the realization of the direct channels. As was noted previously, the aligned solutions depend only on the cross channels $\boldsymbol{H}_{kj}$ with $k \neq j$; therefore, if we select a random aligned solution without considering the direct channels $\boldsymbol{H}_{kk}$ for each channel realization, the asymptotic slope is $s = K$ and the mean and variance of the sum rate offset r are $-K\gamma$ and $K\pi^2/6$, respectively.

If more than one aligned solution is known, selecting one at random is not a very good strategy; instead, we would like to choose the aligned solution for which r is highest. In the following we approximate the performance of this strategy.

5.3.5.1 Approximation for the Best of L Aligned Solutions

As was discussed in Section 5.3.4 and is indicated by the results in Tables 5.1 and 5.2, it appears that the number of aligned solutions grows very rapidly in the system dimensions. Therefore, determining all aligned solutions and comparing the resulting sum rate offset r is considered infeasible for larger systems. Instead, we assume that we are able to determine a random subset of all aligned solutions consisting of L solutions. A method to obtain different aligned solutions with reasonable complexity will be discussed later in this chapter.

We begin with some notation: as we are examining a fully loaded system, the total number of aligned solutions T is finite and the same for almost all channel realizations. We refer to one aligned solution, a $2K$-tuple of unit-norm receive filters and beamforming vectors, as

$$s^t = \left(\boldsymbol{g}_1^t, \ldots, \boldsymbol{g}_K^t, \boldsymbol{v}_1^t, \ldots, \boldsymbol{v}_K^t\right) \tag{5.44}$$

where for this analysis we use the super-script $t \in \{1, \ldots, T\}$ to denote the solution index. The set of all solutions for a given channel realization is

$$\mathcal{S} = \left\{s^1, \ldots, s^T\right\}. \tag{5.45}$$

The solution set $\mathcal{S}$ is a function of all cross channels $\boldsymbol{H}_{kj}$ with $k \neq j$.

We furthermore define the set $\mathcal{L}$ that contains the indices of the L solutions that we compare in terms of sum rate offset, i. e.,

$$\mathcal{L} = \{t_1, \ldots, t_L\} \tag{5.46}$$

where $1 \leq t_\ell \leq T$ for all $\ell \in \{1, \ldots, L\}$. The random variables are the channel matrices $\boldsymbol{H}_{kj}$ for all $(k, j) \in \{1, \ldots, K\}^2$, which are distributed according to the i. i. d. Gaussian channel model, and the set $\mathcal{L}$, which is a random subset with cardinality L of the set of all indices $\{1, \ldots, T\}$, where each possible subset has the same probability; consequently, for each new channel realization, a new subset $\mathcal{L}$ is drawn.

The sum rate offset r is a function of the direct channel matrices $\boldsymbol{H}_{kk}$ as well as the $2K$-tuple of receive filters and beamformers, i. e., the aligned solution; the aligned solution in turn is a function of the cross channel matrices $\boldsymbol{H}_{kj}$ with $k \neq j$ and the solution index. Altogether, r therefore is a function of all channel matrices and the solution index t. To express this relationship, we denote the offset as $r(\boldsymbol{H}, t)$.

Our goal is to compute the mean offset of the best solution index in the set $\mathcal{L}$, i. e.,

$$\bar{r} = \mathrm{E}\left[\max_{\ell \in \{1,\ldots,L\}} r(\boldsymbol{H}, t_\ell)\right]. \tag{5.47}$$

The mean offset $\bar{r}$ is not straightforward to compute. In particular, we do not know how the offset r of the different aligned solutions is distributed for a given channel realization. With some simplifying assumptions, however, we are able to approximate $\bar{r}$.

First of all, note that considering a fixed index ℓ effectively means choosing a random solution, as t_ℓ for a fixed ℓ is a random index between one and the total number of solutions T. From (5.43) it is therefore clear that

$$\mathrm{E}\left[r(\boldsymbol{H}, t_\ell)\right] = -K\gamma \qquad \text{and} \qquad \mathrm{E}\left[\left(r(\boldsymbol{H}, t_\ell) + K\gamma\right)^2\right] = \frac{K\pi^2}{6} \tag{5.48}$$

for any fixed $\ell \in \{1, \ldots, L\}$. Furthermore, let us make the following simplifying approximations:

- The distribution of $r(\boldsymbol{H}, t_\ell)$ for a fixed $\ell \in \{1, \ldots, L\}$ is Gaussian with mean $-K\gamma$ and variance $K\pi^2/6$. The rationale behind this approximation is that r is the sum of K independent random variables; as K becomes large, the Gaussian approximation becomes more accurate according to the *central limit theorem,* cf. [81, Chapter 7].
- The offsets for different fixed indices ℓ, i. e., $r(\boldsymbol{H}, t_1), \ldots, r(\boldsymbol{H}, t_L)$, are statistically independent. This means that knowing the offset of a randomly chosen aligned solution does not contain any information about the offset of the other aligned solutions. Clearly, if the aligned

solutions tend to be "clustered", i. e., if, say, $\boldsymbol{g}_1^{t_1}$ and $\boldsymbol{g}_1^{t_2}$ as well as $\boldsymbol{v}_1^{t_1}$ and $\boldsymbol{v}_1^{t_2}$ tend to be similar, the offsets are not independent, as the offset $r(\boldsymbol{H}, t_1)$ and $r(\boldsymbol{H}, t_2)$ are also similar. Therefore, one assumption behind this approximation is that the vectors $\boldsymbol{g}_k^1, \ldots, \boldsymbol{g}_k^T$ and $\boldsymbol{v}_k^1, \ldots, \boldsymbol{v}_k^T$ are in some way evenly distributed on the M_k- or N_k-dimensional unit sphere for a given channel realization.

The second assumption behind this approximation is that the spectrum of singular values of $\boldsymbol{H}_{kk}$ does not vary strongly from one realization to the next. Assume, e. g., that the offset $r(\boldsymbol{H}, t_1)$ is unusually high: this could contain the information that the singular values of the realization of the direct channels are high; therefore the offset of another aligned solution $r(\boldsymbol{H}, t_2)$ is more likely to be high as well. As we discuss in Section 5.4.2.1, with growing system dimensions M and N, the variations in the spectrum of singular values decrease. Therefore, this assumption in a sense is also a large-system approximation.

For the time being we do not experimentally investigate the validity of these approximations, as we have not yet discussed how to obtain different random aligned solutions; the numerical justification will be given later on, in Section 5.6.1.

With these two approximations, comparing the L solutions with the indices from $\mathcal{L}$ is equivalent to comparing L independent realizations of a Gaussian random variable with mean $-K\gamma$ and variance $K\pi^2/6$. We can now use a result from extreme statistics that characterizes the distribution of the maximum of L independent realizations of a Gaussian random variable, where L is large.

Theorem 5.2 (cf. [82, Example 10.5.3])**.** *Let*

$$u = \max_{n \in \{1,\ldots,L\}} w_n \tag{5.49}$$

where w_n are i. i. d. real Gaussian random variables with mean zero and variance one, then, as $L \to \infty$, the PDF $f_x(x)$ of $x = b_L(u - b_L)$ approaches the Gumbel distribution, i. e.,

$$\lim_{L\to\infty} f_x(x) = \mathrm{e}^{-x}\,\mathrm{e}^{-\mathrm{e}^{-x}} \tag{5.50}$$

where $b_L = \mathrm{Q}^{-1}(1/L)$ and $\mathrm{Q}^{-1}(\cdot)$ is the inverse Q-function. Furthermore, since the mean of the Gumbel distribution is γ,

$$\lim_{L\to\infty} b_L(\mathrm{E}[u] - b_L) = \gamma \tag{5.51}$$

and

$$\mathrm{E}[u] = b_L + \frac{\gamma}{b_L} + \mathrm{o}(1) \tag{5.52}$$

as $L \to \infty$.

The concept of the limiting Gumbel distribution is illustrated in Figure 5.2: with an increasing number of samples L, the limiting distribution is shifted to the right and the variance decreases.

With Theorem 5.2, we approximate the mean of the maximum of L zero-mean unit-variance Gaussian random variables $w_1, \ldots, w_L$ as $b_L + \gamma/b_L$. To obtain random variables with mean $-K\gamma$ and variance $K\pi^2/6$, we scale and shift the w_n:

$$r_{\text{approx}} = \max_{n \in \{1,\ldots,L\}} \pi\sqrt{\frac{K}{6}} \cdot w_n - K\gamma = \pi\sqrt{\frac{K}{6}} \cdot u - K\gamma. \tag{5.53}$$

Our approximated mean for the best out of L randomly chosen aligned solutions in an i. i. d. Gaussian channel model therefore is

$$\bar{r}_{\text{approx}} = -K\gamma + \pi\sqrt{\frac{K}{6}}\left(b_L + \frac{\gamma}{b_L}\right). \tag{5.54}$$

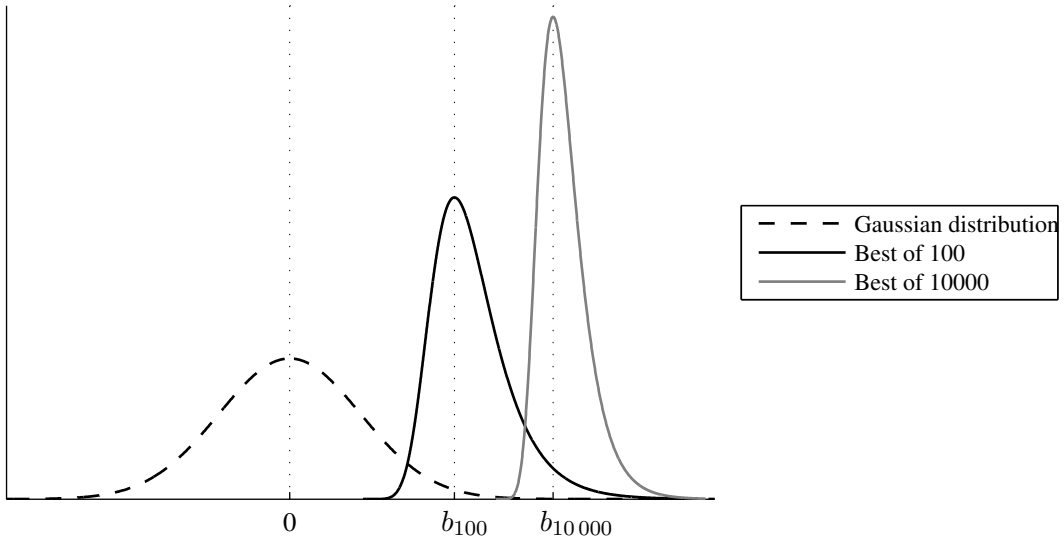


Figure 5.2: Limiting distribution of the maximum of $L_1 = 100$ and $L_2 = 10\,000$ realizations of a unit-variance zero-mean Gaussian random variable.

Some remarks on this approximation:

- The parameter b_L, which governs the gain of trying out L different realizations, grows very slowly with L. In fact, as is shown in [82], $b_L \in \mathrm{O}\big(\sqrt{\log L}\big)$ as $L \to \infty$.
- When L is extremely high, it is uncertain how good the approximation is. In particular, the assumption that the distribution of the offset is Gaussian may be an oversimplification. When the parameter b_L is high, the best of L realizations is on the far right of the "tail" of the Gaussian PDF. Intuitively, small absolute deviations of the actual PDF from the Gaussian approximation may have a large effect on the statistics of the maximum; consider, e. g., the case where the actual PDF is exactly zero above a certain value: the maximum of L realizations can never be above this value, whereas with the Gaussian approximation the shift b_L can be arbitrarily high. Numerically investigating this effect is computationally very difficult, as this would require a very large number L of aligned solutions.
- As the number of solutions is not known for arbitrarily large MIMO scenarios, the approximation cannot be used to characterize the average sum rate offset of the globally high-SNR optimal strategy with increasing system dimensions. We could, however, apply the upper bound from Bézout's theorem, i. e., assume $L = T = 2^{K(K-1)}$. Then, $b_L \in \mathrm{O}(K)$ as $K \to \infty$ and $\bar{r}_{\mathrm{approx}} \in \mathrm{O}\big(K^{3/2}\big)$ as $K \to \infty$.
  To show that this is not a helpful approximation, we make use of a result that is proven later on in Section 5.4.2.1: the squared maximum singular value of the channel matrices $\boldsymbol{H}_{kk}$ is in $\mathrm{O}(N)$ as $N \to \infty$ and therefore in $\mathrm{O}(K)$ for a fully loaded system with $K = 2N - 1$. Consequently, we can upper-bound the sum rate of the global optimum by assuming that all cross channels are zero, in which case each user's individual rate offset is in $\mathrm{O}(\log K)$ and the sum rate offset is in $\mathrm{O}(K \log K)$ as $K \to \infty$; thus, this simple upper bound grows slower than the approximation $\mathrm{O}\big(K^{3/2}\big)$ and therefore applying the bound provided by Bézout's theorem does not yield any useful insights.

Nonetheless, as is shown in Section 5.6.2, for some example scenarios the approximation predicts the numerically simulated behavior very well. In particular, the approximation allows us to better

understand the performance gain that follows from the fact that there are many solutions, as opposed to the MISO case, where there is only one high-SNR optimal solution if the system is fully loaded.

5.4 Non-Iterative Techniques

The aligned strategies discussed in the previous section, which achieve the full potential of the system at infinitely high SNR, can be found by solving a non-trivial system of polynomial equations; it is in general not possible to express these solutions in closed form or determine them with a bounded number of operations such as eigenvalue decompositions or matrix inversions. As the asymptotic case is a simplified special case of the general sum rate maximization problem, it is intuitive that for finite SNR finding good strategies does not become any easier. We therefore cannot expect to obtain close-to-optimal strategies without iterative strategy updates. Nonetheless, the two non-iterative schemes discussed in the following are of special interest: the first will turn out to be a good initialization for the iterative algorithms discussed in Section 5.5 and the second provides some interesting insights in the behavior of the sum rate offset r when the maximal slope s is not achieved.

5.4.1 The Selfish Solution Ignoring the Interference

The maximum eigenvalue/eigenvector (ME) strategy was shown to be optimal at low SNR in Section 5.3.1. Similarly, if all interference channel gains are zero, it is clearly in the best interest of each user k to use $\boldsymbol{g}_k^{\mathrm{ME}}$ and $\boldsymbol{v}_k^{\mathrm{ME}}$, and the ME strategy is the selfish solution. Even when the interference is not zero, we may choose to ignore the interference in order to obtain a simple strategy; intuitively, this is a valid simplification if the noise power is large compared to the interference power.

However, in contrast to the MISO case, for multi-antenna receivers the low-SNR optimal strategy does not in general lead to the selfish solution: if interference is present, user k can improve its own SINR γ_k by changing the beamformer $\boldsymbol{v}_k$ to make the desired signal component "more distinguishable" from the interference components at receiver k and by employing the optimal receive filter $\boldsymbol{g}_k^{\mathrm{opt}}$, which takes into account the received interference.

As the selfish strategy of user k depends on the other users' beamformers $\boldsymbol{v}_j$ with $j \neq k$, the selfish solution cannot be found without iterations. An algorithm in which the strategies are iteratively updated in a selfish way will be discussed in Section 5.5.1.

5.4.2 Successive Zero-Forcing Strategies

Similar to the technique discussed in Section 4.3.3, we can also obtain a Pareto optimal strategy in the MIMO case by assigning strategies to the users one after another and requiring that a newly allocated strategy does not decrease the SINR of previously allocated users. We again assume for ease of notation that the strategies are allocated in the order of the user indices $1, \ldots, K$; the extension to arbitrary orders is straightforward.

As the first user experiences no interference, it can maximize its SINR γ_1 by applying the ME strategy. The objective of the second user is to maximize its own SINR γ_2 with the constraint that no interference may be caused to the first user, i. e., that $\boldsymbol{g}_1^{\mathrm{T}} \boldsymbol{H}_{12} \boldsymbol{v}_2 = 0$, as otherwise γ_1 would be

decreased. Similarly, the strategy of user k solves the optimization problem

$$\max_{\boldsymbol{g}_k, \boldsymbol{v}_k} \frac{|\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k|^2}{\sum_{j<k} |\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kj} \boldsymbol{v}_j|^2 + \sigma^2} \quad \text{s.t.:} \quad \boldsymbol{g}_j^{\mathrm{T}} \boldsymbol{H}_{jk} \boldsymbol{v}_k = 0 \quad \text{for all active users} \quad j < k, \quad \|\boldsymbol{g}_k\|_2^2 = 1 \quad \text{and} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1 \tag{5.55}$$

where a user j is considered to be active if $\gamma_j > 0$. Since the zero-forcing constraints do not affect the receive filter $\boldsymbol{g}_k$, the optimal receive filter can be derived analogously to (5.13) and is

$$\boldsymbol{g}_k^* = \frac{1}{\left\| \left(\sum_{j<k} \boldsymbol{H}_{kj} \boldsymbol{v}_j \boldsymbol{v}_j^{\mathrm{H}} \boldsymbol{H}_{kj}^{\mathrm{H}} + \sigma^2 \mathbf{I} \right)^{-1} \boldsymbol{H}_{kk} \boldsymbol{v}_k \right\|_2} \left(\sum_{j<k} \boldsymbol{H}_{kj} \boldsymbol{v}_j \boldsymbol{v}_j^{\mathrm{H}} \boldsymbol{H}_{kj}^{\mathrm{H}} + \sigma^2 \mathbf{I} \right)^{-1} \boldsymbol{H}_{kk} \boldsymbol{v}_k. \tag{5.56}$$

Similarly, from inserting the expression for the optimal receive filter into the objective function it follows that the problem of choosing the beamformer is

$$\max_{\boldsymbol{v}_k} \boldsymbol{v}_k^{\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} \left(\sum_{j<k} \boldsymbol{H}_{kj} \boldsymbol{v}_j \boldsymbol{v}_j^{\mathrm{H}} \boldsymbol{H}_{kj}^{\mathrm{H}} + \sigma^2 \mathbf{I} \right)^{-1} \boldsymbol{H}_{kk} \boldsymbol{v}_k \quad \text{s.t.:} \quad \boldsymbol{g}_j^{\mathrm{T}} \boldsymbol{H}_{jk} \boldsymbol{v}_k = 0 \quad \text{for all active users} \quad j < k \quad \text{and} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1. \tag{5.57}$$

The solution to this optimization problem can be found with the method used for the derivation of the zero-forcing beamformer in the MISO case, cf. Appendix A5: we stack the row vectors $\boldsymbol{g}_j^{\mathrm{T}} \boldsymbol{H}_{jk}$ for all active users $j < k$ in the matrix $\boldsymbol{H}_k$ and define the projection matrix $\boldsymbol{\Pi}_k = \mathbf{I} - \boldsymbol{H}_k^{+} \boldsymbol{H}_k$. Next, we replace $\boldsymbol{v}_k$ with $\boldsymbol{\Pi}_k \boldsymbol{v}_k$ in the objective function of (5.57), which results in an equivalent optimization problem, as $\boldsymbol{v}_k = \boldsymbol{\Pi}_k \boldsymbol{v}_k$ when the zero-forcing constraints $\boldsymbol{H}_k \boldsymbol{v}_k = \mathbf{0}$ are fulfilled. Without the zero-forcing constraints, the solution to the optimization problem is straightforward: the optimal beamformer $\boldsymbol{v}_k$ is the principal eigenvector of the matrix

$$\boldsymbol{P}_k = \boldsymbol{\Pi}_k \boldsymbol{H}_{kk}^{\mathrm{H}} \left(\sum_{j<k} \boldsymbol{H}_{kj} \boldsymbol{v}_j \boldsymbol{v}_j^{\mathrm{H}} \boldsymbol{H}_{kj}^{\mathrm{H}} + \sigma^2 \mathbf{I} \right)^{-1} \boldsymbol{H}_{kk} \boldsymbol{\Pi}_k. \tag{5.58}$$

Since $\boldsymbol{H}_k \boldsymbol{v}_k = \lambda^{-1} \boldsymbol{H}_k \boldsymbol{P}_k \boldsymbol{v}_k = \mathbf{0}$, where λ is the highest eigenvalue of $\boldsymbol{P}_k$, this solution also fulfilles the zero-forcing constraints. Therefore, the principal eigenvector of $\boldsymbol{P}_k$ is also the maximizer of the optimization problem (5.57).

Pareto optimality of the successive zero-forcing strategies follows from the same argument as in Section 4.3.3:

- The SINR of user one cannot be improved.
- The SINR of user two can only be improved by violating the constraint of zero interference to user one or by changing the beamformer of user one, i. e., γ_2 can be increased only at the cost of decreasing γ_1.
- Similarly, the SINR of user k can be improved only by allowing interference to previously allocated users or by changing previously allocated users' beamformers, thus decreasing at least one SINR γ_j with $j < k$.
- Consequently, if $(\gamma_1, \ldots, \gamma_K)$ is the SINR K-tuple resulting from the successive zero-forcing strategy, there is no feasible SINR K-tuple $(\gamma_1', \ldots, \gamma_K')$ with $\gamma_k' \geq \gamma_k$ for all users k and $\gamma_j' > \gamma_j$ for at least one user j.

The successive zero-forcing strategies are on the Pareto boundary of the SINR region and the rate region regardless of the SNR. This does not mean, however, that they also achieve the sum-rate optimal slope at high SNR. In the following we analyze the parameters of the high-SNR asymptote; again, the slope s is the same for almost all channel realizations with a given antenna configuration, whereas characterizing the sum rate offset r requires a probabilistic analysis.

We begin by examining the high-SNR limit of the optimal receive filter vector. Similar to Section 5.2.1, we define $\boldsymbol{Y}_k$ to contain the vectors $\boldsymbol{H}_{kj}\boldsymbol{v}_j$ of the active users with $j < k$ as columns. With the matrix-inversion lemma, the receive filter can be expressed as

$$\boldsymbol{g}_k^* = \alpha \left(\boldsymbol{Y}_k \boldsymbol{Y}_k^{\mathrm{H}} + \sigma^2 \mathbf{I}\right)^{-1} \boldsymbol{H}_{kk}\boldsymbol{v}_k = \alpha \left(\sigma^{-2}\mathbf{I} - \sigma^{-2}\boldsymbol{Y}_k \left(\boldsymbol{Y}_k^{\mathrm{H}}\boldsymbol{Y}_k + \sigma^2 \mathbf{I}\right)^{-1} \boldsymbol{Y}_k^{\mathrm{H}}\right) \boldsymbol{H}_{kk}\boldsymbol{v}_k \tag{5.59}$$

where α ensures that $\|\boldsymbol{g}_k\|_2^2 = 1$. If $\boldsymbol{Y}_k\boldsymbol{Y}_k^{\mathrm{H}}$ is invertible, i. e., if there exists no vector $\boldsymbol{g}_k \neq \mathbf{0}$ with $\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{Y}_k = \mathbf{0}$, the high-SNR limit is $\boldsymbol{g}_k^* = \alpha \left(\boldsymbol{Y}_k\boldsymbol{Y}_k^{\mathrm{H}}\right)^{-1} \boldsymbol{H}_{kk}\boldsymbol{v}_k$; clearly, this receive filter does not completely remove interference and user k does not contribute to the asymptotic slope s.

If, on the other hand, zero-forcing is possible, i. e., there exists a unit-norm receive filter $\boldsymbol{g}_k$ fulfilling $\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{Y}_k = \mathbf{0}$, we can define the non-zero projection matrix $\boldsymbol{\Pi}_k' = \mathbf{I} - \boldsymbol{Y}_k\boldsymbol{Y}_k^{+}$. Furthermore, we note that

$$\lim_{\sigma^2 \to 0} \left(\boldsymbol{Y}_k^{\mathrm{H}}\boldsymbol{Y}_k + \sigma^2 \mathbf{I}\right)^{-1} \boldsymbol{Y}_k^{\mathrm{H}} = \boldsymbol{Y}_k^{+} \tag{5.60}$$

and with (5.59) the high-SNR optimal receive filter is

$$\lim_{\sigma^2 \to 0} \boldsymbol{g}_k^* = \frac{1}{\|\boldsymbol{\Pi}_k'\boldsymbol{H}_{kk}\boldsymbol{v}_k\|_2} \boldsymbol{\Pi}_k'\boldsymbol{H}_{kk}\boldsymbol{v}_k. \tag{5.61}$$

From $\boldsymbol{\Pi}_k'\boldsymbol{Y}_k = \mathbf{0}$ it follows that $\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{Y}_k = \mathbf{0}$, i. e., the high-SNR receive filter removes all interference. Therefore, user k contributes to the slope s, cf. (5.29).

Assuming that receiver k is able to cancel all interference, we insert the high-SNR optimal receive filter $\boldsymbol{g}_k$ into the SINR of user k:

$$\gamma_k = \frac{|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_{kk}\boldsymbol{v}_k|^2}{\sigma^2} = \frac{|\boldsymbol{v}_k^{\mathrm{H}}\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{\Pi}_k'\boldsymbol{H}_{kk}\boldsymbol{v}_k|^2}{\|\boldsymbol{\Pi}_k'\boldsymbol{H}_{kk}\boldsymbol{v}_k\|_2^2 \cdot \sigma^2} = \sigma^{-2}\boldsymbol{v}_k^{\mathrm{H}}\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{\Pi}_k'\boldsymbol{H}_{kk}\boldsymbol{v}_k. \tag{5.62}$$

Comparing this expression with the objective function of (5.57), it becomes evident that the optimal beamformer $\boldsymbol{v}_k$ is the principal eigenvector of the matrix

$$\boldsymbol{P}_k = \sigma^{-2}\boldsymbol{\Pi}_k\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{\Pi}_k'\boldsymbol{H}_{kk}\boldsymbol{\Pi}_k. \tag{5.63}$$

With the SVD of the projected direct channel matrix $\boldsymbol{\Pi}_k'\boldsymbol{H}_{kk}\boldsymbol{\Pi}_k = \boldsymbol{U}_k\boldsymbol{\Sigma}_k\boldsymbol{V}_k^{\mathrm{H}}$, we can therefore express the beamformer as $\boldsymbol{v}_k = \boldsymbol{V}_k\boldsymbol{e}_1$, and for the receive filter it follows that $\boldsymbol{g}_k^* = \boldsymbol{U}_k\boldsymbol{e}_1$. We conclude that the optimal zero-forcing strategy at high SNR can be found by applying the ME strategy to a projected version of the direct channel matrix $\boldsymbol{H}_{kk}$, where the projections from the left and the right ensure that the receiver removes all interference and the beamformer causes no interference, respectively.

In an i. i. d. Gaussian channel model and for any fixed set of previously allocated receive filters $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_{k-1}$, the rank of the matrix $\boldsymbol{H}_k$ with probability one is the number of active users $j < k$ or N_k, whichever is smaller. Therefore, the projection matrix $\boldsymbol{\Pi}_k$ is non-zero with probability one if the number of previously allocated active users is strictly less than N_k and zero with probability one otherwise. By the same argument, the projection matrix $\boldsymbol{\Pi}_k'$ is non-zero with probability one

if the number of non-zero beamformers $\boldsymbol{v}_j$ with $j < k$ is strictly less than M_k. And finally, if both $\boldsymbol{\Pi}_k$ and $\boldsymbol{\Pi}'_k$ are non-zero, clearly the matrix $\boldsymbol{\Pi}'_k \boldsymbol{H}_{kk} \boldsymbol{\Pi}_k$ is non-zero with probability one as well.

In a symmetric i. i. d. Gaussian scenario with N antennas at every transmitter and M antennas at every receiver, every transmitter can almost always fulfill $N-1$ zero-forcing contraints and every receiver can almost always fulfill $M-1$ zero-forcing constraints. Therefore, the first $\min\{N.M\}$ allocated users in the successive zero-forcing scheme are active without causing interference to previously allocated users and are capable of completely removing the interference received from previously allocated users. Consequently, the high-SNR slope of the successive zero-forcing strategy in a symmetric setting almost always is $s = \min\{N, M\}$. The optimal slope in a symmetric setting with sufficiently many users K, on the other hand, is $s = N+M-1$, cf. (5.40); unless either $N = 1$ or $M = 1$, the slope achieved by successive zero-forcing is therefore strictly suboptimal.

The successive zero-forcing procedure yields different strategies depending on the order in which the beamformers and receive filters are allocated. Up to $K!$ different strategies can be found by trying out all different permutations of user indices. If we are interested in finding a strategy with a high sum rate without trying out all $K!$ different permutations, we can take a "greedy" approach: in each allocation step the user with the highest resulting SINR is selected; in order to be able to determine this user, the update, i. e., the solution to the optimization problem (5.57), must be computed for every user that has not yet been allocated, and the resulting SINRs compared. In the high-SNR case, instead of comparing SINRs, we can equivalently project all remaining direct channel matrices from the left and right according to the zero-forcing constraints and select the user for which the resulting maximum singular value is highest.

5.4.2.1 Large-System Approximation of the Offset with Successive Zero-Forcing

While the slope s achieved by the successive zero-forcing scheme is the same for almost all channel realizations, the sum rate offset r is a random variable, which we examine in the following. We assume an i. i. d. Gaussian channel model with unit-variance channel coefficients and examine the case where all transmitters and receivers have N antennas. The analysis can also be applied to general antenna configurations, but the resulting expressions become considerably more involved. The following results were briefly discussed in [80].

Our analysis relies on a result from the theory of random matrices, which states that under certain conditions the spectrum of singular values of a random matrix becomes deterministic as the size of the matrix increases. By assuming asymptotically large system dimensions, we are able to give analytical expressions for the mean offset r; these expressions can then be used as an approximation for the offset for finite system dimensions.

Theorem 5.3. *Let $\boldsymbol{H}_{kk} \in \mathbb{C}^{N\times N}$ be a random matrix with i. i. d. complex Gaussian elements with mean zero and variance one and let ν denote the highest eigenvalue of $\boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{H}_{kk}$. Then*

$$\lim_{N\to\infty} \mathrm{E}\left[\frac{\nu}{N}\right] = 4 \qquad \textit{and} \qquad \lim_{N\to\infty} \mathrm{E}\left[\left(\frac{\nu}{N} - 4\right)^2\right] = 0. \tag{5.64}$$

Proof. It is shown in [83] and [84] that the distribution of the scaled and shifted maximum eigenvalue $\dfrac{\nu - \mu_N}{\sigma_N}$ with

$$\mu_N = 4N \qquad \text{and} \qquad \sigma_N = \sqrt[3]{16N} \tag{5.65}$$

converges towards the Tracy-Widom F_2 distribution as $N \to \infty$. The PDF of this distribution is defined by means of the Painlevé II differential equation and cannot be expressed in closed form;

it was, however, numerically determined in [85] that the mean and variance are $\mu_0 \approx -1.7711$ and $\sigma_0^2 \approx 0.8132$, respectively. Therefore,

$$\begin{aligned}\lim_{N\to\infty} \mathrm{E}\left[\frac{\nu}{N}\right] &= \lim_{N\to\infty} \mathrm{E}\left[\left(\frac{\nu-\mu_N}{\sigma_N}\sigma_N + \mu_N\right)/N\right] = \lim_{N\to\infty} \frac{\mu_0\sigma_N + \mu_N}{N} \\ &= 4 + \lim_{N\to\infty} \frac{\mu_0\sqrt[3]{16}}{N^{2/3}} = 4.\end{aligned} \tag{5.66}$$

Similarly,

$$\begin{aligned}\lim_{N\to\infty} \mathrm{E}\left[\left(\frac{\nu}{N} - 4\right)^2\right] &= \lim_{N\to\infty} \mathrm{E}\left[\left(\frac{\sigma_N}{N}\cdot\frac{\nu-\mu_N}{\sigma_N}\right)^2\right] \\ &= \lim_{N\to\infty} \frac{\sigma_N^2}{N^2}\,\mathrm{E}\left[\left(\frac{\nu-\mu_N}{\sigma_N} - \mu_0 + \mu_0\right)^2\right] \\ &= \lim_{N\to\infty} \frac{\sigma_N^2}{N^2}\left(\mathrm{E}\left[\left(\frac{\nu-\mu_N}{\sigma_N} - \mu_0\right)^2\right] + 2\mu_0\,\mathrm{E}\left[\frac{\nu-\mu_N}{\sigma_N} - \mu_0\right] + \mu_0^2\right) \\ &= \lim_{N\to\infty} \frac{\sigma_N^2}{N^2}\left(\sigma_0^2 + \mu_0^2\right) = \lim_{N\to\infty} \frac{\sqrt[3]{256}}{N^{4/3}}\left(\sigma_0^2 + \mu_0^2\right) = 0.\end{aligned} \tag{5.67}$$

□

As a consequence of Theorem 5.3, we can approximate the highest singular value of $\boldsymbol{H}_{kk}/\sqrt{N}$ as 2 if N is large. This result is confirmed by the Marchenko-Pastur law [86], which states that the *empirical distribution* of the eigenvalues of $\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{H}_{kk}/N$ approaches a known limiting distribution, i. e., for large N the singular value spectrum of $\boldsymbol{H}_{kk}/\sqrt{N}$ "becomes deterministic."

Lemma 5.1. *Let $\boldsymbol{H}_{kk} \in \mathbb{C}^{N\times N}$ be a random matrix with i. i. d. complex Gaussian elements with mean zero and variance one and let $\sqrt{\nu_k}$ denote the highest singular value of the matrix $\boldsymbol{\Pi}_k'\boldsymbol{H}_{kk}\boldsymbol{\Pi}_k$, where the projection matrices $\boldsymbol{\Pi}_k'$ and $\boldsymbol{\Pi}_k$ are statistically independent of $\boldsymbol{H}_{kk}$ and have rank $N-k+1$. Then*

$$\lim_{N\to\infty} \mathrm{E}\left[\frac{\nu_k}{N-k+1}\right] = 4 \qquad \textit{and} \qquad \lim_{N\to\infty} \mathrm{E}\left[\left(\frac{\nu_k}{N-k+1} - 4\right)^2\right] = 0. \tag{5.68}$$

The result holds for fixed k, but also if k is a function of N, provided that $N-k \to \infty$ as $N \to \infty$.

Proof. We can express the projection matrices as $\boldsymbol{\Pi}_k' = \boldsymbol{W}_k'\boldsymbol{W}_k'^{\mathrm{H}}$ and $\boldsymbol{\Pi}_k = \boldsymbol{W}_k\boldsymbol{W}_k^{\mathrm{H}}$ where $\boldsymbol{W}_k' \in \mathbb{C}^{N\times N-k+1}$ and $\boldsymbol{W}_k \in \mathbb{C}^{N\times N-k+1}$ are matrices with orthogonal unit-norm columns, i. e., $\boldsymbol{W}_k'^{\mathrm{H}}\boldsymbol{W}_k' = \boldsymbol{W}_k^{\mathrm{H}}\boldsymbol{W}_k = \mathbf{I}$. Clearly, the maximum singular value of $\boldsymbol{W}_k'^{\mathrm{H}}\boldsymbol{H}_{kk}\boldsymbol{W}_k$ is also $\sqrt{\nu_k}$. It is furthermore straightforward to show that $\boldsymbol{W}_k'^{\mathrm{H}}\boldsymbol{H}_{kk}\boldsymbol{W}_k \in \mathbb{C}^{N-k+1\times N-k+1}$ has zero-mean unit-variance i. i. d. complex Gaussian elements. Therefore, we can apply Theorem 5.3 to $\boldsymbol{W}_k'^{\mathrm{H}}\boldsymbol{H}_{kk}\boldsymbol{W}_k$ and the proof is complete. □

With these properties of the highest singular value of the projected direct channel matrices, we are ready to analyze the sum rate offset in a large system. We recall that in the symmetric setting with $M = N$ antennas at every node a slope of $s = N$ is achievable, so that we assume that our

system has $K = N$ users. Since at high SNR the optimal receive filter and beamformer of user k select the maximum singular value of the projected direct channel $\boldsymbol{\Pi}_k' \boldsymbol{H}_{kk} \boldsymbol{\Pi}_k$, the sum rate offset is

$$r = \sum_{k=1}^{N} \log|\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k|^2 = \sum_{k=1}^{N} \log \nu_k \tag{5.69}$$

where ν_k is defined as in Lemma 5.1. Note that since the projection matrices $\boldsymbol{\Pi}_k'$ and $\boldsymbol{\Pi}_k$ are functions of the channel matrices $\boldsymbol{H}_{jk}$ and $\boldsymbol{H}_{kj}$ for all $j < k$, they are indeed statistically independent of $\boldsymbol{H}_{kk}$.

Theorem 5.4. *In the large system limit, i. e., as $N \to \infty$, the mean sum rate offset for $K = N$ users is*

$$\mathrm{E}[r] = 2N \log 2 + \log N! + \mathrm{o}(N). \tag{5.70}$$

Proof. We define $Q = \lfloor \log N \rfloor$ and split the N users into two groups; the first group consists of $N - Q$ users with a total rate offset of q_1, the second group consists of Q users with a total rate offset of q_2:

$$r = \sum_{k=1}^{N-Q} \log \nu_k + \sum_{k=N-Q+1}^{N} \log \nu_k = q_1 + q_2. \tag{5.71}$$

Note that $\lim_{N\to\infty} Q = \infty$, but that $\lim_{N\to\infty} (N - Q)/N = 1$, i. e., the number of users in the second group grows without bound with the system dimension, but in the limit almost all users are in the first group.

We begin by examining $\mathrm{E}[q_1]$; since Q grows without bound with N, we can apply the large system limit to every user in the first group. From Lemma 5.1 it follows that in the limit $\nu_k/(N - k + 1)$ is deterministic. Therefore, $\log \nu_k - \log(N - k + 1)$ approaches $\log 4$ and we can state that

$$\mathrm{E}[\log \nu_k] = 2 \log 2 + \log(N - k + 1) + \mathrm{o}(1) \tag{5.72}$$

and

$$\mathrm{E}[q_1] = \sum_{k=1}^{N-Q} \mathrm{E}[\log \nu_k] = 2(N - Q) \log 2 + \log N! - \log Q! + \mathrm{o}(N). \tag{5.73}$$

It remains to be shown that $2Q \log 2 \in \mathrm{o}(N)$ and $\log Q! \in \mathrm{o}(N)$. While the former follows directly from the definition of Q, for the latter we apply the bound

$$\log Q! \leq Q \log Q \leq \log N \cdot \log \log N \tag{5.74}$$

from which it follows that

$$\lim_{N\to\infty} \frac{\log Q!}{N} = 0. \tag{5.75}$$

Therefore,

$$\mathrm{E}[q_1] = \sum_{k=1}^{N-Q} \mathrm{E}[\log \nu_k] = 2N \log 2 + \log N! + \mathrm{o}(N). \tag{5.76}$$

For the remaining Q users in the second group, we make use of the intuitive fact that the mean individual rate offset $\mathrm{E}[\log \nu_k]$ is non-increasing in the index k, as adding more constraints cannot increase the mean achievable rate. Therefore, we can upper-bound $\mathrm{E}[q_2]$ by assuming that all Q users perform as well as the first user in the second group, i. e., user $k = N - Q + 1$. Since

N	$\mathrm{E}[r]$ (numerical)	$\bar{r}_{\mathrm{MSA}}$	$\bar{r}_{\mathrm{LSA}}$
2	0.54	0.87	3.47
5	7.13	7.29	11.7
10	22.8	22.9	29.0
20	61.9	61.9	70.0
100	488	488	502

Table 5.3: Approximations of the average sum rate offset of the successive zero-forcing strategy compared to the numerically measured average offset.

$N-k=Q-1\to\infty$ as $N\to\infty$, this user still falls under the large system limit and we can state that

$$\mathrm{E}[q_2]\leq Q\,\mathrm{E}\left[\nu_{N-Q+1}\right]=Q(2\log 2+\log Q+\mathrm{o}(1)) \tag{5.77}$$

and

$$\lim_{N\to\infty}\frac{\mathrm{E}[q_2]}{N}=0. \tag{5.78}$$

With

$$\mathrm{E}[r]=\mathrm{E}[q_1]+\mathrm{E}[q_2] \tag{5.79}$$

the proof is complete. □

With the result of Theorem 5.4, we formulate our large-system approximation by ignoring the terms that grow slower than N:

$$\bar{r}_{\mathrm{LSA}}=2N\log 2+\log N!. \tag{5.80}$$

In addition, we propose a more accurate approximation for moderate system sizes that takes into account the shift parameter μ_0 that appears in the proof of Theorem 5.3. According to the second line of (5.66), for an i. i. d. complex Gaussian matrix of size $N-k+1\times N-k+1$, we approximate the highest squared singular value as

$$\bar{\nu}_k=\left(4+\frac{\mu_0\sqrt[3]{16}}{(N-k+1)^{2/3}}\right)(N-k+1) \tag{5.81}$$

for $k\in\{1,\ldots,N-1\}$. We omit the term $\bar{\nu}_N$, as above approximation is negative for $N-k+1=1$. Our "moderate system-size approximation" therefore is

$$\bar{r}_{\mathrm{MSA}}=\sum_{k=1}^{N-1}\log\left(4+\frac{\mu_0\sqrt[3]{16}}{(N-k+1)^{2/3}}\right)+\log N! \tag{5.82}$$

where $\mu_0\approx -1.7711$. Table 5.3 shows the result of the two approximations for different system dimensions N along with a numerical estimate of the actual value of $\mathrm{E}[r]$; the estimate was obtained by generating random direct channels and random projection matrices and taking the sample mean over 10 000 realizations. For the quality of the approximation at finite SNR we refer to the numerical results in Section 5.6.2.

The offset approximations for the successive zero-forcing strategy and the offset approximation of an aligned solution show fundamentally different behavior: while the former is always positive

and increases with the system size, $\bar{r}_{\mathrm{approx}}$, cf. (5.54), is negative and decreases with the system size (unless L grows extremely rapidly with N, in which case the validity of the approximation is questionable). On the other hand, the former procedure yields a slope of $s = N$, while the latter strategy achieves $s = 2N - 1$. This tradeoff between offset and slope will become more apparent in Section 5.6.2, where we also examine the behavior of the crossing point of the two asymptotes.

5.5 Distributed Algorithms

The discussed analytical results concerning the performance of different strategies in MIMO interference networks are valid in the high-SNR limit; at finite SNR, we are largely dependent on numerical methods to evaluate the potential of the system. The goal of the algorithms discussed in the following is to determine close-to-optimal solutions regardless of the noise power; in particular, this means that at high SNR a good algorithm should converge to an aligned solution. As in the previous chapters, we also require that the algorithms are suitable for distributed operation, i. e., that the individual users perform their share of the computations locally and exchange only a limited amount of information. We focus on the maximization of the sum rate utility.

The following algorithms for single-stream MIMO interference networks for the most part build on the algorithms known from the MISO case: for a given transmit strategy, i. e., a K-tuple of beamformers, the optimal receive filters are straightforward to compute. For a fixed K-tuple of receive filters, on the other hand, we can effectively view the system as a MISO interference network with the combination of the receive filters and the channel matrices as equivalent channels. Consequently, the basic approach of the algorithms for single-stream MIMO systems is as follows: we alternate between updating the transmit strategies by applying MISO updates on the equivalent MISO channels using, e. g., the techniques discussed in Section 4.4, and updating the receive filters, e. g., as discussed in Section 5.2. Clearly, this procedure implies additional information exchange requirements: in order to compute the equivalent MISO channels, the transmitters must have knowledge of all users' current receive filters, which must therefore be announced after each update.

Again, the algorithms can be operated in a sequential or a parallel mode: when updating sequentially, the users take turns updating first their beamformer and then their receive filter; after user k has updated both its beamformer and its receive filter, it announces the new receive filter to the other users. At this point all users also measure their SINR and, if necessary, compute and exchange their interference prices, etc., which have changed due to the beamformer update of user k; the next user $k + 1$ can then perform its beamformer and receive filter update with up-to-date information concerning the other receivers' current conditions. When considering parallel updates, on the other hand, we assume that all K beamformers are updated simultaneously, and then all K receive filters are updated simultaneously, followed by the exchange of the new receive filters and other necessary information among the users.

5.5.1 Selfish Updates

As already noted in Section 5.4.1, a selfish solution in general requires iterative updates, as the selfish behavior in the MIMO case depends on the other users' strategies. In order to selfishly update its strategy, user k must adjust its beamformer $\boldsymbol{v}_k$ and receive filter $\boldsymbol{g}_k$ to maximize γ_k. The receiver update therefore consists of computing the SINR-optimal receive filter from (5.13) using

the current beamformers:

$$\boldsymbol{g}_k^{(\ell+1)} = \boldsymbol{g}_k^{\mathrm{opt}}\Big|_{\boldsymbol{v}_j=\boldsymbol{v}_j^{(\ell)}\ \forall j} = \frac{\left(\left(\sum_{j\neq k}\boldsymbol{H}_{kj}\boldsymbol{v}_j^{(\ell)}\boldsymbol{v}_j^{(\ell)\mathrm{H}}\boldsymbol{H}_{kj}^{\mathrm{H}}+\sigma^2\mathbf{I}\right)^{-1}\boldsymbol{H}_{kk}\boldsymbol{v}_k^{(\ell)}\right)^*}{\left\|\left(\sum_{j\neq k}\boldsymbol{H}_{kj}\boldsymbol{v}_j^{(\ell)}\boldsymbol{v}_j^{(\ell)\mathrm{H}}\boldsymbol{H}_{kj}^{\mathrm{H}}+\sigma^2\mathbf{I}\right)^{-1}\boldsymbol{H}_{kk}\boldsymbol{v}_k^{(\ell)}\right\|_2}. \tag{5.83}$$

For a given receive filter $\boldsymbol{g}_k^{(\ell)}$, the direct channel can be equivalently seen as a MISO channel with the coefficient vector $\boldsymbol{g}_k^{(\ell)\mathrm{T}}\boldsymbol{H}_{kk}$. Thus, with (4.18) the SINR-optimal transmitter update is

$$\boldsymbol{v}_k^{(\ell+1)} = \frac{\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{g}_k^{(\ell)*}}{\left\|\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{g}_k^{(\ell)*}\right\|_2}. \tag{5.84}$$

The beamformer update only requires knowledge of the direct channel matrix and the current receive filter of the desired receiver; the former is estimated initially by the receiver and then fed back to the transmitter, the latter is fed back after each update. The receive filter update requires knowledge of the combination of beamformer and direct channel as well as the covariance matrix of interference plus noise; we assume that both can be estimated locally without further information exchange among the users.

When the iterative selfish updates have converged, i. e., when the beamformers and receive filters do not change when updated as above, the so-found stationary strategy is a selfish solution and a Nash equlibrium: no user k can improve its own SINR γ_k by unilaterally adjusting its strategy consisting of $\boldsymbol{g}_k$ and $\boldsymbol{v}_k$. Convergence of the selfish updates cannot be guaranteed, however; oscillations can be observed in many channel realizations in all of our numerically evaluated channel models, cf. Section 5.6. Also, it is clear that a selfish solution does not achieve close-to-optimal performance at high SNR: in order to reach an aligned solution, the caused interference must be taken into account when updating the beamformers.

Without the restriction of single-stream transmission, the procedure of selfishly maximizing the own rate is known as "iterative waterfilling" and has been the object of study for a decade [87, 88, 89, 90, 91]. Different measures have been proposed to improve performance by artificially reducing the number of streams [88, 90] and preventing oscillations [90]; a sufficient condition for convergence was derived in [91]. These results are, however, not directly applicable to the single-stream case.

5.5.2 Sum Interference Power Minimization: The "Min-Leakage" Algorithm

The first algorithm designed specifically to find aligned solutions was proposed in [50]: the objective of receive filter and beamformer design is to minimize the sum interference power

$$I_{\mathrm{sum}} = \sum_k \sum_{j\neq k} |\boldsymbol{g}_j^{\mathrm{T}}\boldsymbol{H}_{jk}\boldsymbol{v}_k|^2. \tag{5.85}$$

It is clearly necessary to require that all beamformers have unit power and that the receive filters have a fixed norm, as otherwise I_{sum} would be minimized by switching off all users.

The receive filter update therefore is

$$\boldsymbol{g}_k^{(\ell+1)} = \underset{\mathrm{g}_k}{\arg\min}\ I_{\mathrm{sum}}\Big|_{\boldsymbol{v}_j=\boldsymbol{v}_j^{(\ell)}\ \forall j} \qquad \text{s. t.:} \quad \|\boldsymbol{g}_k\|^2 = 1. \tag{5.86}$$

The part of the cost function that depends on $\boldsymbol{g}_k$ is $\sum_{j\neq k}|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_{kj}\boldsymbol{v}_j|^2$; consequently, the KKT condition necessary for optimality is

$$\frac{\partial I_{\mathrm{sum}}}{\partial \boldsymbol{g}_k} = \sum_{j\neq k} \boldsymbol{H}_{kj}\boldsymbol{v}_j\boldsymbol{v}_j^{\mathrm{H}}\boldsymbol{H}_{kj}^{\mathrm{H}}\boldsymbol{g}_k^* = \lambda \boldsymbol{g}_k^* \tag{5.87}$$

with the Lagrangian multiplier $\lambda \in \mathbb{R}$, i. e., the optimal (complex conjugate) receive filter $\boldsymbol{g}_k^*$ is an eigenvector of the matrix $\sum_{j\neq k} \boldsymbol{H}_{kj}\boldsymbol{v}_j\boldsymbol{v}_j^{\mathrm{H}}\boldsymbol{H}_{kj}^{\mathrm{H}}$. By multiplying from the left with $\boldsymbol{g}_k^{\mathrm{T}}$ it also follows that $\sum_{j\neq k}|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_{kj}\boldsymbol{v}_j|^2 = \lambda$; therefore, the smallest eigenvalue leads to the smallest sum interference power I_{sum}. Consequently, the updated receive filter $\boldsymbol{g}_k^{(\ell+1)^*}$ is the eigenvector corresponding to the smallest eigenvalue of the matrix $\sum_{j\neq k} \boldsymbol{H}_{kj}\boldsymbol{v}_j^{(\ell)}\boldsymbol{v}_j^{(\ell)^{\mathrm{H}}}\boldsymbol{H}_{kj}^{\mathrm{H}}$. If the smallest eigenvalue is not unique, e. g., because the matrix has a multi-dimensional null space, any eigenvector can be chosen. In the same way, the updated beamformer $\boldsymbol{v}_k^{(\ell+1)}$ can be shown to be any eigenvector corresponding to the smallest eigenvalue of the matrix $\sum_{j\neq k} \boldsymbol{H}_{jk}^{\mathrm{H}}\boldsymbol{g}_j^{(\ell)^*}\boldsymbol{g}_j^{(\ell)^{\mathrm{T}}}\boldsymbol{H}_{jk}$.

With parallel or sequential updates, the value of I_{sum} cannot be increased by an update; therefore, the sum interference power must converge. It was furthermore shown experimentally in [50] that the min-leakage algorithm reliably finds aligned solutions if the system is proper.

We note that neither the noise power σ^2 nor the direct channel matrices $\boldsymbol{H}_{kk}$ appear in the min-leakage updates. When the initial strategy is independent of the direct channels, the strategy determined by the min-leakage algorithm is therefore also independent of the direct channels. We can thus use the min-leakage algorithm to determine a random aligned solution, of which the performance was analyzed in Section 5.3.5. Furthermore, by using L random independent isotropically distributed[6] initialization strategies, where L is significantly lower than the total number of solutions, we can assume that the min-leakage algorithm returns L different (independent) aligned solutions. In Sections 5.6.1 and 5.6.2, we use this method to numerically evaluate the approximation discussed in Section 5.3.5.1.

In order to perform the update, receiver k must have knowledge of the current covariance matrix of received interference; we assume that this can be estimated. Transmitter k must have knowledge of all cross channels $\boldsymbol{H}_{jk}$ with $j \neq k$ as well as the current receive filters $\boldsymbol{g}_j^{(\ell)}$ with $j \neq k$. The cross channels can be exchanged once among the users, the receive filters, however, must be announced and exchanged over the signaling links after each update.

The min-leakage algorithm was originally proposed for a time-division duplex setting, in which receivers and transmitters periodically switch roles: assuming reciprocal channel conditions, in the "reverse direction" the channel matrix from transmitter k (normally receiver k) to receiver j (normally transmitter j) is $\boldsymbol{H}_{kj}^{\mathrm{T}}$; furthermore, transmitter k in the reverse direction uses $\boldsymbol{g}_k$ as its beamformer. Therefore, the update of $\boldsymbol{v}_k$ can be viewed as a receiver update in the reverse direction and can be performed using the estimated received interference covariance matrix. No signaling links are necessary in such a setting.

We finally note that in a MISO scenario in which zero-forcing is possible the min-leakage algorithm converges to a strategy with zero interference power in a single iteration.

[6] We refer to a random strategy as *isotropically distributed* when all elements of the beamformers and receive filters are drawn independently from an i. i. d. complex Gaussian distribution and the beamformers and receive filters are then normalized to have unit norm.

5.5.3 Virtual SINR Updates: The "Max-SINR" Algorithm

A second algorithm, which additionally takes into account the direct channel matrices and the noise power, was also proposed in [50]. It can be viewed as an extension of the virtual SINR maximization method discussed in Section 4.4.2 to MIMO systems.

The objective of the receive filter update of user k is to maximize the own SINR γ_k, cf. (5.13), which leads to the same receive filter update as in the selfish algorithm in Section 5.5.1. The objective of the beamformer update is to maximize the virtual SINR ξ_k, defined as in (4.50). Since we can view the channel from transmitter k to receiver j in combination with the receive filter of user j as a MISO channel with the coefficient vector $\boldsymbol{g}_j^{(\ell)\mathrm{T}} \boldsymbol{H}_{jk}$, the transmitter update solves the optimization problem

$$\boldsymbol{v}_k^{(\ell+1)} = \underset{\boldsymbol{v}_k}{\arg\max} \; \frac{\left|\boldsymbol{g}_k^{(\ell)\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k\right|^2}{\sum_{j \neq k} \left|\boldsymbol{g}_j^{(\ell)\mathrm{T}} \boldsymbol{H}_{jk} \boldsymbol{v}_k\right|^2 + \sigma^2} \qquad \text{s. t.:} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1. \tag{5.88}$$

In the same way as in Section 4.4.2 we obtain

$$\boldsymbol{v}_k^{(\ell+1)} = \frac{\left(\sum_{j \neq k} \boldsymbol{H}_{jk}^{\mathrm{H}} \boldsymbol{g}_j^{(\ell)*} \boldsymbol{g}_j^{(\ell)\mathrm{T}} \boldsymbol{H}_{jk} + \sigma^2 \mathbf{I}\right)^{-1} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^{(\ell)*}}{\left\| \left(\sum_{j \neq k} \boldsymbol{H}_{jk}^{\mathrm{H}} \boldsymbol{g}_j^{(\ell)*} \boldsymbol{g}_j^{(\ell)\mathrm{T}} \boldsymbol{H}_{jk} + \sigma^2 \mathbf{I}\right)^{-1} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^{(\ell)*} \right\|_2}. \tag{5.89}$$

The beamformers in the max-SINR algorithm therefore always have full power.

Again, the max-SINR algorithm was originally proposed for a time-division duplex setting. In the reverse direction, the beamformer update is simply a receive filter update that maximizes the received SINR and all quantities can be estimated without the need for signaling links. In our signaling framework, the transmitters must exchange the cross channel matrices once and the current receive filters after each update.

General convergence of the max-SINR algorithm has not been proven, but it was shown experimentally in [50] that convergence is very reliable in an i. i. d. Gaussian channel model. In [92], it was later shown that if the algorithm is initialized sufficiently close to a stationary strategy and if the SNR is sufficiently high, it converges exponentially. Furthermore, it was rigorously proven in [92] that in a proper system the globally sum-rate optimal strategy is a stationary strategy of the max-SINR algorithm at asymptotically high SNR, as are all other aligned solutions.

5.5.4 Global SINR Updates

The use of the global SINR objective was proposed in [68]; in Section 4.4.3 we applied this algorithm to the MISO case and therefore only briefly discuss the general MIMO case here.

The objective of both the transmitter and the receiver updates is to maximize the global SINR ξ_{sum} with equality power/norm constraints. For the MISO case the global SINR is defined in (4.60); the extension to the MIMO case is

$$\xi_{\mathrm{sum}} = \frac{\sum_k |\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k|^2}{\sum_k \sum_{j \neq k} |\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kj} \boldsymbol{v}_j|^2 + K\sigma^2}. \tag{5.90}$$

The receive filter is therefore updated according to the optimization problem

$$\boldsymbol{g}_k^{(\ell+1)} = \underset{\boldsymbol{g}_k}{\arg\max} \; \xi_{\mathrm{sum}} \Big|_{\substack{\boldsymbol{v}_j = \boldsymbol{v}_j^{(\ell)} \; \forall j \\ \boldsymbol{g}_j = \boldsymbol{g}_j^{(\ell)} \; \forall j \neq k}} \qquad \text{s. t.:} \quad \|\boldsymbol{g}_k\|_2^2 = 1 \tag{5.91}$$

and the beamformer update is

$$\boldsymbol{v}_k^{(\ell+1)} = \underset{\boldsymbol{v}_k}{\arg\max}\ \xi_{\mathrm{sum}}\Big|_{\substack{\boldsymbol{g}_j=\boldsymbol{g}_j^{(\ell)}\ \forall j \\ \boldsymbol{v}_j=\boldsymbol{v}_j^{(\ell)}\ \forall j\neq k}} \qquad \text{s. t.:} \quad \|\boldsymbol{v}_k\|_2^2 = 1. \tag{5.92}$$

In order to compactly express the solutions to the above optimization problems, we define the following abbreviations:

$$a_k^{(\ell)} = \sum_{j\neq k}\left|\boldsymbol{g}_j^{(\ell)^{\mathrm{T}}}\boldsymbol{H}_{jj}\boldsymbol{v}_j^{(\ell)}\right|^2 \tag{5.93}$$

$$b_k^{(\ell)} = \sum_{j\neq k}\sum_{i\neq j}\left|\boldsymbol{g}_i^{(\ell)^{\mathrm{T}}}\boldsymbol{H}_{ij}\boldsymbol{v}_j^{(\ell)}\right|^2 + K\sigma^2 \tag{5.94}$$

$$c_k^{(\ell)} = \sum_{j\neq k}\sum_{i\neq j}\left|\boldsymbol{g}_j^{(\ell)^{\mathrm{T}}}\boldsymbol{H}_{ji}\boldsymbol{v}_i^{(\ell)}\right|^2 + K\sigma^2. \tag{5.95}$$

Note that $a_k^{(\ell)}$ is the part of the numerator of ξ_{sum} that does not depend on $\boldsymbol{v}_k$ and $\boldsymbol{g}_k$, while $b_k^{(\ell)}$ consists of the terms in the denominator that do not depend on $\boldsymbol{v}_k$ and $c_k^{(\ell)}$ consists of the terms in the denominator that do not depend on $\boldsymbol{g}_k$. The optimization problems are solved as in Section 4.4.3, by noting that the objectives are generalized Rayleigh quotients. The resulting updated (complex conjugate) receive filter $\boldsymbol{g}_k^{(\ell+1)^*}$ is the principal eigenvector of the matrix

$$\left(\sum_{j\neq k}\boldsymbol{H}_{kj}\boldsymbol{v}_j^{(\ell)}\boldsymbol{v}_j^{(\ell)^{\mathrm{H}}}\boldsymbol{H}_{kj}^{\mathrm{H}} + c_k^{(\ell)}\mathbf{I}\right)^{-1}\left(\boldsymbol{H}_{kk}\boldsymbol{v}_k^{(\ell)}\boldsymbol{v}_k^{(\ell)^{\mathrm{H}}}\boldsymbol{H}_{kk}^{\mathrm{H}} + a_k^{(\ell)}\mathbf{I}\right) \tag{5.96}$$

and the updated beamformer $\boldsymbol{v}_k^{(\ell+1)}$ is the principal eigenvector of the matrix

$$\left(\sum_{j\neq k}\boldsymbol{H}_{jk}^{\mathrm{H}}\boldsymbol{g}_j^{(\ell)^*}\boldsymbol{g}_j^{(\ell)^{\mathrm{T}}}\boldsymbol{H}_{jk} + b_k^{(\ell)}\mathbf{I}\right)^{-1}\left(\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{g}_k^{(\ell)^*}\boldsymbol{g}_k^{(\ell)^{\mathrm{T}}}\boldsymbol{H}_{kk} + a_k^{(\ell)}\mathbf{I}\right). \tag{5.97}$$

For sequential updates, it is clear that ξ_{sum} cannot be decreased by updating either a beamformer or a receive filter, i. e., convergence of the global SINR utility is guaranteed. For parallel updates, convergence is reliable as well, as we show in our numerical evaluation in Section 5.6.3.

As with the previously discussed algorithms, transmitter k must have knowledge of the channel matrices $\boldsymbol{H}_{jk}$ and the current receive filters $\boldsymbol{g}_j^{(\ell)}$ for all $j \in \{1,\ldots,K\}$. Additionally, $a_k^{(\ell)}$ and $b_k^{(\ell)}$ must be known; to this end, all users j must announce the numerators and denominators of their respective SINRs $\gamma_j^{(\ell)}$ after each update.

On the receiver side, we again assume that $\sum_{j\neq k}\boldsymbol{H}_{kj}\boldsymbol{v}_j^{(\ell)}\boldsymbol{v}_j^{(\ell)^{\mathrm{H}}}\boldsymbol{H}_{kj}^{\mathrm{H}}$ and $\boldsymbol{H}_{kk}\boldsymbol{v}_k^{(\ell)}\boldsymbol{v}_k^{(\ell)^{\mathrm{H}}}\boldsymbol{H}_{kk}^{\mathrm{H}}$ can be estimated. The regularization factors $a_k^{(\ell)}$ and $c_k^{(\ell)}$ cannot be estimated, however, and require additional signaling. The global SINR maximization algorithm is the only of our discussed schemes in which the receiver update requires information that cannot be estimated.

5.5.5 MMSE Updates

The use of the sum MSE objective for multi-antenna interference networks, which we explored for the MISO case in Section 4.4.4, was proposed independently in [69, 70, 68]. As the generalization from MISO to MIMO systems is fairly straightforward, we again only briefly state the updates and refer to Section 4.4.4 for the details.

The MSE ε_k of user k is defined in (5.16). As ε_j with $j \neq k$ does not depend on $\boldsymbol{g}_k$, the goal of the receiver update is to minimize ε_k. With (5.18), the updated receive filter therefore is

$$\boldsymbol{g}_k^{(\ell+1)^*} = \left(\sum_j \boldsymbol{H}_{kj} \boldsymbol{v}_j^{(\ell)} \boldsymbol{v}_j^{(\ell)^{\mathrm{H}}} \boldsymbol{H}_{kj}^{\mathrm{H}} + \sigma^2 \mathbf{I} \right)^{-1} \boldsymbol{H}_{kk} \boldsymbol{v}_k^{(\ell)}. \tag{5.98}$$

Note that, contrary to the previous algorithms, the MSE-optimal receive filter does not have unit norm.

The sum-MSE optimal beamformer update solves the problem

$$\boldsymbol{v}_k^{(\ell+1)} = \arg\min_{\boldsymbol{v}_k} \sum_j \varepsilon_j \qquad \text{s.t.:} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1. \tag{5.99}$$

For the solution we again make use of the notion of equivalent MISO channels and simply replace $g_j^{(\ell)} \boldsymbol{h}_{jk}^{\mathrm{T}}$ in Section 4.4.4 with $\boldsymbol{g}_j^{(\ell)^{\mathrm{T}}} \boldsymbol{H}_{jk}$ for all $j \in \{1, \ldots, K\}$ to obtain the beamformer update

$$\boldsymbol{v}_k^{(\ell+1)} = \left(\sum_j \boldsymbol{H}_{jk}^{\mathrm{H}} \boldsymbol{g}_j^{(\ell)^*} \boldsymbol{g}_j^{(\ell)^{\mathrm{T}}} \boldsymbol{H}_{jk} + \lambda \mathbf{I} \right)^{+} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^{(\ell)^*} \tag{5.100}$$

where the Lagrangian multiplier $\lambda \geq 0$ ensures that the power constraint is fulfilled. As for the MISO case, if $\lambda = 0$ does not fulfill the power constraint, we must determine λ with a line search, e. g., using Newton's method. If $\lambda > 0$, the pseudo-inverse in (5.100) can be replaced by the inverse.

For both sequential and parallel updates, the MSE cannot be increased by an update of a beamformer or a receive filter and therefore converges. Also, a stationary strategy of the MMSE-algorithm clearly fulfills the KKT conditions of the sum MSE minimization problem.

We can also use this procedure to minimize a weighted sum MSE $\sum_k \alpha_k \varepsilon_k$, where α_k is the weight associated with user k: the receiver update is identical to the unweighted case and the beamformer update is

$$\boldsymbol{v}_k^{(\ell+1)} = \left(\sum_j \frac{\alpha_j}{\alpha_k} \boldsymbol{H}_{jk}^{\mathrm{H}} \boldsymbol{g}_j^{(\ell)^*} \boldsymbol{g}_j^{(\ell)^{\mathrm{T}}} \boldsymbol{H}_{jk} + \lambda \mathbf{I} \right)^{+} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^{(\ell)^*}. \tag{5.101}$$

By adaptively choosing the weights as

$$\alpha_k = \frac{1}{\varepsilon_k} \bigg|_{\substack{\boldsymbol{v}_j = \boldsymbol{v}_j^{(\ell)} \ \forall j \\ \boldsymbol{g}_j = \boldsymbol{g}_j^{(\ell)} \ \forall j}} \tag{5.102}$$

we can locally mimic the behavior of the sum rate utility. Furthermore, since the rate $R_k = -\log \varepsilon_k$ is convex in the MSE, the weighted-MSE approximation under-estimates the sum rate, so that from

one weight update to the next the sum rate is non-decreasing and must converge. Finally, a strategy that is stationary in beamformers, receive filters, and weights fulfills the KKT conditions of the sum rate maximization problem.

The receive filter updates require knowledge of the covariance matrix of the received signal and the combination of direct channel and desired beamformer; we assume that both can be estimated. For the beamformer update of user k, the channel matrices $\boldsymbol{H}_{jk}$ must be known as well as the current receive filters of all users j and the weights α_j, which must be communicated over the signaling links after each update.

5.5.6 Interference Pricing

The objective of the interference pricing algorithm is to maximize the own achievable rate minus a cost term that is linear in the caused interference power. As the interference caused by user k does not depend on the receive filter $\boldsymbol{g}_k$, the receive filter is updated to maximize the own SINR γ_k with a unit norm constraint, leading to the same update as for the selfish and max-SINR algorithms, cf. (5.83).

For the beamformer update of user k, we replace the channel vectors $\boldsymbol{h}_{jk}^{\mathrm{T}}$ in Section 4.4.5 with the equivalent MISO channel vectors ${\boldsymbol{g}_j^{(\ell)}}^{\mathrm{T}}\boldsymbol{H}_{jk}$ and obtain the update problem

$$\boldsymbol{v}_k^{(\ell+1)} = \underset{\boldsymbol{v}_k}{\arg\max}\ \log\left(1+\frac{\left|{\boldsymbol{g}_k^{(\ell)}}^{\mathrm{T}}\boldsymbol{H}_{kk}\boldsymbol{v}_k\right|^2}{\sum_{j\neq k}\left|{\boldsymbol{g}_k^{(\ell)}}^{\mathrm{T}}\boldsymbol{H}_{kj}\boldsymbol{v}_j^{(\ell)}\right|^2+\sigma^2}\right)-\sum_{j\neq k}\pi_j^{(\ell)}\left|{\boldsymbol{g}_j^{(\ell)}}^{\mathrm{T}}\boldsymbol{H}_{jk}\boldsymbol{v}_k\right|^2$$
$$\text{s.t.:}\quad \|\boldsymbol{v}_k\|_2^2\leq 1 \tag{5.103}$$

where the interference price of user $j\neq k$ is

$$\begin{aligned}\pi_j^{(\ell)} &= -\frac{\partial R_j}{\partial|\boldsymbol{g}_j^{\mathrm{T}}\boldsymbol{H}_{jk}\boldsymbol{v}_k|^2}\bigg|_{\substack{\boldsymbol{g}_j=\boldsymbol{g}_j^{(\ell)}\\ \boldsymbol{v}_i=\boldsymbol{v}_i^{(\ell)}\ \forall i}}\\ &= \frac{1}{\sum_{i\neq j}\left|{\boldsymbol{g}_j^{(\ell)}}^{\mathrm{T}}\boldsymbol{H}_{ji}\boldsymbol{v}_i^{(\ell)}\right|^2+\sigma^2}-\frac{1}{\sum_{i}\left|{\boldsymbol{g}_j^{(\ell)}}^{\mathrm{T}}\boldsymbol{H}_{ji}\boldsymbol{v}_i^{(\ell)}\right|^2+\sigma^2}.\end{aligned} \tag{5.104}$$

The derivation of the solution to the beamformer update problem (5.103) is identical to the MISO case, which is discussed in detail in Section 4.4.5 and Appendix A6; to obtain the relevant expressions, we must only replace the MISO channel vectors with the equivalent MISO channels consisting of the MIMO channel and the current receive filter. For reference, we briefly state the resulting update of the beamformer $\boldsymbol{v}_k$ for the rate utility in the following.

We first define the matrix $\boldsymbol{B}$ to contain the vectors $\sqrt{\pi_j^{(\ell)}}\boldsymbol{H}_{jk}^{\mathrm{H}}{\boldsymbol{g}_j^{(\ell)}}^*$ for all $j\neq k$ as columns and perform the reduced SVD $\boldsymbol{B}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}$ such that $\boldsymbol{\Sigma}$ is invertible. We can now use the left singular basis $\boldsymbol{U}$ to check whether zero-forcing is possible for user k and distinguish between the following two cases:

- $(\mathbf{I}-\boldsymbol{U}\boldsymbol{U}^{\mathrm{H}})\boldsymbol{H}_{kk}^{\mathrm{H}}{\boldsymbol{g}_k^{(\ell)}}^*\neq\mathbf{0}$, i. e., zero-forcing is possible:
  We determine the unit-norm vector $\boldsymbol{w}$ and the positive scalar ρ such that $\boldsymbol{w}$ is the eigenvector corresponding to the unique positive eigenvalue of the matrix

$$\boldsymbol{A}(\rho)=\rho\cdot\boldsymbol{H}_{kk}^{\mathrm{H}}{\boldsymbol{g}_k^{(\ell)}}^*{\boldsymbol{g}_k^{(\ell)}}^{\mathrm{T}}\boldsymbol{H}_{kk}-\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}} \tag{5.105}$$

and

$$\rho = \frac{1}{\left|\boldsymbol{g}_k^{(\ell)^{\mathrm{T}}} \boldsymbol{H}_{kk} \boldsymbol{w}\right|^2 + \sum_{j \neq k} \left|\boldsymbol{g}_k^{(\ell)^{\mathrm{T}}} \boldsymbol{H}_{kj} \boldsymbol{v}_j^{(\ell)}\right|^2 + \sigma^2}. \tag{5.106}$$

The solution can be found with a bisection line search on the values of $\left|\boldsymbol{g}_k^{(\ell)^{\mathrm{T}}} \boldsymbol{H}_{kk} \boldsymbol{w}\right|^2$ in the interval $\left[0, \left\|\boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^{(\ell)^*}\right\|_2^2\right]$. The updated beamformer is $\boldsymbol{v}_k^{(\ell+1)} = \boldsymbol{w}$.

- $(\mathbf{I} - \boldsymbol{U}\boldsymbol{U}^{\mathrm{H}})\boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^{(\ell)^*} = \mathbf{0}$, i. e., zero-forcing is not possible:
  In this case it may be necessary to perform power control. To this end we compute the scalar

$$\zeta_2 = \left\|\boldsymbol{\Sigma}^{-1} \boldsymbol{U}^{\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^{(\ell)^*}\right\|_2^2 - \sum_{j \neq k} \left|\boldsymbol{g}_k^{(\ell)^{\mathrm{T}}} \boldsymbol{H}_{kj} \boldsymbol{v}_j^{(\ell)}\right|^2 - \sigma^2 \tag{5.107}$$

  and distinguish between three cases:

  1) $\zeta_2 \leq 0$
     The updated beamformer is $\boldsymbol{v}_k^{(\ell+1)} = \mathbf{0}$.
  2) $0 < \zeta_2 \leq \left\|\boldsymbol{\Sigma}^{-1} \boldsymbol{U}^{\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^{(\ell)^*}\right\|_2^4 \Big/ \left\|\boldsymbol{\Sigma}^{-2} \boldsymbol{U}^{\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^{(\ell)^*}\right\|_2^2$
     The updated beamformer is

$$\boldsymbol{v}_k^{(\ell+1)} = \sqrt{\zeta_2} \frac{\boldsymbol{U} \boldsymbol{\Sigma}^{-2} \boldsymbol{U}^{\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^{(\ell)^*}}{\left\|\boldsymbol{\Sigma}^{-1} \boldsymbol{U}^{\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^{(\ell)^*}\right\|_2^2}. \tag{5.108}$$

  3) $\zeta_2 > \left\|\boldsymbol{\Sigma}^{-1} \boldsymbol{U}^{\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^{(\ell)^*}\right\|_2^4 \Big/ \left\|\boldsymbol{\Sigma}^{-2} \boldsymbol{U}^{\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^{(\ell)^*}\right\|_2^2$
     The updated beamformer is determined in the same way as for the case in which zero-forcing is possible, i. e., we determine $\boldsymbol{w}$ and ρ such that $\boldsymbol{w}$ is the eigenvector corresponding to the unique positive eigenvalue of the matrix

$$\boldsymbol{A}(\rho) = \rho \cdot \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^{(\ell)^*} \boldsymbol{g}_k^{(\ell)^{\mathrm{T}}} \boldsymbol{H}_{kk} - \boldsymbol{B}\boldsymbol{B}^{\mathrm{H}} \tag{5.109}$$

     and

$$\rho = \frac{1}{\left|\boldsymbol{g}_k^{(\ell)^{\mathrm{T}}} \boldsymbol{H}_{kk} \boldsymbol{w}\right|^2 + \sum_{j \neq k} \left|\boldsymbol{g}_k^{(\ell)^{\mathrm{T}}} \boldsymbol{H}_{kj} \boldsymbol{v}_j^{(\ell)}\right|^2 + \sigma^2}. \tag{5.110}$$

     Again, we can be apply a bisection line search of the values of $\left|\boldsymbol{g}_k^{(\ell)^{\mathrm{T}}} \boldsymbol{H}_{kk} \boldsymbol{w}\right|^2$ in the interval $\left[0, \min\left\{\zeta_2, \left\|\boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{g}_k^{(\ell)^*}\right\|_2^2\right\}\right]$. The updated beamformer is $\boldsymbol{v}_k^{(\ell+1)} = \boldsymbol{w}$.

Since the rate utility is convex in the received interference power, the linearization of the other users' utilities leads to an under-estimation of the sum rate, assuming that the other users' strategies remain constant. As the beamformer update maximizes the under-estimated sum rate, the sum rate cannot be decreased by an update. Clearly, also the receiver update cannot decrease the sum rate. With sequential updates, the sum rate therefore converges. Furthermore, it is straightforward to show that a stationary strategy of the pricing algorithm fulfills the KKT conditions of the sum rate maximization problem.

Again, the receive filter updates rely only on information that can be locally estimated. For the transmitter update of user k, knowledge of the cross channels $\boldsymbol{H}_{jk}$ is necessary for all $j \in \{1, \ldots, K\}$ as well as knowledge of the current receive filters weighted with the square root of the prices, i. e., $\sqrt{\pi_j^{(\ell)}}\boldsymbol{g}_j^{(\ell)}$ for all $j \in \{1, \ldots, K\}$, which must be exchanged among the users after each update. Furthermore, for (5.106) and (5.107) transmitter k must know the current interference-plus-noise power at the desired receiver k. We assume, though, that the current SINR $\gamma_k^{(\ell)}$ must be fed back to the transmitter in order to enable the coded transmission of data; therefore the interference-plus-noise power can be computed from $\gamma_k^{(\ell)}$, the direct channel $\boldsymbol{H}_{kk}$, the current beamformer $\boldsymbol{v}_k^{(\ell)}$, and the current receive filter $\boldsymbol{g}_k^{(\ell)}$ without any additional information exchange.

5.5.7 Interference Pricing with Incremental SNR

In [93], we proposed a heuristic procedure to improve the outcome of the simplified pricing algorithm, which can be applied to the (not simplified) pricing algorithm as discussed here as well. We first describe the method and then proceed to discuss the intuition behind it.

For the "incremental SNR" method, we define a decreasing series of noise powers that ends with the actual noise power σ^2:

$$\sigma_{(1)}^2 > \sigma_{(2)}^2 > \ldots > \sigma^2. \tag{5.111}$$

The initialization of the algorithm is the ME strategy. Assuming that the noise power is $\sigma_{(1)}^2$, the pricing algorithm is run until convergence is detected according to some convergence criterion. Next, the resulting strategy is used as an initialization and the pricing algorithm is run until convergence assuming a noise power of $\sigma_{(2)}^2$. This process is continued until convergence is reached for the actual noise power σ^2. The only additional signaling overhead in the incremental SNR method is the detection of convergence and the synchronization of the SNR decrements.

We can equivalently view the incremental SNR algorithm as a procedure in which the transmitters slowly increase their transmit power while the noise power remains constant; as we have assumed unit transmit power throughout this work, denoting the incremental SNR as decreasing noise power leads to a more consistent notation. In a practical implementation, a gradual transmit power increase would perhaps be more suitable than transmitting with full power and (falsely) assuming higher noise power levels.

As is shown in Section 5.6.3, the incremental SNR method generally yields higher sum-rate performance with a lower number of total iterations compared to the pricing algorithm with fixed SNR. We explain this result with the following observations and intuitions:

- At low SNR, i. e., for a high noise power σ^2, the sum rate objective is comparatively "well-behaved": the optimal strategy for asymptotically low SNR can be computed in a straightforward way, cf. Section 5.3.1, and for moderately low SNR most algorithms converge within a comparatively small number of iterations, as is shown in Section 5.6.3.
- At high SNR, on the other hand, the sum rate maximization problem in a fully loaded system has several undesirable properties: first of all, it is intuitively clear that for sufficiently high SNR the aligned solutions correspond to local optima, i. e., the objective function potentially has a very large number of distinct and separate local optima. Second, numerical experiments show that there exist many additional local optima that correspond to strategies that are not aligned solutions; these other local optima can, for example, have the property that the interference is aligned at some, but not all, receivers or that individual users are inactive, i. e., a suboptimal slope is achieved.

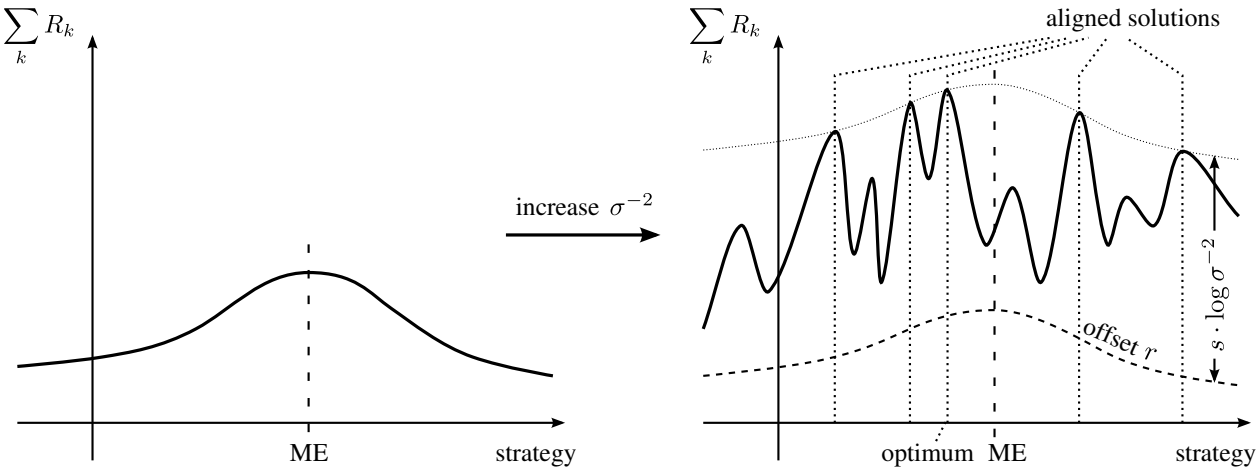


Figure 5.3: Simplified illustration of the sum rate objective function for low SNR (left) and high SNR (right). The strategy consists of all beamformers $\boldsymbol{v}_k$ and receive filters $\boldsymbol{g}_k$ and is therefore actually multi-dimensional; to be able to illustrate the relevant properties of the objective function, we assume a one-dimensional optimization variable.

- Also, the different aligned solutions, i. e., the desirable local optima, may differ significantly in sum-rate performance. In particular, we recall that the performance of an aligned solution in a fully loaded system is described by the sum rate offset

$$r = \sum_k \left| \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{H}_{kk} \boldsymbol{v}_k \right|^2 \tag{5.112}$$

  cf. (5.30), i. e., the performance depends on how well the aligned beamformers and receive filters are "matched" to the direct channel matrices.
- Altogether, it is intuitive that the quality of the strategy determined with a distributed algorithm that uses "local" information, i. e., information that is obtained from measuring the interference power, etc., at the current operating point, depends strongly on the initialization. Algorithms aimed at local optimality in terms of sum rate, such as the pricing algorithm, can converge to local optima that do not achieve the optimal slope s.
- The expression for r is maximized by the ME strategy. We can infer that the best aligned solution will be in some sense "close" to the ME strategy and that the ME strategy may be a good starting point when searching for the best aligned solution.

Therefore, with the incremental SNR method we attempt to "track" a local optimum as it develops with increasing SNR while staying close to the ME solution, which suggests that the so-found aligned solution will be a particularly good aligned solution in terms of the offset r. We illustrate this intuition for a one-dimensional optimization problem in Figure 5.3.

The concept of gradually transforming a problem from a more manageable form into its original, difficult form, is known as "homotopy continuation". In [78], homotopy continuation is also used to find aligned solutions, albeit in a different way: the "start system" is a rank-one approximation of the channel matrices, for which finding a solution to the zero-interference conditions is equivalent to solving a linear system of equations; then, the channel matrices are gradually transformed to their original full-rank form, while the zero-interference solutions are tracked with Newton iterations. In contrast to our incremental SNR method, however, the homotopy method of [78] is not designed to find a good aligned solution in terms of sum rate offset.

5.5.8 Min-Leakage Combined with Sum-Rate Gradient

In [94], an algorithm was proposed that is based on the min-leakage updates, but adds a gradient component in order to find an aligned solution with good sum-rate performance. Each receiver and transmitter update consists of two steps: the first is identical to the update in the min-leakage algorithm and the second is based on moving in the direction of the sum-rate gradient. The size of the second step is initially large and then gradually reduced, or "annealed", so that finally the convergence behavior of the min-leakage algorithm is reached.

For the receiver update, the intermediate receive filter $\bar{\boldsymbol{g}}_k^{(\ell+1)}$ is determined to minimize the sum interference power, i. e., it is the eigenvector corresponding to the smallest eigenvalue of the matrix $\sum_{j\neq k} \boldsymbol{H}_{kj}\boldsymbol{v}_j^{(\ell)}\boldsymbol{v}_j^{(\ell)\mathrm{H}}\boldsymbol{H}_{kj}^{\mathrm{H}}$. Next, the sum-rate gradient at the current operating point is computed:

$$\boldsymbol{d}_k^{(\ell+1)} = \left.\frac{\partial R_k}{\partial \boldsymbol{g}_k^*}\right|_{\substack{\boldsymbol{v}_j=\boldsymbol{v}_j^{(\ell)}\ \forall j \\ \boldsymbol{g}_k=\bar{\boldsymbol{g}}_k^{(\ell+1)}.}} \tag{5.113}$$

Since the receive filter $\bar{\boldsymbol{g}}_k^{(\ell+1)}$ has unit norm, the rate of user k is

$$R_k = \log\left(\sum_j |\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_{kj}\boldsymbol{v}_j|^2 + \sigma^2\right) - \log\left(\sum_{j\neq k} |\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_{kj}\boldsymbol{v}_j|^2 + \sigma^2\right) \tag{5.114}$$

and the complex conjugate gradient vector is

$$\begin{aligned}\boldsymbol{d}_k^{(\ell+1)^*} &= \frac{1}{\sum_i \left|\bar{\boldsymbol{g}}_k^{(\ell+1)^{\mathrm{T}}}\boldsymbol{H}_{ki}\boldsymbol{v}_i^{(\ell)}\right|^2 + \sigma^2}\boldsymbol{H}_{kk}\boldsymbol{v}_k^{(\ell)}\boldsymbol{v}_k^{(\ell)\mathrm{H}}\boldsymbol{H}_{kk}^{\mathrm{H}}\bar{\boldsymbol{g}}_k^{(\ell+1)^*} \\ &- \left(\frac{1}{\sum_{i\neq k}\left|\bar{\boldsymbol{g}}_k^{(\ell+1)^{\mathrm{T}}}\boldsymbol{H}_{ki}\boldsymbol{v}_i^{(\ell)}\right|^2 + \sigma^2} - \frac{1}{\sum_i \left|\bar{\boldsymbol{g}}_k^{(\ell+1)^{\mathrm{T}}}\boldsymbol{H}_{ki}\boldsymbol{v}_i^{(\ell)}\right|^2 + \sigma^2}\right)\sum_{j\neq k}\boldsymbol{H}_{kj}\boldsymbol{v}_j^{(\ell)}\boldsymbol{v}_j^{(\ell)\mathrm{H}}\boldsymbol{H}_{kj}^{\mathrm{H}}\bar{\boldsymbol{g}}_k^{(\ell+1)^*}.\end{aligned} \tag{5.115}$$

The gradient vector is then projected into the space orthogonal to the intermediate receive filter $\bar{\boldsymbol{g}}_k^{(\ell+1)}$:

$$\bar{\boldsymbol{d}}_k^{(\ell+1)} = \left(\mathbf{I} - \bar{\boldsymbol{g}}_k^{(\ell+1)}\bar{\boldsymbol{g}}_k^{(\ell+1)\mathrm{H}}\right)\boldsymbol{d}_k^{(\ell+1)}. \tag{5.116}$$

With the step size $\kappa^{(\ell)}$, the updated receive filter is

$$\boldsymbol{g}_k^{(\ell+1)} = \bar{\boldsymbol{g}}_k^{(\ell+1)}\cos\left(\kappa^{(\ell)}\left\|\bar{\boldsymbol{d}}_k^{(\ell+1)}\right\|_2\right) + \frac{\bar{\boldsymbol{d}}_k^{(\ell+1)}}{\left\|\bar{\boldsymbol{d}}_k^{(\ell+1)}\right\|_2}\sin\left(\kappa^{(\ell)}\left\|\bar{\boldsymbol{d}}_k^{(\ell+1)}\right\|_2\right). \tag{5.117}$$

Note that for $\kappa^{(\ell)} = 0$ the update is simply the min-leakage update $\bar{\boldsymbol{g}}_k^{(\ell+1)}$. For $\kappa^{(\ell)} > 0$ an orthogonal term obtained from the gradient vector is added; weighting the summands with cosine and sine ensures that the result has unit norm. As noted in [94], a step is taken following the "geodesic on the Grassmanian manifold."

For the beamformer update, the principle of switching the roles of transmitters and receivers is applied, as was already discussed for the min-leakage and max-SINR algorithms: in the reverse

direction, receiver j acts as a transmitter and uses the vector $\boldsymbol{g}_j$ as a beamformer. Transmitter k then applies the above receive filter update to its beamformer $\boldsymbol{v}_k$. The expression for $\boldsymbol{v}_k^{(\ell+1)}$ can therefore be found by replacing $\boldsymbol{g}_j^*$ with $\boldsymbol{v}_j$ and vice versa for all $j \in \{1, \ldots, K\}$, and replacing all channel matrices $\boldsymbol{H}_{jk}$ with $\boldsymbol{H}_{kj}^{\mathrm{H}}$.

In [94], it was suggested that a reasonable choice for the initial step size is $\kappa^{(0)} = 0.1$; after each beamformer and receive filter has been updated, the step size is then decreased by a multiplication with 0.995. Thus, in the course of the iterations the step size approaches zero and the algorithm converges in the same way as the min-leakage algorithm.

The receiver update requires knowledge of $\sum_{j \neq k} \boldsymbol{H}_{kj} \boldsymbol{v}_j^{(\ell)} \boldsymbol{v}_j^{(\ell)^{\mathrm{H}}} \boldsymbol{H}_{kj}^{\mathrm{H}}$ and of $\boldsymbol{H}_{kk} \boldsymbol{v}_k^{(\ell)} \boldsymbol{v}_k^{(\ell)^{\mathrm{H}}} \boldsymbol{H}_{kk}^{\mathrm{H}}$; we assume that both can be locally estimated. In our signaling framework without time-division duplex, the beamformer update of user k can be performed by exchanging the channel matrices $\boldsymbol{H}_{jk}$ initially and the receive filters $\boldsymbol{g}_j^{(\ell)}$ after each update.

5.5.9 Other Algorithms

The above list of algorithms is not exhaustive. We chose to omit some further algorithms so that our comparison does not to become too cluttered. In [95], for example, a gradient method was proposed, which was found in [94] to have poor convergence properties; this is consistent with our observations in the MISO case, where for certain SNR values the projected gradient algorithm takes very many iterations to converge and the issue of choosing a universally suitable step size appears to be a problem. In [96], the min-leakage objective was augmented with an additional summand that rewards a high power gain of the desired signal; it is not clear, however, how to choose the weight of the additional term other than by trial and error, and a good choice of the weight appears to be strongly dependent on the noise power. In [97], an MMSE-based algorithm was proposed for a setting in which the sum power of all users is constrained instead of the individual users' powers.

Some other algorithms, e. g., those proposed in [89, 75], are based on optimizing the transmit covariance matrices $\boldsymbol{Q}_k$ instead of the beamformers $\boldsymbol{v}_k$. To apply these algorithms to the single-stream case, however, an additional constraint would be necessary that allows the covariance matrices to have at most a rank of one; adding such a constraint is not trivial. On the other hand, it should be noted that many of the algorithms discussed above have also been applied or can be extended to the multi-beam case.

5.5.10 Comparison of the Distributed Algorithms

An overview of the discussed distributed algorithms is given in Table 5.4. As the second column shows, many of the algorithms do not have the ability to reduce the power of the beamformers or deactivate users; these algorithms depend on being initialized with the correct number of users. Also, as shown in the third column, some algorithms are aimed specifically at optimizing the sum rate, while others are based on a different objective that happens to also yield good sum-rate performance. Those algorithms that maximize the sum rate can be easily extended to handle similar objectives, such as the weighted sum rate. Finally, the table shows whether convergence in some metric has been proven and what sort of computations are necessary for one beamformer update.

Table 5.5 shows the information exchange requirements of the different algorithms for the signaling model discussed in Section 3.5. The numerator and denominator of $\gamma_k^{(\ell)}$ are denoted by $n_k^{(\ell)}$ and $d_k^{(\ell)}$, respectively. Notably, the global SINR maximization algorithm also requires information to be signaled from the transmitters to their respective desired receivers for the receive filter update.

Algorithm	Power control	Sum rate	Convergence	Complexity
Selfish updates	No	No	No	Inverse
Min-leakage	No	No	Yes	EVD
Max-SINR	No	No	Local	Inverse
Global SINR	No	No	Seq. updates	Inverse, EVD
MMSE	Yes	No	Yes	Inverse, line search
Adaptive MMSE	Yes	Yes	Yes	Inverse, line search
Pricing	Yes	Yes	Seq. updates	EVD, line search
Incremental SNR	Yes	Yes	Seq. updates	EVD, line search
Min-leakage/gradient	No	Yes	Yes	EVD

Table 5.4: Overview of the discussed algorithms for single-stream MIMO interference networks

Algorithm	Rx $k \to$ Tx k once	Tx $k \to$ Tx j once	Rx $k \to$ Tx k per iteration	Tx $k \to$ Tx j per iteration	Tx $k \to$ Rx k per iteration
Selfish updates	$\boldsymbol{H}_{kk}$	—	$\gamma_k^{(\ell)}, \boldsymbol{g}_k^{(\ell)}$	—	—
Min-leakage	$\boldsymbol{H}_{kj}\ \forall j \neq k$	$\boldsymbol{H}_{kj}$	$\gamma_k^{(\ell)}, \boldsymbol{g}_k^{(\ell)}$	$\boldsymbol{g}_k^{(\ell)}$	—
Max-SINR	$\boldsymbol{H}_{kj}\ \forall j$	$\boldsymbol{H}_{kj}$	$\gamma_k^{(\ell)}, \boldsymbol{g}_k^{(\ell)}$	$\boldsymbol{g}_k^{(\ell)}$	—
Global SINR	$\boldsymbol{H}_{kj}\ \forall j$	$\boldsymbol{H}_{kj}$	$\gamma_k^{(\ell)}, \boldsymbol{g}_k^{(\ell)}$	$\boldsymbol{g}_k^{(\ell)}, n_k^{(\ell)}, d_k^{(\ell)}$	$a_k^{(\ell)}, c_k^{(\ell)}$
MMSE	$\boldsymbol{H}_{kj}\ \forall j$	$\boldsymbol{H}_{kj}$	$\gamma_k^{(\ell)}, \boldsymbol{g}_k^{(\ell)}$	$\boldsymbol{g}_k^{(\ell)}$	—
Adaptive MMSE	$\boldsymbol{H}_{kj}\ \forall j$	$\boldsymbol{H}_{kj}$	$\gamma_k^{(\ell)}, \boldsymbol{g}_k^{(\ell)}$	$\boldsymbol{g}_k^{(\ell)}, \alpha_k$	—
Pricing	$\boldsymbol{H}_{kj}\ \forall j$	$\boldsymbol{H}_{kj}$	$\gamma_k^{(\ell)}, \boldsymbol{g}_k^{(\ell)}$	$\sqrt{\pi_k^{(\ell)}}\boldsymbol{g}_k^{(\ell)}$	—
Incremental SNR	$\boldsymbol{H}_{kj}\ \forall j$	$\boldsymbol{H}_{kj}$	$\gamma_k^{(\ell)}, \boldsymbol{g}_k^{(\ell)}$	$\sqrt{\pi_k^{(\ell)}}\boldsymbol{g}_k^{(\ell)}$	—
Min-leakage/grad.	$\boldsymbol{H}_{kj}\ \forall j$	$\boldsymbol{H}_{kj}$	$\gamma_k^{(\ell)}, \boldsymbol{g}_k^{(\ell)}$	$\boldsymbol{g}_k^{(\ell)}$	—

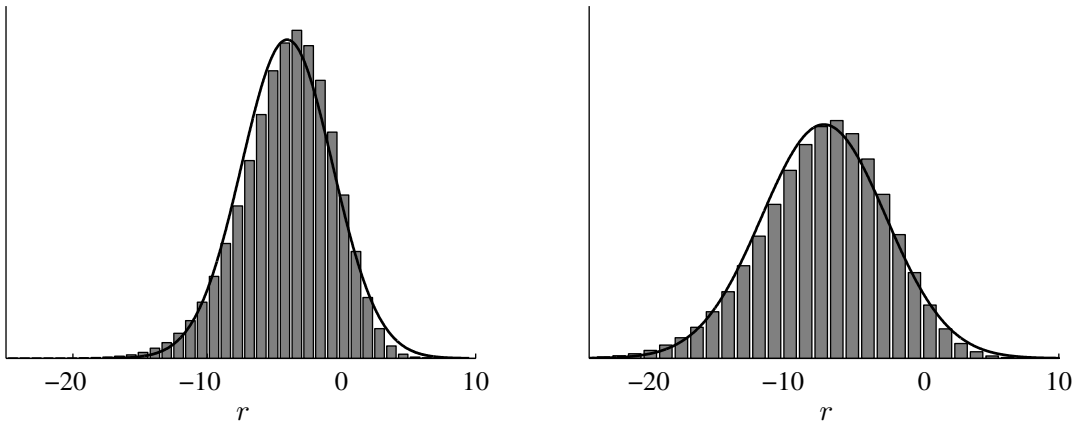
Table 5.5: Information exchange requirements of the MIMO algorithms assuming that the transmitters compute the updates and are connected via signaling links, cf. Figure 3.6(a).

We again note that the weights α_k in the adaptively weighted MMSE algorithm only need to be exchanged when they are updated, which can be significantly less frequently than the beamformer/receive filter updates. Also, the incremental SNR algorithm requires some sort of signaling mechanism for detecting convergence and moving on to the next SNR value.

5.6 Numerical Evaluation of the Approximations and Algorithms

In this section we numerically compare the sum-rate performance of the discussed distributed algorithms for single-stream MIMO interference networks in an i. i. d. Gaussian channel model. As in the MISO case, we distinguish between systems where zero interference is possible and systems where users must be deactivated at high SNR; also, we again include results for scenarios in which the cross channel gains are significantly lower than the direct channel gains.

The discussed algorithms differ significantly in their convergence properties. As the number of iterations required for satisfactory performance is perhaps the most important metric for the practical relevance of the algorithms, we pay special attention to this issue as well.



(a) $K = 7$ (b) $K = 13$

Figure 5.4: Measured distribution of the sum rate offset r of a random aligned solution compared to the Gaussian approximation

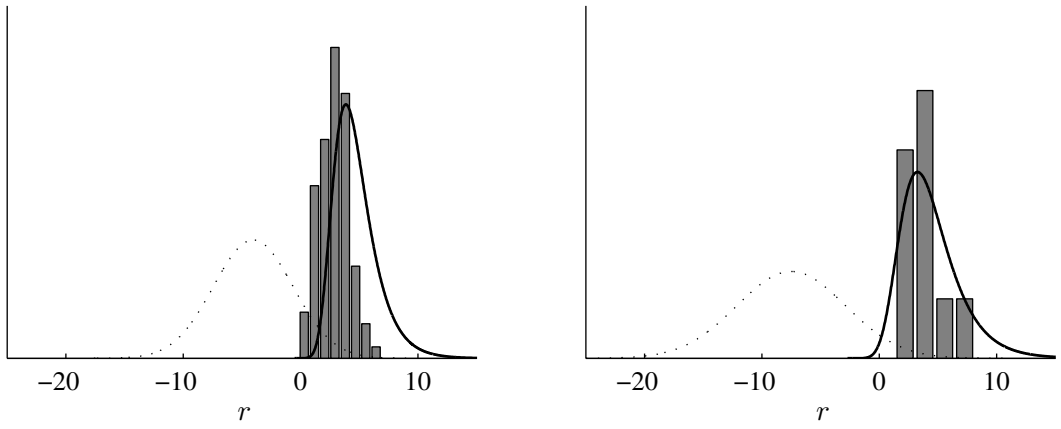
To begin with, however, we revisit the approximation for the mean sum rate offset of the best of L aligned solutions discussed in Section 5.3.5.1. We recall that, in order to obtain the expression for $\bar{r}_{\text{approx}}$ in (5.54), some assumptions were made for which we did not provide any numerical evidence. Now that the min-leakage algorithm provides us with a method of generating different random aligned solutions, we can examine the validity of the approximation more closely.

5.6.1 Numerical Validation of the High-SNR Offset Approximations

We made two simplifying assumptions in order to be able to approximate the limiting distribution of the offset of the best of L aligned solutions in Theorem 5.2: first, we assumed that the distribution of the offset of a random aligned solution is Gaussian; second, we assumed that the offsets of different aligned solutions of one channel realization are statistically independent. In the following we perform some experiments to gain some intuition concerning the validity of these assumptions.

The rationale behind the Gaussianity assumption is that the offset r is the sum of K independent random variables and therefore approaches a Gaussian distribution for large K according to the central limit theorem; since we know the distribution of the individual summands, cf. (5.42), we can numerically generate realizations of r and compare the histogram with the idealized Gaussian PDF. The result of this experiment for two fully loaded systems with $K = 7$ and $K = 13$ users, respectively, is shown in Figure 5.4. The figures indicate that the shape of the measured PDF is reasonably close to the Gaussian approximation, but that it appears to be slightly skewed to the left. The visible difference between measured and approximated PDF decreases with a growing number of users K.

The assumption of statistical independence of the aligned solutions is more difficult to validate. To obtain some limited insight into this issue, we perform the following experiment: for every channel realization, we determine two different aligned solutions by running the min-leakage algorithm until convergence from two random (isotropically distributed) initializations. We then estimate the correlation coefficient [81, Chapter 6] by computing the sample mean over C channel realizations. Since we know that the mean of r is $-K\gamma$ and the variance is $K\pi^2/6$, the estimated correlation



(a) $K = 7$ and $N = M = 4$ (b) $K = 13$ and $N = M = 7$

Figure 5.5: Measured distribution of the sum offset r of the best of $L = 100$ aligned solutions compared with the limiting Gumbel approximation. The dotted line represents the Gaussian approximation of the sum offset of a random aligned solution.

coefficient is

$$\bar{\rho} = \frac{1}{C} \sum_{c=1}^{C} \frac{\left(r_c^{(1)} + K\gamma\right)\left(r_c^{(2)} + K\gamma\right)}{\frac{K\pi^2}{6}} \tag{5.118}$$

where $r_c^{(1)}$ and $r_c^{(2)}$ are the sum rate offset of the first and the second aligned solution of the cth channel realization, respectively, and C is the number of total channel realizations. In a fully loaded system with $N = M = 4$ antennas and $K = 7$ users, we measured a correlation coefficient of 0.07 averaged over $C = 1000$ channel realizations; for $N = M = 7$ and $K = 13$, the measured correlation coefficient was even below 0.01. These results indicate that the correlation between different aligned solutions is not very high; we note, however, that pairwise independence is not sufficient for mutual independence and that this experiment is only meant to provide an intuition and not conclusive evidence.

Finally, we investigate how well the properly scaled and shifted limiting Gumbel distribution of Theorem 5.2 matches the experimentally obtained distribution of the offset of the best of L aligned solutions. For the results in Figure 5.5 we determined the best of $L = 100$ aligned solutions, again using different random initializations for the min-leakage algorithm. For the system with $N = M = 4$ antennas and $K = 7$ users, we generated the histogram from 100 channel realizations, for the larger system with $N = M = 7$ and $K = 13$ the histogram was computed using 20 realizations. It appears that for $K = 4$ the approximation yields an over-estimation of the offset; the reason most likely is the left skew of the distribution of r, leading to a "shorter" right tail of the PDF. For $K = 13$ the Gumbel distribution matches the histogram very well.

Altogether, the results indicate that the approximation is reasonable and that it becomes more accurate with growing system dimensions. Again, for very large values of L the approximation should be used with caution; as discussed in Section 5.3.5.1, it is not known how accurate the Gaussian approximation of the distribution of r is far on the right tail of the PDF and it is computationally infeasible to conclusively validate the approximation in this case, as each execution of the min-leakage algorithm requires a considerable computational effort.

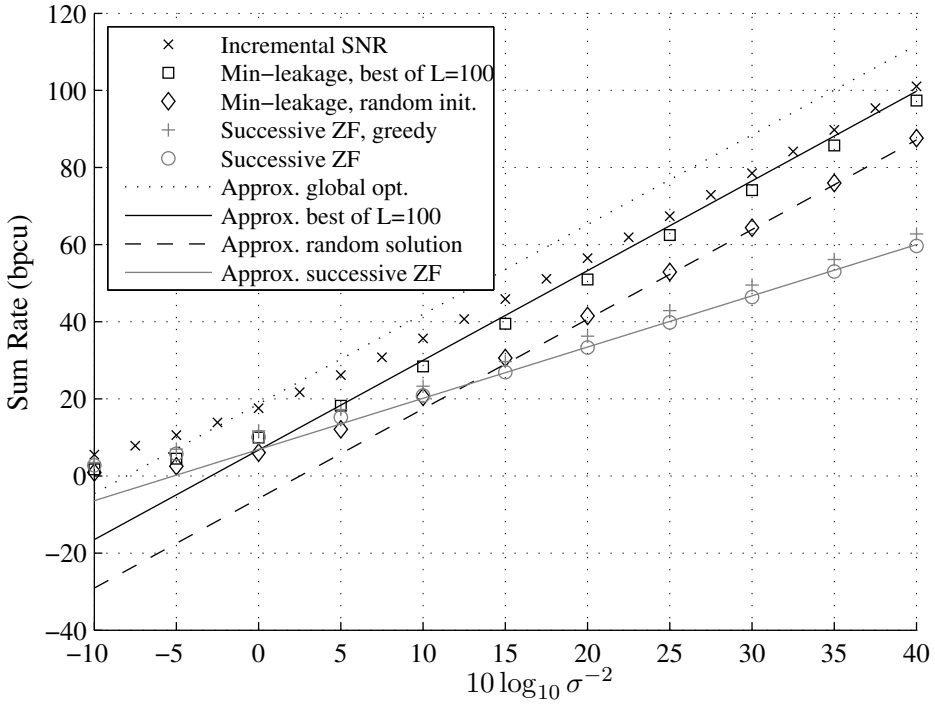


Figure 5.6: $K = 7$ users, $N = M = 4$ antennas at each transmitter and receiver, sum rate averaged over 100 channel realizations.

5.6.2 Performance of the Best of L Aligned Solutions

The performance approximation by means of slope and offset is a high-SNR approximation. In this section we compare the sum-rate performance of the strategy of selecting the best of L solutions (determined with the min-leakage algorithm) with the corresponding high-SNR asymptote, not only for high SNR, but across the whole range of SNRs. We also include the successive zero-forcing scheme, for which we derived an approximate high-SNR asymptote in Section 5.4.2.1.

Figure 5.6 shows the result of averaging the sum rate over 100 channel realizations for a fully loaded system with $K = 7$ users and $N = M = 4$ antennas at each node. It is known that in this setting the total number of aligned solutions is $T = 1\,975\,560$, cf. Table 5.2; we can therefore approximate the asymptote of the global optimum by assuming $L = T$. The approximated asymptote for the best of $L = 100$ solutions slightly over-estimates the measured performance, which is consistent with Figure 5.5. For the random aligned solution, the measured performance and the asymptote match very well at high SNR; this is to be expected since the derivation of the offset and slope of a random aligned solution does not contain any simplifying assumptions (other than $\sigma^{-2} \to \infty$). For comparison with the other distributed algorithms we also included the measured performance of the incremental SNR algorithm, which in the following section will be shown to yield the best performance among all algorithms in this setting. Assuming that our approximation for the global optimum is reasonable, all investigated distributed algorithms are therefore clearly suboptimal. It is, however, also remarkable that the incremental SNR algorithm achieves better performance than the min-leakage algorithm run until convergence a hundred times. The parameters used for the iterative algorithms, such as the convergence criterion and the maximum number of iterations, are discussed in the next section.

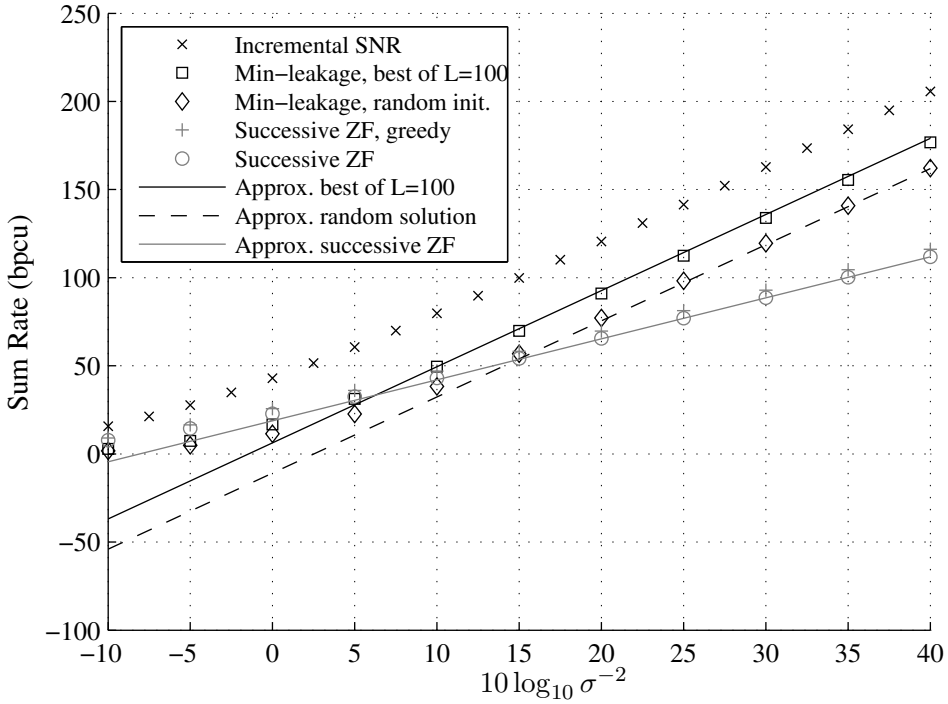


Figure 5.7: $K = 13$ users, $N = M = 7$ antennas at each transmitter and receiver, sum rate averaged over 20 channel realizations.

For the approximated asymptote of the successive zero-forcing strategy, we used the offset $\bar{r}_{\mathrm{MSA}}$, cf. (5.82), which was already shown to approximate the numerically measured offset very well, cf. Table 5.3. Figure 5.6 shows that the performance of the successive zero-forcing strategy is accurately approximated by its asymptote for moderate SNRs as well. A small performance gain can be achieved by allocating the strategies to the users in a greedy fashion, as explained in Section 5.4.2.

Figure 5.7 shows the same comparison for a system with $K = 13$ users and $N = M = 7$ antennas at each node. Due to the higher computational complexity of running the iterative algorithms for these system dimensions, the results are averaged only over 20 channel realizations. For this setting the total number of aligned solutions is not known, and we therefore cannot approximate the global optimum at high SNR. The behavior is for the most part similar to Figure 5.6; the approximation of the best of $L = 100$ aligned solutions in this scenario is very accurate. Also, the incremental SNR method has far better performance than the best out of 100 random aligned solutions.

In Figures 5.6 and 5.7, we observe that the SNR at which the asymptote of the successive zero-forcing strategy and the best-of-L strategy cross is higher for the larger system: while for $K = 7$ the two asymptotes cross around 0 dB, for $K = 13$ the crossing point is around 6 dB. Since the approximations are fairly simple to compute, we can investigate how the crossing point evolves with growing system dimensions: in a generic system with $N = M$ antennas at each node, we know that the best-of-L strategy has a slope of $s = 2N-1$ and the successive zero-forcing strategy achieves a slope of $s = N$. The noise power $\sigma^2_{\mathrm{cross}}$ at which the two approximated asymptotes cross

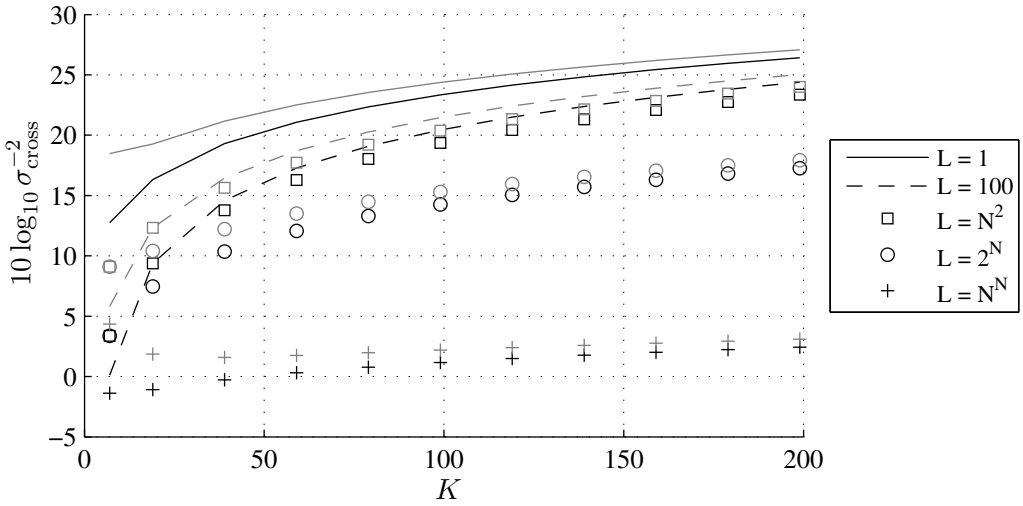


Figure 5.8: Evolution of the SNR at which the approximated asymptotes for the best-of-L strategy and the successive zero-forcing strategy cross. For the black plots, the approximation $\bar{r}_{\text{MSA}}$ was used for the offset of the successive zero-forcing strategy, for the gray plots $\bar{r}_{\text{LSA}}$ was used.

is therefore defined by

$$\bar{r}_{\text{approx}} + (2N-1)\log\sigma_{\text{cross}}^{-2} = \bar{r}_{\text{MSA}} + N\log\sigma_{\text{cross}}^{-2} \quad \Rightarrow \quad \log\sigma_{\text{cross}}^{-2} = \frac{\bar{r}_{\text{MSA}} - \bar{r}_{\text{approx}}}{N-1}. \tag{5.119}$$

The behavior of σ_{cross}^2 for up to $K = 2N - 1 = 199$ users is shown in Figure 5.8, where the gray plots result from using the offset approximation $\bar{r}_{\text{LSA}}$ instead of $\bar{r}_{\text{MSA}}$. Since the number of aligned solutions grows rapidly in the system dimensions, we also investigate different increasing relationships between L and N. Remarkably, even for $L = N^N$, i. e., for a super-exponential growth of the number of sampled aligned solutions with the system dimensions, the crossing SNR $\sigma_{\text{cross}}^{-2}$ increases with the system dimensions. We conclude from this brief investigation that achieving the optimal slope in a MIMO interference network can incur a significant penalty in the offset, so that at moderately high SNR searching for aligned solutions may not be a good approach.

5.6.3 Performance of the Distributed Algorithms

For the numerical comparison of the distributed algorithms discussed in Section 5.5 we use the i. i. d. Gaussian channel model, where all elements of all channel matrices are independent and complex Gaussian. The elements of the direct channel matrices $\boldsymbol{H}_{kk}$ have unit variance and the elements of the cross channel matrices $\boldsymbol{H}_{kj}$ with $k \neq j$ have either variance one or variance 0.01, depending on the scenario.

To detect convergence of the algorithms, we again examine the sum of the Euclidean norms of the differences between the previous and the current beamformers, as is explained in Section 4.5.2; we do not use the receive filters for our convergence criterion. The convergence threshold and the maximum iteration number were chosen to ensure that the sum-rate performance is as good as possible across the whole examined range of noise powers; to determine the values given in Table 5.6 we compared different convergence criteria for a fully loaded system with $K = 7$ users and $N = M = 4$ antennas at each node and selected a criterion that did not lead to a significant performance loss. The resulting iteration numbers are rather high, which can be argued to limit the

Update mode	Parallel
Max. iteration number	10 000 / 100 000
Convergence threshold	10^{-4} / 10^{-5}
Newton steps	10
Bisection steps	20
Initial noise power $\sigma^2_{(1)}$	10
SNR increments	2.5 dB
Initial step size $\kappa^{(0)}$	0.1
Step size reduction factor	0.995

Table 5.6: Parameters used for the numerical evaluation of the algorithms: the MMSE-based algorithms and the pricing algorithm (without incremental SNR) are allowed 100 000 iterations and have a convergence threshold of 10^{-5}, the other algorithms are allowed 10 000 iterations and have a convergence threshold of 10^{-4}.

practical relevance of the algorithms. This is, however, intended, as we wish to investigate the full potential of the algorithms at this point and rule out negative effects due to prematurely terminating the iterations. Later, in Section 5.6.3.4, we briefly examine the performance when a lower number of iterations is allowed.

In contrast to the MISO case, we found that it is crucial for the performance of the adaptively weighted MMSE algorithm that the weights are only updated when convergence of the beamformers has been detected. The adaptive MMSE algorithm is therefore operated with an inner loop and an outer loop: in the inner loop the weights are fixed and the weighted MMSE updates are run until convergence of the beamformers, in the outer loop the weights are updated according to the current operating point; convergence of the algorithm is reached when the change of the beamformers from one iteration of the outer loop to the next is below the threshold.

The sequence of noise powers used for the incremental SNR algorithm is also given in Table 5.6. We start with a noise power of $\sigma^2_{(1)} = 10$ and reduce the noise power after convergence of the pricing algorithm by 2.5 dB, which corresponds to a multiplication with approximately 0.562. The limit on the number of iterations applies to each value of the noise power separately, i. e., the total number of iterations to reach convergence for the desired noise power may be higher.

The step size parameters for the min-leakage/gradient algorithm are adopted from [94], i. e., the initial step size is $\kappa^{(0)} = 0.1$ and after each beamformer and receiver has been updated, the step size is multiplied with 0.995. All algorithms are operated with parallel updates; the performance and convergence behavior with sequential updates is very similar. In the successive zero-forcing scheme the strategies are allocated in the order of the user indices, i. e., not in a greedy fashion.

In the following we compare nine iterative algorithms, each of which may take well over a thousand iterations to converge, for a broad range of scenarios and noise powers; furthermore, some of the algorithms require eigenvalue decompositions and/or line searches in every iteration. Altogether, the following comparisons are computationally very demanding, so that the average is taken only over 100 channel realizations instead of 1000 in the previous chapters. Nonetheless, we believe the results to be meaningful, as simulations with 1000 channel realizations for selected algorithms and noise powers indicate that the "long term" average is already practically achieved after 100 realizations.

Algorithm	–10 dB	15 dB	40 dB	–10 dB	15 dB	40 dB
Selfish updates	18 (0)	10 000 (67)	10 000 (78)	8 (0)	21 (1)	10 000 (70)
Min-leakage	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)
Max-SINR	24 (0)	1103.5 (0)	8594 (35)	9 (0)	37 (0)	2017 (1)
Global SINR	27 (0)	1334 (0)	4255 (13)	10 (0)	40 (0)	2337 (2)
MMSE	36 (0)	2026 (0)	96 363 (46)	25 (1)	287 (0)	51 527 (4)
Adaptive MMSE	80.5 (0)	5282.5 (0)	100 000 (75)	26 (1)	640.5 (2)	68 966.5 (14)
Pricing	33.5 (0)	1816 (0)	70 016.5 (23)	14 (0)	50.5 (0)	4009 (0)
Incremental SNR	25 (0)	796.5 (0)	920 (0)	10 (0)	194 (0)	885 (0)
Min-leakage/gradient	1210.5 (0)	882 (0)	1165.5 (0)	1221 (0)	1030.5 (0)	1139 (0)

Table 5.7: Median number of iterations until convergence for Figure 5.9 (left) and Figure 5.10 (right). In parentheses is the percentage of channel realizations for which the algorithm did not converge before reaching the maximum iteration number.

5.6.3.1 Underutilized System: $K = 4$ Users and $N = M = 4$ Antennas

Throughout the following numerical evaluation, we focus on systems with four antennas at each node. We begin with a scenario with $K = 4$ users; clearly, this is a proper system, i. e., a slope of $s = K = 4$ is almost always achievable. Furthermore, the optimal slope can be achieved without iterative algorithms: if we consider any K-tuple of fixed receive filters, we effectively have a MISO system with four users and four transmit antennas, for which a zero-forcing solution achieving the optimal slope can be straightforwardly computed. Nonetheless, iterative algorithms involving an optimization of the receive filters may still improve the sum rate offset as well as the performance at moderate and low SNR.

Figures 5.9 and 5.10 show the sum-rate performance of the discussed non-iterative techniques and iterative algorithms for strong and weak cross channels, respectively; the corresponding median iteration numbers are shown in Table 5.7, with the percentage of channel realizations for which convergence was not achieved within the maximum iteration number in parentheses. We observe that all iterative algorithms appear to achieve the optimal slope. Even the selfish update procedure, where the caused interference is not taken into account in the beamformer updates, is able to remove all interference, as the receivers have sufficiently many antennas to seperate the desired signal from the interfering signals; however, the selfish updates very often fail to converge.

The min-leakage algorithm leads to a significantly suboptimal offset; convergence is reached after one update, however, as all interference can be removed immediately with zero-forcing beamformers. Adding the sum-rate gradient step to the min-leakage algorithm yields an improvement, but the performance of the other iterative algorithms is not reached. The other iterative algorithms perform approximately equally, but exhibit substantial differences in the convergence behavior: the MMSE-based algorithms and the interference pricing algorithm require a very high number of iterations at high SNR, whereas the incremental SNR algorithm consistently converges comparatively quickly. Altogether, the iterative algorithms yield an improvement of no more than 5 bpcu over the successive zero-forcing strategy, so that it is disputable whether such computationally demanding iterative procedures are worthwile in these scenarios.

5.6.3.2 Fully Loaded System: $K = 7$ Users and $N = M = 4$ Antennas

With $K = 7$ users the four-antenna system is fully loaded: consequently, interference alignment is achievable without deactivating users and there is a finite number of aligned solutions. The

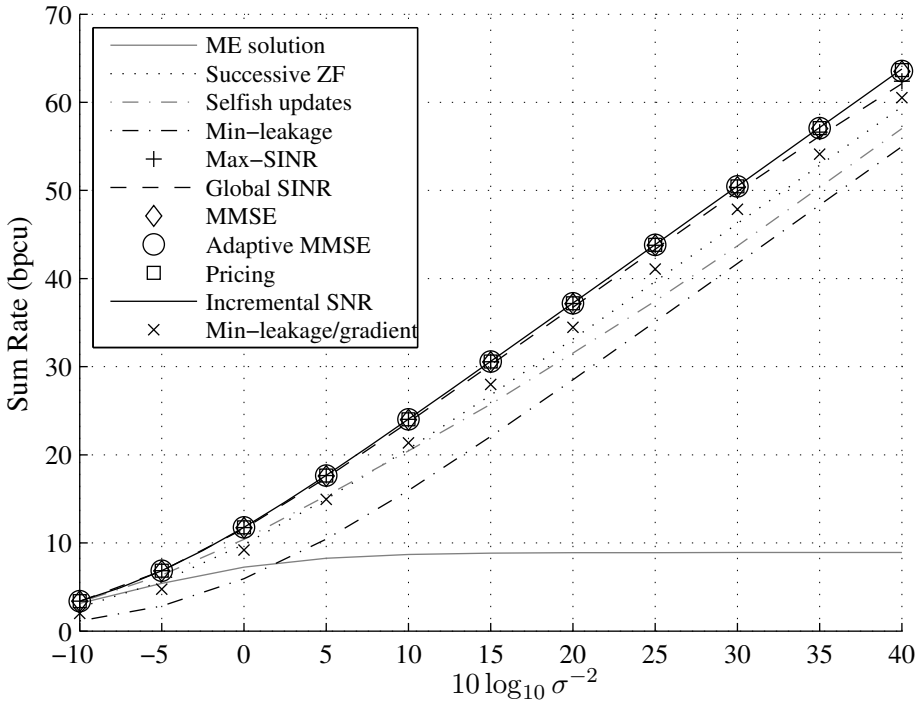


Figure 5.9: $K = 4$ users, $N = M = 4$ antennas at each transmitter and receiver, sum rate averaged over 100 channel realizations. The iterative algorithms are initialized with the ME solution.

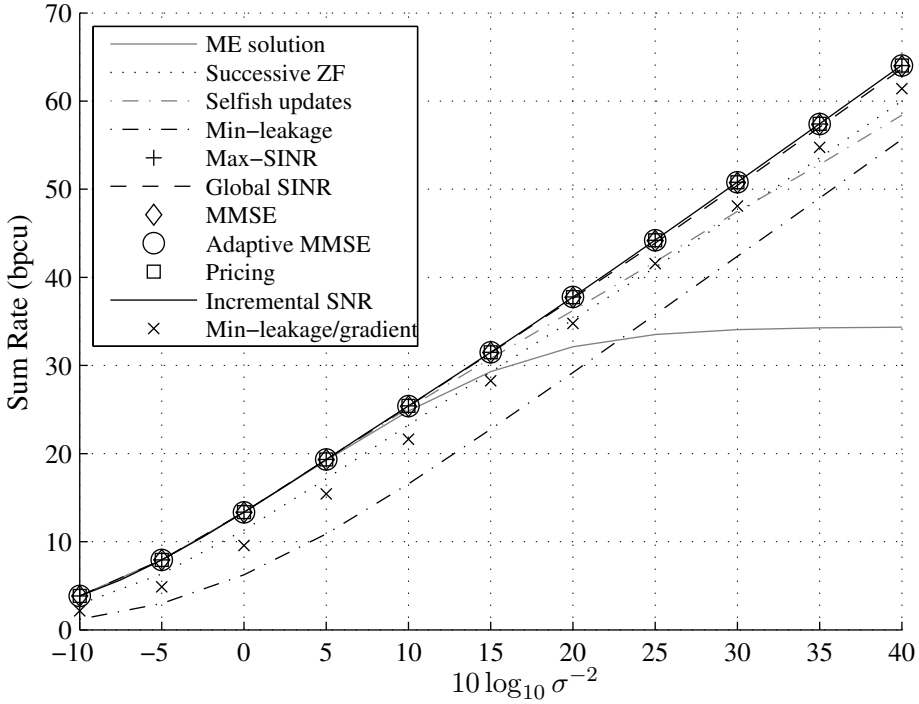


Figure 5.10: $K = 4$ users, $N = M = 4$ antennas at each transmitter and receiver, sum rate averaged over 100 channel realizations of a channel model where the variance of the cross channels is 0.01 and the variance of the direct channels is 1. The iterative algorithms are initialized with the ME solution.

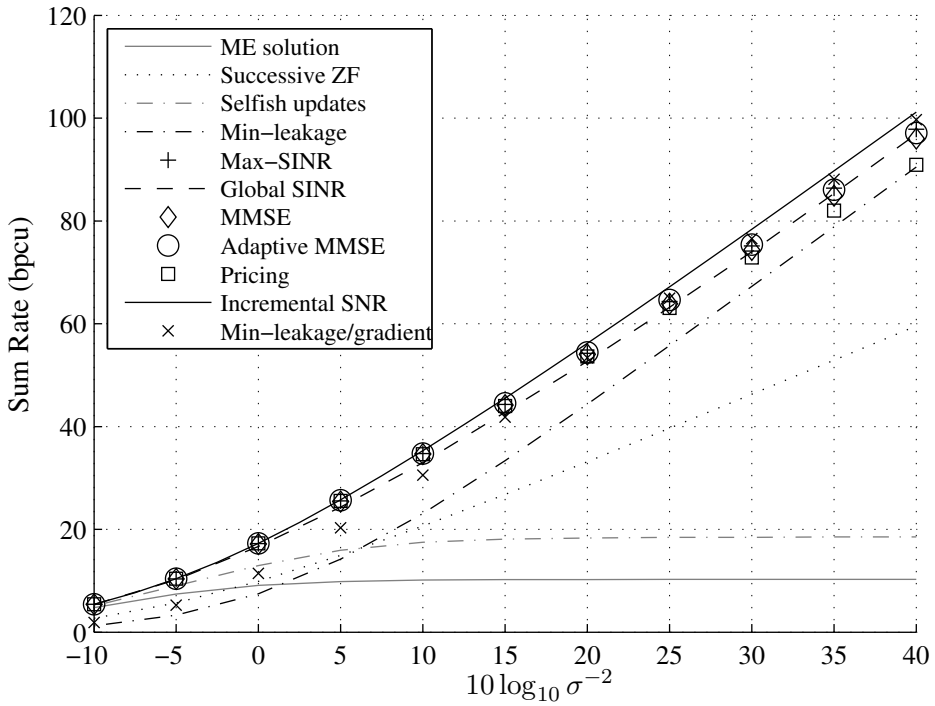


Figure 5.11: $K = 7$ users, $N = M = 4$ antennas at each transmitter and receiver, sum rate averaged over 100 channel realizations. The iterative algorithms are initialized with the ME solution.

sum-rate performance of the different strategies is plotted over the SNR for strong and weak cross channels in Figures 5.11 and 5.12, respectively, the iteration numbers are given in Table 5.8.

In contrast to the previously discussed underutilized scenario, here the sum rate achieved by the selfish updates saturates for high SNR, as alignment can only be achieved by taking into account the caused interference. For the successive zero-forcing strategy we can observe that an asymptotic slope of $4\log_2 10 \approx 13.3$ bpcu per 10 dB increase in SNR is achieved, which corresponds to $s = 4$ when rate and noise power are expressed with the natural logarithm, whereas the iterative schemes achieve a slope of $s = 7$, or approximately 23.3 bpcu per 10 dB.

Of the algorithms that achieve intereference alignment, the incremental SNR algorithm on average appears to find the best solution at the cost of a higher number of iterations than, e. g., the max-SINR algorithm. For the case of strong cross channels in Figure 5.11, the performance of the incremental SNR method is closely followed by that of the min-leakage/gradient and max-SINR algorithms; for weak cross channels, the min-leakage/gradient heuristic places too much emphasis on removing all interference and performance is significantly poorer. The global SINR maximization algorithm in general has convergence properties that are very similar to those of the max-SINR algorithm, the average sum-rate performance is, however, marginally lower.

The MMSE-based algorithms yield good performance, but, as was also observed in the MISO case, require extremely many iterations to converge when the SNR is high. Similarly, the pricing algorithm without incremental SNR has difficulty finding an aligned solution at high SNR in the case of strong cross channels; this has a visible effect on the slope above an SNR of 30 dB in Figure 5.11 and can also be seen from the iteration numbers in Table 5.8.

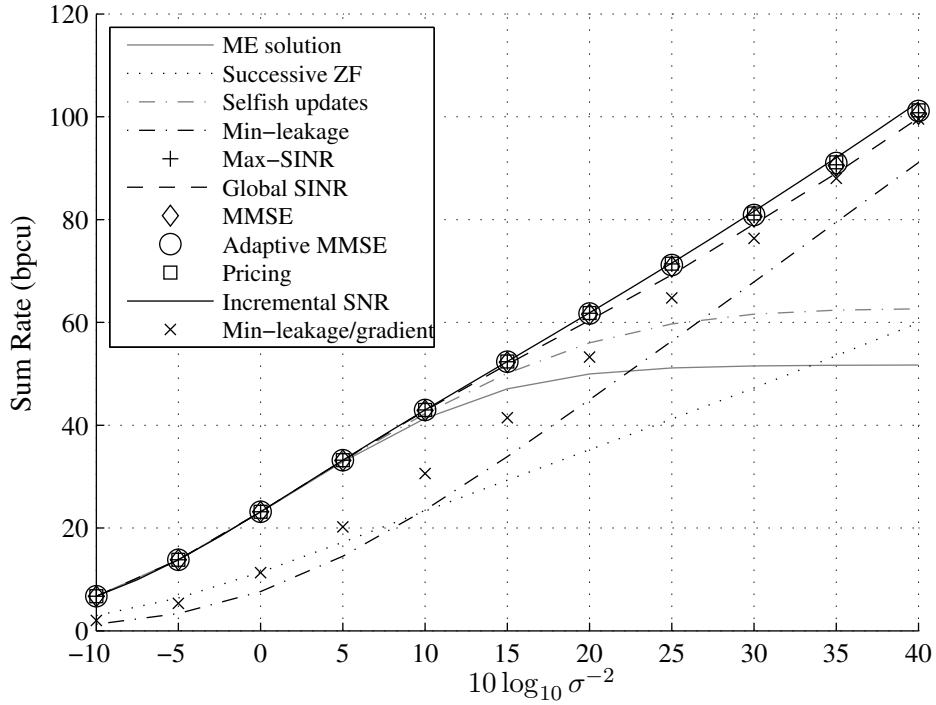


Figure 5.12: $K = 7$ users, $N = M = 4$ antennas at each transmitter and receiver, sum rate averaged over 100 channel realizations of a channel model where the variance of the cross channels is 0.01 and the variance of the direct channels is 1. The iterative algorithms are initialized with the ME solution.

Algorithm	–10 dB	15 dB	40 dB	–10 dB	15 dB	40 dB
Selfish updates	22 (0)	51 (10)	51.5 (9)	11 (0)	30 (0)	51.5 (14)
Min-leakage	1066.5 (0)	1066.5 (0)	1066.5 (0)	1270.5 (0)	1270.5 (0)	1270.5 (0)
Max-SINR	36.5 (0)	792.5 (0)	1009 (0)	12 (0)	69 (0)	1150.5 (0)
Global SINR	43 (0)	817 (0)	1081.5 (0)	13 (0)	78.5 (0)	1115.5 (1)
MMSE	52 (0)	1258.5 (0)	27 441 (1)	27 (0)	383.5 (0)	25 010 (0)
Adaptive MMSE	136.5 (0)	4558 (0)	29 690.5 (3)	28 (0)	1064.5 (0)	38 311.5 (4)
Pricing	51 (0)	1698 (0)	100 000 (66)	18 (0)	96.5 (0)	2982 (0)
Incremental SNR	39.5 (0)	2140.5 (0)	6242 (12)	12 (0)	292.5 (0)	3363.5 (0)
Min-leakage/gradient	1834.5 (0)	2236.5 (0)	2246.5 (0)	1771.5 (0)	2081 (0)	2195.5 (0)

Table 5.8: Median number of iterations until convergence for Figure 5.11 (left) and Figure 5.12 (right). In parentheses is the percentage of channel realizations for which the algorithm did not converge before reaching the maximum iteration number.

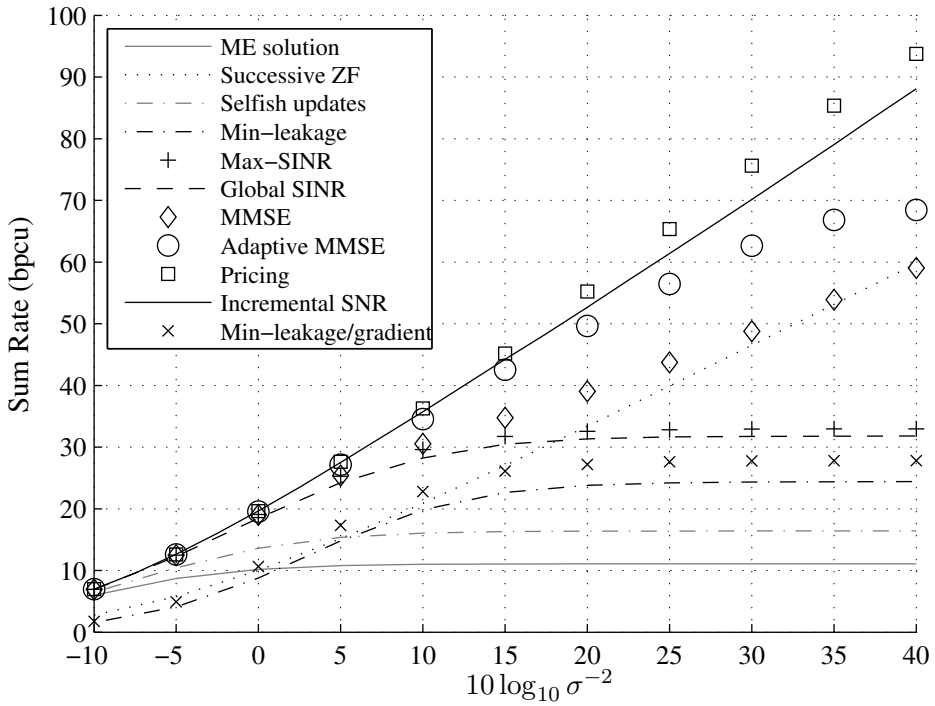


Figure 5.13: $K = 10$ users, $N = M = 4$ antennas at each transmitter and receiver, sum rate averaged over 100 channel realizations. The iterative algorithms are initialized with the ME solution.

5.6.3.3 Improper System: $K = 10$ Users and $N = M = 4$ Antennas

When all nodes in a MIMO interference network have four antennas, it is almost surely impossible to achieve a slope higher than $s = 7$ with linear single-stream strategies, regardless of the number of users K. If $K > 7$, users must therefore be deactivated in order to achieve the optimal slope. Consequently, as is shown in Figures 5.13 and 5.14, the sum rate of the algorithms that are not capable of power control, i. e., the min-leakage, max-SINR, global SINR, and min-leakage/gradient algorithms, cf. Table 5.4, saturates at high SNR.

The MMSE algorithms have the capability of reducing the transmit power of individual users and therefore achieve a higher sum rate at high SNR. The unweighted MMSE algorithm does not achieve the optimal slope, however, as can be seen in Figure 5.13, whereas the adaptively weighted MMSE algorithm has difficulty converging within 100 000 iterations. The best performance is achieved by the pricing-based algorithms, where, again, the incremental SNR technique significantly improves the convergence behavior; remarkably, however, the performance of the pricing algorithm with fixed SNR is slightly higher in Figure 5.13. A comparison of the absolute value of the sum rate achieved at high SNR in Figure 5.13 with that achieved in Figure 5.11 shows that the slope achieved for $K = 7$ users is higher than for $K = 10$ users; we conclude that even the best of the known iterative techniques is not able to reliably select the appropriate number of users for high-SNR sum-rate optimality. Instead, convergence to local optima that do not achieve the optimal slope can occur when the algorithm is initialized with “too many” active users.

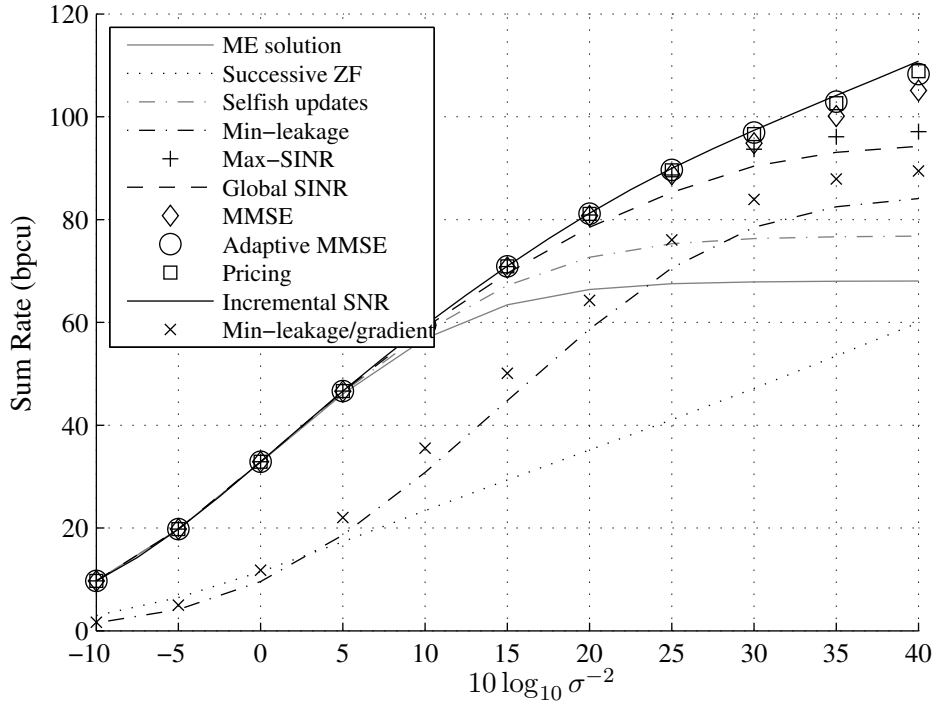


Figure 5.14: $K = 10$ users, $N = M = 4$ antennas at each transmitter and receiver, sum rate averaged over 100 channel realizations of a channel model where the variance of the cross channels is 0.01 and the variance of the direct channels is 1. The iterative algorithms are initialized with the ME solution.

Algorithm	–10 dB	15 dB	40 dB	–10 dB	15 dB	40 dB
Selfish updates	27 (0)	40 (3)	40 (2)	12 (0)	32 (0)	44 (3)
Min-leakage	268.5 (0)	268.5 (0)	268.5 (0)	279.5 (0)	279.5 (0)	279.5 (0)
Max-SINR	45 (0)	243.5 (0)	255.5 (0)	13 (0)	93 (0)	225 (0)
Global SINR	61 (0)	243 (0)	259.5 (0)	14 (0)	115.5 (0)	257.5 (0)
MMSE	63 (0)	628 (0)	16 398 (0)	28 (0)	395 (0)	8942.5 (0)
Adaptive MMSE	182.5 (0)	9731.5 (0)	100 000 (99)	29 (0)	1478.5 (0)	100 000 (79)
Pricing	68 (0)	1180 (0)	100 000 (52)	20 (0)	129.5 (0)	2135.5 (0)
Incremental SNR	52.5 (0)	2074 (0)	5724.5 (9)	14 (0)	390.5 (0)	2842.5 (0)
Min-leakage/gradient	905 (0)	1343.5 (0)	1340.5 (0)	893 (0)	1397 (0)	1450.5 (0)

Table 5.9: Median number of iterations until convergence for Figure 5.13 (left) and Figure 5.14 (right). In parentheses is the percentage of channel realizations for which the algorithm did not converge before reaching the maximum iteration number.

5.6.3.4 Discussion of the Numerical Properties of the Algorithms

The simulation results indicate that if it the system is fully loaded, the fixed-power algorithms are very useful for finding aligned solutions. The max-SINR and global SINR algorithms show the best convergence behavior with the former achieving a marginally higher sum-rate performance. Somewhat surprisingly, however, these algorithms exhibit convergence difficulties at high SNR in the underutilized setting, cf. Table 5.7. Also, it is not possible to adapt them to a different utility function: the update procedures are not based on maximizing the sum rate, but happen to yield good sum-rate performance; it is not possible to use the max-SINR or global SINR algorithms to maximize, e. g., the weighted sum rate.

The interference pricing technique offers the additional possibility of deactivating users and automatically finding a "good" number of users. This comes at a cost, however: not always is the optimal slope achieved and the convergence behavior at high SNR is generally problematic. Both issues can be somewhat mitigated with the incremental SNR method: with properly chosen SNR increments we observe excellent performance in all examined scenarios and far more reliable convergence behavior. Furthermore, it is straightforward to extend the pricing-based algorithms to other utility functions, such as the weighted sum rate.

The remaining algorithms seem to be less practical: while the MMSE-based methods also have the capability of deactivating users, they generally suffer from very slow convergence at high SNR. The performance of the min-leakage strategy clearly suffers from the fact that the direct power gain is not maximized and that at moderate and low SNR attempting to completely remove the interference is far from optimal; instead, the noise power must be taken into account in order to be able to provide a good tradeoff between interference cancelation and maximization of the direct power gain. The gradient extension of the min-leakage algorithm proposed in [94] leads to an improvement, but does not consistently perform as well as, e. g., the max-SINR algorithm; presumably, the performance of this method can be further improved by choosing different step size parameters, but it is questionable whether a universally suitable set of parameters can be found.

In a practical system, allowing up to 100 000 iterations would most likely be infeasible. In Figure 5.15, we show how the algorithms perform when we terminate the algorithms after 100 iterations for the fully loaded system with $K = 7$ users and $N = M = 4$ antennas per node with equally strong direct and cross channels. For the adaptive MMSE algorithm, we now update the weights after each iteration; for the incremental SNR strategy, we use the same series of noise powers as in the previous simulations (cf. Table 5.6), but permit only 5 iterations for each SNR value, so that at 40 dB 105 iterations will have been performed, while for an SNR of 30 dB only 85 iterations are allowed.

Clearly, interference alignment and the optimal slope cannot be achieved by any algorithm in so few iterations. The incremental SNR algorithm appears to come close; however, when we compare the absolute values of the sum rate to those in Figure 5.11, where the algorithms are run until convergence in the same scenario, it becomes clear that the optimal slope is not reached here. The max-SINR and global SINR algorithms also exhibit good performance for moderately high SNR. The min-leakage/gradient algorithm, on the other hand, performs surprisingly poorly: the choice of the step size parameters has the effect that, when the algorithm is terminated, the size of the gradient steps is still relatively large; in the words of [94], the algorithm is still in the "exploration phase" and has not yet reached the "convergence phase".

Finally, we briefly address the issues of parallel/sequential updates and the choice of the initialization: further experiments show that sequential updates do not yield a substantial improvement

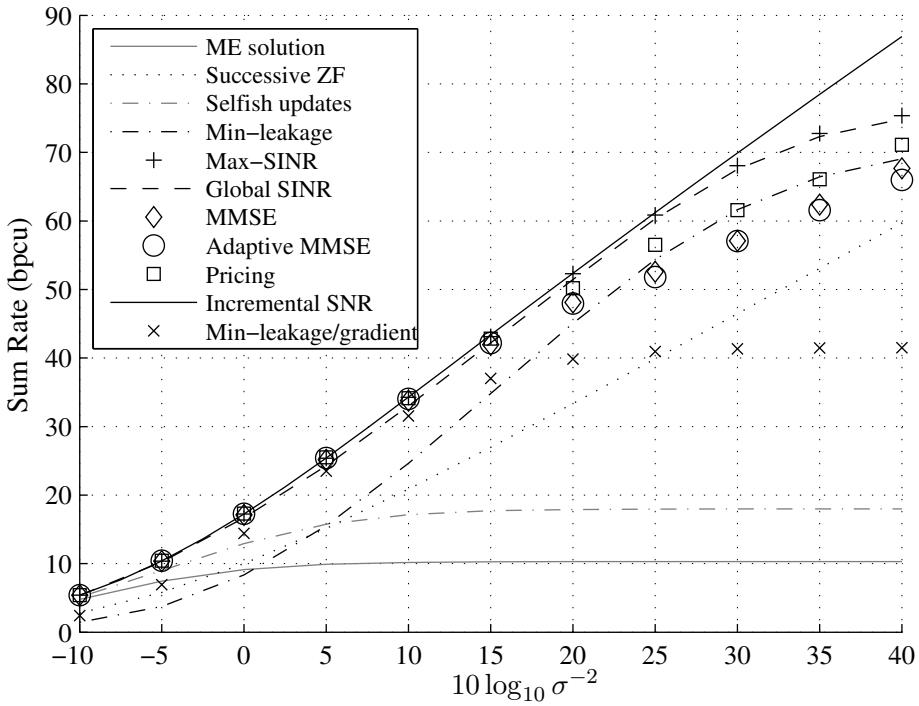


Figure 5.15: $K = 7$ users, $N = M = 4$ antennas at each transmitter and receiver, sum rate averaged over 100 channel realizations. The iterative algorithms are initialized with the ME solution and are allowed only 100 iterations.

for any of the examined algorithms, either in sum-rate performance or in convergence behavior; we omit the comparison, as it is not particularly instructive. Second, in all simulations of Section 5.6.3 we initialized the algorithms with the ME solution. In Figure 5.16, we exemplarily show for the max-SINR and min-leakage algorithms that this leads to a significant performance gain over random isotropically distributed initial strategies. We note that the other algorithms behave similarly.

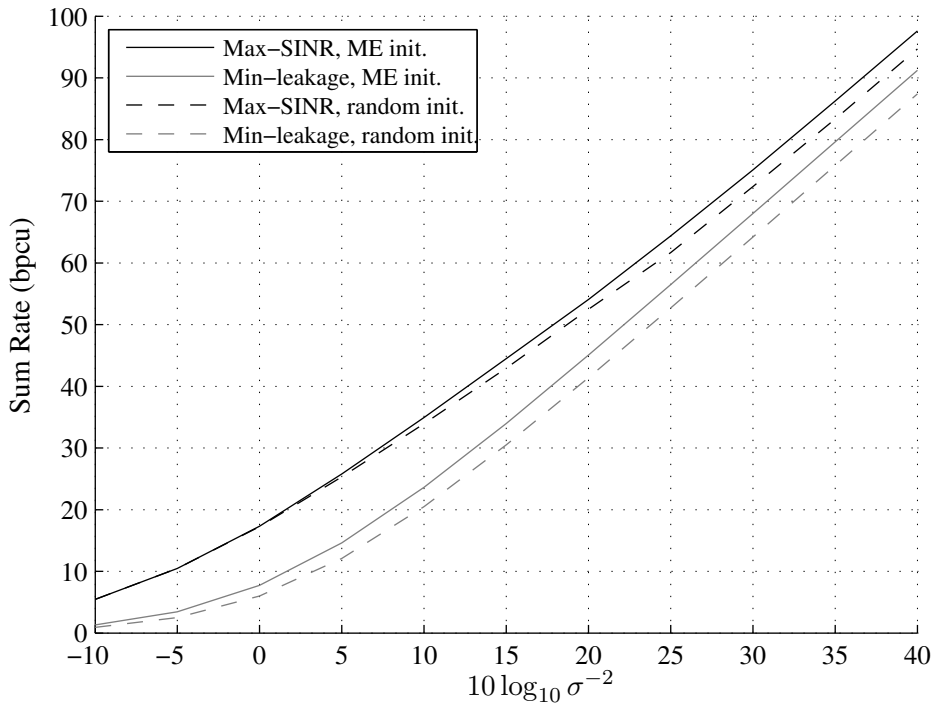


Figure 5.16: $K = 7$ users, $N = M = 4$ antennas at each transmitter and receiver, sum rate averaged over 100 channel realizations.

6. Conclusion

In this work we examined networks of multiple interfering transmitter-receiver pairs, where the transmitters and the receivers may have multiple antennas. Understanding this system model is important, e. g., for managing the inter-cell interference in cellular communication networks through base station cooperation. We gave an overview of the analytical results concerning the achievable performance of such networks and of the algorithms proposed to find strategies that come close to the achievable performance; we thoroughly compared the properties of the algorithms and evaluated their average performance in numerical experiments; we thereby showed that some algorithms are significantly more useful than others.

The overview of information theoretic results in Chapter 2 showed that the capacity limits for interference scenarios are in general difficult to analyze. As follows from the discussed achievable schemes and upper bounds for asymptotically high SNR, the interference between the users has the effect of cutting in half the degrees of freedom compared to the case where the users do not interfere with each other. At the same time, however, these results indicate that very complex coding schemes are necessary to achieve this asymptotically optimal behavior for more than three users. In particular, the transmitters must jointly design their codebooks depending on the current channel conditions, the receivers must know the codebooks used by all interfering transmitters, and the process of encoding the data for transmission and decoding the received data is considerably more difficult than "conventional" channel coding and decoding. It appears that currently such "DoF-achieving" schemes are mainly of academic interest and most likely not suitable for practical systems in the near future; therefore, suboptimal strategies are necessary that provide a good compromise between performance and simplicity.

Of the different possibilities to balance simplicity and performance, the one extreme is that the users do not cooperate at all, i. e., the transmitters do not coordinate their transmit strategies and the receivers do not attempt to decode the interference; this leads to the paradigm of *selfish* behavior. Our numerical simulations show that, as a rule of thumb, unless the power of the interference that cannot be removed by the linear receive filters is in the same order of magnitude or lower than the power of the background noise, selfish behavior leads to strongly suboptimal performance. Furthermore, when transmitters and receivers have multiple antennas, selfish behavior can lead to strategy oscillations.

This shows that cooperation among the users is in general desirable. In particular, the gain that can be achieved by cooperation grows with decreasing power of the background noise, i. e., with increasing SNR. The closest possible mode of cooperation allowed by our system model, and thus the other extreme of the tradeoff between performance and simplicity, requires the transmitters to jointly design their codesbooks and the receivers to decode the signal with the knowledge of all transmitters' codebooks. This type of cooperation is necessary to achieve the capacity limits of the interference channels, but is challenging to implement in a practical system, as noted above. A somewhat "looser" and less demanding mode of cooperation results from using the linear strategies discussed in Chapters 3–5. Here, conventional channel coding techniques are applied to individual scalar data streams; the transmitters cooperate only in power control and beamformer design and the receivers decode their individual data stream without having to decode any interfering signals.

While the linear strategies are less complex in terms of coding and information exchange among the users, their analysis is by no means simple: some of the encountered problems are convex optimization problems, such as the maximization of certain sum utilities in single-antenna networks or the balancing problem in the MISO case; the sum rate maximization problems, however, which are of particular interest, are non-convex and proven to be NP-hard. Also, we observe that the difficulty of the analysis increases when we allow the nodes to have more than one antenna: for SISO systems, we can give a sufficient condition for convexity of the achievable utility region as well as necessary and sufficient conditions for Pareto optimality of a strategy; also, the feasibility and SINR balancing problems can be solved in a straightforward way and the high-SNR sum-rate optimal strategy is trivial.

In the MISO case, on the other hand, even utility functions that guarantee convex utility regions for single-antenna systems may lead to a non-convex utility region. Also, the Pareto boundary can only be characterized with necessary conditions, which are in general not sufficient. The feasibility problem cannot be solved in closed form, but can be reformulated so that it can be solved with general purpose optimization tools; and the sum-rate optimal strategy at high SNR can be expressed in closed form for settings in which zero-forcing is possible.

For general MIMO intereference networks, finally, the structure of the underlying problems is so complex that analysis is limited to the regime of asymptotically low or high SNR for the rate utility; these asymptotic results are not valid for all channels, but only for *almost all* channel realizations of a random channel model. The high-SNR optimal strategies, which have the property that the interference is *aligned,* in general cannot be expressed in closed form and numerically finding the global high-SNR optimum appears to be computationally very demanding and only feasible for some small scenarios: the only known method is to determine and compare all aligned solutions, the number of which grows extremely fast with the system dimensions.

While the high-SNR analysis in the literature focuses on the characterization of the asymptotic slope, in this work we presented a probabilistic analysis of the *offset* that makes use of large-system approximations and extreme statistics. Our analysis allows us to understand the average performance gain that can be achieved by trying out many different aligned solutions in fully loaded systems and also indicates that for larger systems the successive zero-forcing strategy, which achieves only half of the optimal slope, can outperform aligned strategies even at moderately high SNR.

Numerous algorithms have been proposed for determining good suboptimal strategies in terms of sum rate. These algorithms can be implemented in a distributed manner, i. e., the individual users iteratively compute strategy updates using locally available information and exchange a limited amount of information with the other users after each iteration. We can divide the algorithms into two classes: the first type of algorithm adjusts only the direction of the beamformers and every user transmits with the highest possible power; the second type of algorithm optimizes over the beamformer direction and the power.

The most convincing of the algorithms without power control according to our numerical simulations is the max-SINR algorithm, which achieves excellent sum-rate performance in many scenarios and generally converges quickly and reliably. The strategy of maximizing the virtual SINR, which is the MISO equivalent of the max-SINR technique and also performs very well in systems where zero-forcing is possible, does not even require iterative updates. Since the algorithms in this class are not capable of deactivating individual users when it is in the common interest, they rely on being initialized with the correct number of users, i. e., they depend on the system being *proper* at high SNR. In settings where zero-interference solutions are not possible it may be necessary to precede the fixed-power algorithms with a user selection procedure. It is not clear, however,

how such a user selection would determine which and how many users should be active: while it is known that at high SNR the system should be proper, at moderate SNR it could be beneficial to have more active users. Another drawback of the fixed-power algorithms is that they are not designed to achieve local sum-rate optimality. This does not imply that their sum-rate performance is lower, but it does mean that they cannot be adapted to other related optimization objectives, such as the weighted sum rate. Maximizing the weighted sum rate can be of significance when, e. g., the users are prioritized according to scheduling considerations. Finally, fixed-power algorithms are not applicable to single-antenna interference networks.

Algorithms in the second class, on the other hand, can be specifically designed to at least locally maximize a certain utility. The most straightforward method to achieve this is to perform a gradient ascent on the sum utility function. However, numerical evaluation of the proposed gradient-based algorithms shows that choosing a universally suitable step size is an open problem and that a poor choice of the step size can greatly affect the convergence behavior. The interference pricing method, on the other hand, does not require any parameters to be adjusted; it shows good performance across a wide range of scenarios and is applicable in SISO, MISO, and MIMO networks, regardless of whether the system is proper or whether users must be deactivated. The interference pricing algorithm is therefore able to inherently perform the necessary user selection. The algorithm does have difficulties in fully loaded MIMO interference networks at high SNR, though; this appears to be a consequence of the properties of the sum rate utility, which has many local optima that are not aligned solutions. The heuristic of gradually increasing the SNR appears to somewhat relieve this problem: in nearly all examined MIMO scenarios the pricing algorithm with incremental SNR achieves the best performance among all strategies. The number of iterations, however, in general is higher than for the fixed-power algorithms.

We conclude that, due to the difficulty of the sum rate maximization problem in MIMO interference networks, there is no “one size fits all” algorithm. Also, it appears that even determining suboptimal strategies with any of the examined algorithms requires a rather large computational effort if we take into account the high iteration numbers. In practical interference networks the goal consequently cannot be to achieve the optimum, but rather to improve an initial strategy as much as possible within a limited number of iterations. Therefore, the results of Figure 5.15 are particularly promising: they indicate that the benefits that can be realistically achieved with the techniques presented in this work are considerable.

It remains to extend the results of this work to the general multi-beam case. Many of the known fixed-power algorithms are discussed in the literature for more than one stream per user, but again they rely on being initialized with a good configuration of streams per user. The challenge to be met by the algorithms with power control is to reliably converge to close-to-optimal strategies even when the optimal stream configuration is not known. The interference pricing technique may turn out to be a valuable tool for this problem as well.

Appendix

A1. Derivation of the Sato Bound

In Section 2.3.2.2 we wish to solve the optimization problem

$$\bar{r} = \arg\min_{r} \ \log\det\left(\mathbf{I} + \boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}^{-1}\boldsymbol{H}\right) \qquad \text{s.t.:} \quad |r| \leq 1 \tag{A1}$$

with

$$\boldsymbol{R} = \sigma^2 \cdot \begin{bmatrix} 1 & r \\ r^* & 1 \end{bmatrix} \tag{A2}$$

and a full rank matrix $\boldsymbol{H} \in \mathbb{C}^{2\times 2}$. We define

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{12}^* & a_{22} \end{bmatrix} = \sigma^{-2}\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}. \tag{A3}$$

Since the logarithm is increasing, we can equivalently find the minimizer of the above determinant, which we simplify as

$$\begin{aligned}
&\det\left(\mathbf{I} + \boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}^{-1}\boldsymbol{H}\right) = \det\left(\mathbf{I} + \boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}^{-1}\right) \\
&= \det\left(\mathbf{I} + \frac{1}{1-|r|^2}\begin{bmatrix} a_{11} & a_{12} \\ a_{12}^* & a_{22} \end{bmatrix} \cdot \begin{bmatrix} 1 & -r \\ -r^* & 1 \end{bmatrix}\right) \\
&= \frac{1}{(1-|r|^2)^2}\det\left(\begin{bmatrix} a_{11} - r^*a_{12} + 1 - |r|^2 & a_{12} - ra_{11} \\ a_{12}^* - r^*a_{22} & a_{22} - ra_{12}^* + 1 - |r|^2 \end{bmatrix}\right) \\
&= \frac{(a_{11} - r^*a_{12} + 1 - |r|^2)(a_{22} - ra_{12}^* + 1 - |r|^2) - (a_{12} - ra_{11})(a_{12}^* - r^*a_{22})}{(1-|r|^2)^2} \\
&= \frac{(1-|r|^2)(a_{11} + a_{22} + a_{11}a_{22} - |a_{12}|^2 - a_{12}r^* - a_{12}^*r + 1 - |r|^2)}{(1-|r|^2)^2} \\
&= 1 + \frac{d_1 - 2\,\mathrm{Re}(d_2^* r)}{1-|r|^2}
\end{aligned} \tag{A4}$$

with

$$d_1 = a_{11} + a_{22} + a_{11}a_{22} - |a_{12}|^2 \tag{A5}$$

$$d_2 = a_{12} \tag{A6}$$

cf. (2.66) and (2.67). We find candidates for the minimum by taking the derivative with respect to r^*:

$$\frac{\partial}{\partial r^*}\left(\frac{d_1 - 2\,\mathrm{Re}(d_2^* r)}{1-|r|^2}\right) = \frac{-(1-|r|^2)d_2 + (d_1 - 2\,\mathrm{Re}(d_2^* r))r}{(1-|r|^2)^2} = \frac{-d_2^* r^2 + d_1 r - d_2}{(1-|r|^2)^2}. \tag{A7}$$

The zeros of the numerator are

$$\bar{r}_{1,2} = \frac{d_1 \pm \sqrt{d_1^2 - 4|d_2|^2}}{2d_2^*}. \tag{A8}$$

Note that $d_1 > a_{11} + a_{22}$ and $|d_2| < \sqrt{a_{11}a_{22}}$, since $\boldsymbol{A}$ is positive definite. Therefore, $d_1 - 2|d_2| > a_{11} + a_{22} - 2\sqrt{a_{11}a_{22}} = (\sqrt{a_{11}} - \sqrt{a_{22}})^2 \geq 0$, and consequently $d_1 > 2|d_2|$. This means that the numerator in (A8) is always real-valued. Furthermore,

$$d_1 - 2|d_2| \leq \sqrt{d_1^2 - 4|d_2|^2} = \sqrt{(d_1 - 2|d_2|)(d_1 + 2|d_2|)} \leq d_1 + 2|d_2| \tag{A9}$$

and it can be shown that $|\bar{r}_1| > 1$ (plus sign in (A8)) and $|\bar{r}_2| < 1$ (minus sign in (A8)). Since there is exactly one candidate for an extreme value that fulfills the constraint and the cost function is infinite on the boundary of the constraint set, we have found that the minimizer is $\bar{r}_2$. The expression in (2.68) is obtained by plugging (2.65) into (A4).

A2. Proof of Proposition 3.2

Let us assume that we are given an SINR K-tuple $(\gamma_1, \ldots, \gamma_K)$ and would like to determine a power allocation $(p_1, \ldots, p_K)$ that achieves this SINR K-tuple. To begin with, it is clear that if any $\gamma_k = 0$, the corresponding power p_k must also be zero. It is then possible to "remove" this user from the system and solve the equations for the remaining non-zero SINRs.

For ease of notation, we assume in the following that the zero-SINR users have already been removed and that all SINRs γ_k with $k \in \{1, \ldots, K\}$ are non-zero. From (3.5) we obtain the system of equations

$$\frac{|h_{11}|^2}{\gamma_1} p_1 - |h_{12}|^2 p_2 - \ldots - |h_{1K}|^2 p_K = \sigma^2 \tag{A10}$$

$$-|h_{21}|^2 p_1 + \frac{|h_{22}|^2}{\gamma_2} p_2 - \ldots - |h_{2K}|^2 p_K = \sigma^2 \tag{A11}$$

$$\vdots$$

$$-|h_{K1}|^2 p_1 - |h_{K2}|^2 p_2 - \ldots + \frac{|h_{KK}|^2}{\gamma_K} p_K = \sigma^2 \tag{A12}$$

which we can express in matrix-vector notation as

$$\boldsymbol{A}\boldsymbol{p} = \sigma^2 \mathbf{1} \tag{A13}$$

where the matrix

$$\boldsymbol{A} = \begin{bmatrix} \frac{|h_{11}|^2}{\gamma_1} & -|h_{12}|^2 & \cdots & -|h_{1K}|^2 \\ -|h_{21}|^2 & \frac{|h_{22}|^2}{\gamma_2} & \ldots & -|h_{2K}|^2 \\ \vdots & \vdots & \ddots & \vdots \\ -|h_{K1}|^2 & -|h_{K2}|^2 & \ldots & \frac{|h_{KK}|^2}{\gamma_K} \end{bmatrix} \tag{A14}$$

depends on the given SINR K-tuple $(\gamma_1, \ldots, \gamma_K)$ and the channel coefficients, and the vector $\boldsymbol{p} = [p_1, \ldots, p_K]^{\mathrm{T}}$ contains the unknown powers. Therefore, we can state that the SINR K-tuple is in the SINR region if and only if there exists a vector $\boldsymbol{p}$ that fulfills (A13) and of which each element is in the interval $[0, 1]$.

Next, let us assume that the SINR K-tuple $(\gamma_1, \ldots, \gamma_K)$ is indeed feasible and that it results from the power allocation $(p_1, \ldots, p_K)$, where each power is strictly greater than zero. Then, by

inserting (3.5) into (A14), we obtain

$$\boldsymbol{A} = \begin{bmatrix} \left(\sum_{j\neq 1}|h_{1j}|^2 p_j + \sigma^2\right)/p_1 & -|h_{12}|^2 & \cdots & -|h_{1K}|^2 \\ -|h_{21}|^2 & \left(\sum_{j\neq 2}|h_{2j}|^2 p_j + \sigma^2\right)/p_2 & \dots & -|h_{2K}|^2 \\ \vdots & \vdots & \ddots & \vdots \\ -|h_{K1}|^2 & -|h_{K2}|^2 & \dots & \left(\sum_{j\neq K}|h_{Kj}|^2 p_j + \sigma^2\right)/p_K \end{bmatrix}. \tag{A15}$$

We can rewrite $\boldsymbol{A}$ as the product of two matrices

$$\boldsymbol{A} = \boldsymbol{B}\boldsymbol{P}^{-1} \tag{A16}$$

where

$$\boldsymbol{B} = \begin{bmatrix} \sum_{j\neq 1}|h_{1j}|^2 p_j + \sigma^2 & -|h_{12}|^2 p_2 & \cdots & -|h_{1K}|^2 p_K \\ -|h_{21}|^2 p_1 & \sum_{j\neq 2}|h_{2j}|^2 p_j + \sigma^2 & \dots & -|h_{2K}|^2 p_K \\ \vdots & \vdots & \ddots & \vdots \\ -|h_{K1}|^2 p_1 & -|h_{K2}|^2 p_2 & \dots & \sum_{j\neq K}|h_{Kj}|^2 p_j + \sigma^2 \end{bmatrix} \tag{A17}$$

and

$$\boldsymbol{P} = \begin{bmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_K \end{bmatrix}. \tag{A18}$$

The matrix $\boldsymbol{B}$ has positive row sums and only non-positive off-diagonal elements and is therefore an *M-matrix* [45]. Thus, $\boldsymbol{B}^{-1}$ exists and has only non-negative elements. Consequently, for every feasible non-zero SINR K-tuple, $\boldsymbol{A}^{-1}$ exists and has only non-negative elements.

Now, let us assume that $\boldsymbol{A}$ results from a feasible non-zero SINR K-tuple (or, equivalently, a feasible power allocation vector $\boldsymbol{p}$ with elements strictly greater than zero) and let us examine another SINR K-tuple $(\bar{\gamma}_1, \dots, \bar{\gamma}_K)$ with $0 < \bar{\gamma}_k \leq \gamma_k$ for all $k \in \{1, \dots, K\}$. A power allocation vector $\bar{\boldsymbol{p}}$ resulting in this SINR K-tuple must fulfill

$$\bar{\boldsymbol{A}}\bar{\boldsymbol{p}} = \sigma^2 \mathbf{1} \tag{A19}$$

where $\bar{\boldsymbol{A}}$ is obtained by replacing $\gamma_1, \dots, \gamma_K$ with $\bar{\gamma}_1, \dots, \bar{\gamma}_K$ in (A14). Since all $\bar{\gamma}_k$ are smaller than or equal to the respective γ_k, we can express $\bar{\boldsymbol{A}}$ as the sum of $\boldsymbol{A}$ and a diagonal matrix $\boldsymbol{D}$ with non-negative elements, i. e.,

$$\bar{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{D} = (\boldsymbol{B} + \boldsymbol{D}\boldsymbol{P})\boldsymbol{P}^{-1}. \tag{A20}$$

Since $\boldsymbol{B} + \boldsymbol{D}\boldsymbol{P}$ has positive row sums and only non-positive off-diagonal elements, $\bar{\boldsymbol{A}}^{-1}$ exists and has only non-negative elements. Consequently, $\bar{\boldsymbol{p}}$ is unique and has only non-negative elements. Furthermore,

$$\bar{\boldsymbol{p}} = \sigma^2 (\boldsymbol{A} + \boldsymbol{D})^{-1} \mathbf{1} = \sigma^2 \boldsymbol{A}^{-1}\mathbf{1} - \sigma^2 \boldsymbol{A}^{-1}\boldsymbol{D} (\boldsymbol{A} + \boldsymbol{D})^{-1} \mathbf{1} = \boldsymbol{p} - \sigma^2 \boldsymbol{A}^{-1}\boldsymbol{D}\bar{\boldsymbol{A}}^{-1}\mathbf{1}. \tag{A21}$$

Since $\boldsymbol{A}^{-1}$, $\boldsymbol{D}$, and $\bar{\boldsymbol{A}}^{-1}$ have only non-negative elements, each element of the vector $\bar{\boldsymbol{p}}$ is less than or equal to the corresponding element of $\boldsymbol{p}$, i. e., $0 < \bar{p}_k \leq p_k$ holds for all $k \in \{1, \dots, K\}$ and $\bar{\boldsymbol{p}}$ is also a feasible power allocation vector.

To show feasibility of an SINR K-tuple $(\bar{\gamma}_1, \ldots, \bar{\gamma}_K)$ that also contains zero elements, we proceed in two steps. First, we examine the SINR K-tuple $(\hat{\gamma}_1, \ldots, \hat{\gamma}_K)$ that results from the power allocation vector $\hat{\boldsymbol{p}}$. We define the kth element of $\hat{\boldsymbol{p}}$ to be either p_k if $\bar{\gamma}_k \neq 0$ or zero if $\bar{\gamma}_k = 0$, i. e., we switch off the users for which $\bar{\gamma}_k$ is zero and leave the rest of the power allocation unchanged. Clearly, for the users with non-zero power $\hat{p}_k$, $\hat{\gamma}_k \geq \gamma_k$. In the second step, we remove the users for which $\hat{\gamma}_k = \bar{\gamma}_k = 0$ from the system. We are left with a reduced system in which $0 < \bar{\gamma}_k \leq \hat{\gamma}_k$ for all remaining users and can apply the same arguments as above to show that the K-tuple $(\bar{\gamma}_1, \ldots, \bar{\gamma}_K)$ is feasible.

As a result, we can state that if an SINR K-tuple $(\gamma_1, \ldots, \gamma_K)$ is feasible, any SINR K-tuple which is element-wise smaller or equal is also feasible. Thus, the SINR region is comprehensive and the proof is complete.

A3. Proof of Proposition 3.3

Let us assume w. l. o. g. that $h_{21} \neq 0$ and examine the power allocations $(\hat{p}_1 = 1, \hat{p}_2 = 1, \ldots, \hat{p}_K = 1)$, $(\check{p}_1 = 0, \check{p}_2 = 1, \ldots, \check{p}_K = 1)$, and $(p_1 = \frac{1}{2}, p_2 = 1, \ldots, p_K = 1)$ as well as the resulting feasible SINR K-tuples $(\hat{\gamma}_1, \ldots, \hat{\gamma}_K)$, $(\check{\gamma}_1, \ldots, \check{\gamma}_K)$, and $(\gamma_1, \ldots, \gamma_K)$. We will then examine a certain convex combination of the former two SINR K-tuples and show that it is infeasible.

The resulting SINRs are

$$\hat{\gamma}_k = \frac{|h_{kk}|^2}{\sum\limits_{j \neq k} |h_{kj}|^2 + \sigma^2} \qquad \forall k \in \{1, \ldots, K\} \tag{A22}$$

$$\check{\gamma}_1 = 0 \qquad \text{and} \qquad \check{\gamma}_k = \frac{|h_{kk}|^2}{\sum\limits_{j \neq k} |h_{kj}|^2 - |h_{k1}|^2 + \sigma^2} \qquad \forall k \in \{2, \ldots, K\} \tag{A23}$$

$$\gamma_1 = \frac{\frac{1}{2}|h_{11}|^2}{\sum\limits_{j \neq 1} |h_{1j}|^2 + \sigma^2} \qquad \text{and} \qquad \gamma_k = \frac{|h_{kk}|^2}{\sum\limits_{j \neq k} |h_{kj}|^2 - \frac{1}{2}|h_{k1}|^2 + \sigma^2} \qquad \forall k \in \{2, \ldots, K\}. \tag{A24}$$

We observe that

$$\gamma_1 = \frac{1}{2}\hat{\gamma}_1 \qquad \text{and} \qquad \gamma_k = 2\left(\hat{\gamma}_k^{-1} + \check{\gamma}_k^{-1}\right)^{-1} \qquad \forall k \in \{2, \ldots, K\}. \tag{A25}$$

We define the K-tuple $(\bar{\gamma}_1, \ldots, \bar{\gamma}_K)$ to be the convex combination of $(\hat{\gamma}_1, \ldots, \hat{\gamma}_K)$ and $(\check{\gamma}_1, \ldots, \check{\gamma}_K)$ with $\alpha = \frac{1}{2}$, i. e.,

$$\bar{\gamma}_k = \frac{\hat{\gamma}_k + \check{\gamma}_k}{2} \qquad \forall k \in \{1, \ldots, K\}. \tag{A26}$$

Since the harmonic mean of two real non-negative numbers is always smaller than or equal to the arithmetic mean with equality only if the two numbers are identical, we can state that

$$\gamma_1 = \bar{\gamma}_1, \qquad \gamma_2 < \bar{\gamma}_2, \qquad \text{and} \qquad \gamma_k \leq \bar{\gamma}_k \qquad \forall k \in \{3, \ldots, K\}. \tag{A27}$$

Next, we examine whether $(\bar{\gamma}_1, \ldots, \bar{\gamma}_K)$ is feasible. Note that any change of power allocation from $(p_1 = \frac{1}{2}, p_2 = 1, \ldots, p_K = 1)$ that does not violate the power constraints will either

1) include a decrease of one or more powers or

2) consist of only increasing p_1.

In both cases, due to the properties of the SINR, at least one SINR will decrease compared to $(\gamma_1, \ldots, \gamma_K)$. Therefore $(\bar{\gamma}_1, \ldots, \bar{\gamma}_K)$ is not feasible and the region is non-convex.

When all cross channels are zero, the SINR region is simply an orthotope (or hyperrectangle) confined by the inequalities

$$0 \leq \gamma_k \leq \frac{|h_{kk}|^2}{\sigma^2} \qquad \forall k \in \{1, \ldots, K\} \tag{A28}$$

and therefore convex.

A4. Proof of Proposition 3.6

The optimization problem (3.47) can be stated equivalently as

$$\max_{u_1,\ldots,u_K} \sum_{k=1}^{K} u_k \qquad \text{s. t.:} \quad (u_1, \ldots, u_K) \text{ in the utility region} \tag{A29}$$

where two optimization problems are *equivalent,* if from the solution to one of the two problems the solution to the other problem can be found in a straightforward way.

The objective function of (A29) is linear in the optimization variables $u_1, \ldots, u_K$ and the feasible set is convex, therefore (A29) is a *concave maximization problem* and the KKT conditions of (A29) are sufficient for global optimality [36]. In the following we will show that the KKT conditions of (A29) are equivalent to the KKT conditions of (3.47) in the sense that a power allocation or corresponding utility K-tuple that fulfills the KKT conditions of one of the two problems also fulfills the KKT conditions of the other problem.

In order to formulate the KKT conditions of (A29), we first explicitly state the constraints using the functions $f_k(u_1, \ldots, u_K)$ which map the utility K-tuple to a power allocation that achieves these utilities, i. e.,

$$f_k(u_1(\gamma_1), \ldots, u_K(\gamma_K)) = p_k \qquad \forall k \in \{1, \ldots, K\}. \tag{A30}$$

We note that, as is discussed in Section 3.4.3, the mapping from the SINR K-tuple to the power allocation is unique and well-defined for all feasible SINR K-tuples. Therefore, the mapping from utilities to powers is also unique and the functions $f_k(\cdot)$ are well-defined for all feasible utilities.

The optimization problem (A29) with explicitly stated constraints is

$$\max_{u_1,\ldots,u_K} \sum_{k=1}^{K} u_k \qquad \text{s. t.:} \quad 0 \leq f_k(u_1, \ldots, u_K) \leq 1 \quad \forall k \in \{1, \ldots, K\}. \tag{A31}$$

With the Lagrangian multipliers $\mu_1, \ldots, \mu_K$ and $\nu_1, \ldots, \nu_K$, the KKT conditions are

$$1 + \sum_{j=1}^{K} \frac{\partial f_j}{\partial u_k}(\mu_j - \nu_j) = 0 \qquad \forall k \in \{1, \ldots, K\} \tag{A32}$$

$$\mu_k \geq 0 \qquad \forall k \in \{1, \ldots, K\} \tag{A33}$$

$$\nu_k \geq 0 \qquad \forall k \in \{1, \ldots, K\} \tag{A34}$$

$$\mu_k f_k(u_1, \ldots, u_K) = 0 \qquad \forall k \in \{1, \ldots, K\} \tag{A35}$$

$$\nu_k(1 - f_k(u_1, \ldots, u_K)) = 0 \qquad \forall k \in \{1, \ldots, K\} \tag{A36}$$

$$0 \leq f_k(u_1, \ldots, u_K) \leq 1 \qquad \forall k \in \{1, \ldots, K\}. \tag{A37}$$

Clearly, (A33)–(A37) are equivalent to (3.49)–(3.53). In the following we will show that, additionally, (A32) is equivalent to (3.48).

To begin with, we define the Jacobi matrices

$$\boldsymbol{J}_f = \begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \cdots & \frac{\partial f_K}{\partial u_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial u_K} & \cdots & \frac{\partial f_K}{\partial u_K} \end{bmatrix} \quad \text{and} \quad \boldsymbol{J}_u = \begin{bmatrix} \frac{\partial u_1}{\partial p_1} & \cdots & \frac{\partial u_K}{\partial p_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial u_1}{\partial p_K} & \cdots & \frac{\partial u_K}{\partial p_K} \end{bmatrix}. \tag{A38}$$

With the vectors of Lagrangian multipliers $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_K]^{\mathrm{T}}$ and $\boldsymbol{\nu} = [\nu_1, \ldots, \nu_K]^{\mathrm{T}}$, we express (A32) in matrix-vector form as

$$\mathbf{1} + \boldsymbol{J}_f(\boldsymbol{\mu} - \boldsymbol{\nu}) = \mathbf{0} \tag{A39}$$

and (3.48) as

$$\boldsymbol{J}_u\mathbf{1} + \boldsymbol{\mu} - \boldsymbol{\nu} = \mathbf{0}. \tag{A40}$$

We observe that if $\boldsymbol{J}_u$ is invertible and $\boldsymbol{J}_f = \boldsymbol{J}_u^{-1}$, the two conditions are equivalent. From (A30) it is clear that the derivative of f_k w. r. t. p_j is one if $j = k$ and zero otherwise, i. e.,

$$\frac{\partial f_k}{\partial p_j} = \boldsymbol{e}_j^{\mathrm{T}}\boldsymbol{e}_k. \tag{A41}$$

By applying the chain rule to the left-hand side of (A30), we can state the same derivative as

$$\frac{\partial f_k}{\partial p_j} = \sum_{i=1}^{K} \frac{\partial f_k}{\partial u_i} \cdot \frac{\partial u_i}{\partial p_j} = \boldsymbol{e}_j^{\mathrm{T}}\boldsymbol{J}_u\boldsymbol{J}_f\boldsymbol{e}_k. \tag{A42}$$

Since this holds for any pair (j, k), it follows that

$$\boldsymbol{J}_u\boldsymbol{J}_f = \mathbf{I}. \tag{A43}$$

It remains to show that the Jacobi matrices are always invertible. To this end, we define the Jacobi matrix of the SINRs as

$$\boldsymbol{J}_\gamma = \begin{bmatrix} \frac{\partial \gamma_1}{\partial p_1} & \cdots & \frac{\partial \gamma_K}{\partial p_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial \gamma_1}{\partial p_K} & \cdots & \frac{\partial \gamma_K}{\partial p_K} \end{bmatrix} \tag{A44}$$

and note that

$$\boldsymbol{J}_u = \boldsymbol{J}_\gamma \cdot \begin{bmatrix} u_1'(\gamma_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & u_K'(\gamma_K) \end{bmatrix}. \tag{A45}$$

Since the utility functions are strictly increasing in the SINRs, invertibility of $\boldsymbol{J}_\gamma$ implies invertibility of $\boldsymbol{J}_u$. With (3.5), the elements of $\boldsymbol{J}_\gamma$ are

$$\frac{\partial \gamma_k}{\partial p_k} = \frac{|h_{kk}|^2}{\sum_{i \neq k} |h_{ki}|^2 p_i + \sigma^2} \quad \text{and} \quad \frac{\partial \gamma_k}{\partial p_j} = -\frac{|h_{kj}|^2 |h_{kk}|^2 p_k}{\left(\sum_{i \neq k} |h_{ki}|^2 p_i + \sigma^2\right)^2} \quad \forall j \neq k. \tag{A46}$$

We can now express the SINR Jacobi matrix as

$$\boldsymbol{J}_\gamma = \boldsymbol{W} \cdot \boldsymbol{T} \tag{A47}$$

where

$$\boldsymbol{W} = \begin{bmatrix} \sum_{i\neq 1}|h_{1i}|^2 p_i + \sigma^2 & -|h_{21}|^2 p_2 & \cdots & -|h_{K1}|^2 p_K \\ -|h_{12}|^2 p_1 & \sum_{i\neq 2}|h_{2i}|^2 p_i + \sigma^2 & \cdots & -|h_{K2}|^2 p_K \\ \vdots & \vdots & \ddots & \vdots \\ -|h_{1K}|^2 p_1 & -|h_{2K}|^2 p_2 & \cdots & \sum_{i\neq K}|h_{Ki}|^2 p_i + \sigma^2 \end{bmatrix} \tag{A48}$$

and $\boldsymbol{T}$ is an invertible diagonal matrix of which the kth element is $|h_{kk}|^2 \big/ \left(\sum_{i\neq k}|h_{ki}|^2 p_i + \sigma^2\right)^2$. Next, we define $\boldsymbol{P}_\varepsilon$ as an invertible diagonal matrix of which the kth element is $p_k + \varepsilon$, where $\varepsilon > 0$. If we denote the element in row j and column k of $\boldsymbol{P}_\varepsilon \cdot \boldsymbol{W}$ as $[\boldsymbol{P}_\varepsilon \boldsymbol{W}]_{j,k}$, we have

$$[\boldsymbol{P}_\varepsilon \boldsymbol{W}]_{j,k} = \begin{cases} \left(\sum_{i\neq k}|h_{ki}|^2 p_i + \sigma^2\right) \cdot (p_k + \varepsilon) & \text{for } j = k \\ -|h_{kj}|^2 p_k (p_j + \varepsilon) & \text{for } j \neq k. \end{cases} \tag{A49}$$

The sum of the elements in the kth column therefore is

$$\sum_{j=1}^{K} [\boldsymbol{P}_\varepsilon \boldsymbol{W}]_{j,k} = \sum_{j\neq k} |h_{kj}|^2 (p_j - p_k)\varepsilon + \sigma^2 \varepsilon + \sigma^2 p_k. \tag{A50}$$

If $p_k = 0$, this sum is positive regardless of ε; if $p_k > 0$, it is positive for sufficiently small ε. Therefore, for sufficiently small ε, $\boldsymbol{P}_\varepsilon \boldsymbol{W}$ has positive column sums and is an *M-matrix* [45], which is guaranteed to be non-singular. It follows that $\boldsymbol{W}$ is non-singular, $\boldsymbol{J}_\gamma$ is non-singular, and finally that $\boldsymbol{J}_u$ is non-singular. With (A43) we have $\boldsymbol{J}_f = \boldsymbol{J}_u^{-1}$ and all KKT conditions of the problems (3.47) and (A31) are equivalent.

A5. Derivation of the MISO Zero-Forcing Beamformer

In Section 4.3.2, the optimization problem

$$\boldsymbol{v}_k^{\mathrm{ZF}} = \arg\max_{\boldsymbol{v}_k} |\boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k|^2 \qquad \text{s. t.:} \quad \boldsymbol{h}_{jk}^{\mathrm{T}} \boldsymbol{v}_k = 0 \quad \text{for all users } j \neq k \text{ in the active set} \quad \text{and} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1 \tag{A51}$$

must be solved. With the matrix $\boldsymbol{H}_k$ that contains the stacked row vectors $\boldsymbol{h}_{jk}^{\mathrm{T}}$ for all users $j \neq k$ that are in the active set, the optimization problem can be written as

$$\boldsymbol{v}_k^{\mathrm{ZF}} = \arg\max_{\boldsymbol{v}_k} |\boldsymbol{h}_{kk}^{\mathrm{T}} \boldsymbol{v}_k|^2 \qquad \text{s. t.:} \quad \boldsymbol{H}_k \boldsymbol{v}_k = \boldsymbol{0} \quad \text{and} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1. \tag{A52}$$

Also, the projection matrix $\boldsymbol{\Pi}_k$ is defined as

$$\boldsymbol{\Pi}_k = \mathbf{I} - \boldsymbol{H}_k^{+} \boldsymbol{H}_k. \tag{A53}$$

Due to (A53), $\boldsymbol{H}_k\boldsymbol{v}_k = \mathbf{0}$ implies that $\boldsymbol{\Pi}_k\boldsymbol{v}_k = \boldsymbol{v}_k$. Therefore, our optimization problem is equivalent to the optimization problem

$$\boldsymbol{v}_k^{\mathrm{ZF}} = \arg\max_{\boldsymbol{v}_k} |\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{\Pi}_k\boldsymbol{v}_k|^2 \qquad \text{s.t.:} \quad \boldsymbol{H}_k\boldsymbol{v}_k = \mathbf{0} \quad \text{and} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1 \tag{A54}$$

in which $\boldsymbol{v}_k$ was replaced by $\boldsymbol{\Pi}_k\boldsymbol{v}_k$ in the objective function, which is identical as long as the first constraint holds.

Next, we drop the first constraint and solve the optimization problem

$$\boldsymbol{v}_k' = \arg\max_{\boldsymbol{v}_k} |\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{\Pi}_k\boldsymbol{v}_k|^2 \qquad \text{s.t.:} \quad \|\boldsymbol{v}_k\|_2^2 \leq 1. \tag{A55}$$

As for the selfish beamformer $\boldsymbol{v}_k^{\mathrm{MF}}$, cf. (4.17), it is straightforward to show that

$$\boldsymbol{v}_k' = \frac{1}{\|\boldsymbol{\Pi}_k\boldsymbol{h}_{kk}^*\|_2} \cdot \boldsymbol{\Pi}_k\boldsymbol{h}_{kk}^*. \tag{A56}$$

Since, with the properties of the pseudo-inverse,

$$\boldsymbol{H}_k\boldsymbol{v}_k' = \frac{1}{\|\boldsymbol{\Pi}_k\boldsymbol{h}_{kk}^*\|_2} \left(\boldsymbol{H}_k - \boldsymbol{H}_k\boldsymbol{H}_k^+\boldsymbol{H}_k\right) \boldsymbol{v}_k' = \mathbf{0}, \tag{A57}$$

$\boldsymbol{v}_k'$ fulfills the constraint that we dropped and therefore also is the solution to problem (A54). Consequently,

$$\boldsymbol{v}_k^{\mathrm{ZF}} = \frac{1}{\|\boldsymbol{\Pi}_k\boldsymbol{h}_{kk}^*\|_2} \cdot \boldsymbol{\Pi}_k\boldsymbol{h}_{kk}^*. \tag{A58}$$

A6. Derivation of the MISO Pricing Update

With the definitions (4.93)–(4.95), the KKT conditions of problem (4.91) are

$$\boldsymbol{A}(\rho)\boldsymbol{v}_k = \mu\boldsymbol{v}_k \tag{A59}$$

$\mu \geq 0$, $\|\boldsymbol{v}_k\|_2^2 \leq 1$, and $\mu(1 - \|\boldsymbol{v}_k\|_2^2) = 0$.

As in Section 4.4.5, we denote the highest eigenvalue of $\boldsymbol{A}$ as λ and the corresponding unit-norm eigenvector as $\boldsymbol{w}$. If λ has multiplicity higher than one, we define $\boldsymbol{w}$ to be the eigenvector "closest" to $\boldsymbol{h}_{kk}^*$, i. e., the eigenvector that maximizes $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2$. Both λ and $\boldsymbol{w}$ are also functions of ρ and therefore of ζ.

We can distinguish between three different possibilities for fulfilling the KKT conditions:

1) $\boldsymbol{v}_k = \mathbf{0}$: the null vector always fulfills the KKT conditions, but is not necessarily a local maximum.
2) $\mu = 0$ and $\boldsymbol{v}_k \neq \mathbf{0}$: if the solution is in the null space of the matrix $\boldsymbol{A}$, it does not necessarily have full power, i. e., $\|\boldsymbol{v}_k\|_2^2 \leq 1$. Clearly, such a solution can only be superior to the null solution if $\zeta = |\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2 > 0$.
3) $\mu > 0$: since $\boldsymbol{A}$ can have at most one positive eigenvalue due to its structure, and since μ must be a positive eigenvalue of $\boldsymbol{A}$ and $\boldsymbol{v}_k$ must have unit power, in this case $\mu = \lambda$ and $\boldsymbol{v}_k = \boldsymbol{w}$.

In the following we examine the behavior of λ and $\boldsymbol{w}$ when ρ is changed:

$$\boldsymbol{A}\boldsymbol{w} = \lambda\boldsymbol{w} \tag{A60}$$

$$\frac{\partial \boldsymbol{A}}{\partial \rho}\boldsymbol{w} + \boldsymbol{A}\frac{\partial \boldsymbol{w}}{\partial \rho} = \frac{\partial \lambda}{\partial \rho}\boldsymbol{w} + \lambda\frac{\partial \boldsymbol{w}}{\partial \rho} \tag{A61}$$

$$\boldsymbol{w}^{\mathrm{H}}\frac{\partial \boldsymbol{A}}{\partial \rho}\boldsymbol{w} + \boldsymbol{w}^{\mathrm{H}}\boldsymbol{A}\frac{\partial \boldsymbol{w}}{\partial \rho} = \frac{\partial \lambda}{\partial \rho} + \boldsymbol{w}^{\mathrm{H}}\lambda\frac{\partial \boldsymbol{w}}{\partial \rho} \tag{A62}$$

$$\frac{\partial \lambda}{\partial \rho} = |\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2. \tag{A63}$$

The second line is obtained by taking the derivative of both sides and applying the product rule; next, both sides are multiplied from the left with $\boldsymbol{w}^{\mathrm{H}}$; the fourth line results from inserting $\boldsymbol{w}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{w}^{\mathrm{H}}\lambda$ and $\partial\boldsymbol{A}/\partial\rho = \boldsymbol{h}_{kk}^{*}\boldsymbol{h}_{kk}^{\mathrm{T}}$.

Taking the derivative of (A61) w. r. t. ρ and applying the product rule yields

$$\frac{\partial^2 \boldsymbol{A}}{(\partial\rho)^2}\boldsymbol{w} + 2\frac{\partial \boldsymbol{A}}{\partial \rho}\frac{\partial \boldsymbol{w}}{\partial \rho} + \boldsymbol{A}\frac{\partial^2 \boldsymbol{w}}{(\partial\rho)^2} = \frac{\partial^2 \lambda}{(\partial\rho)^2}\boldsymbol{w} + 2\frac{\partial \lambda}{\partial \rho}\frac{\partial \boldsymbol{w}}{\partial \rho} + \lambda\frac{\partial^2 \boldsymbol{w}}{(\partial\rho)^2} \tag{A64}$$

$$2\boldsymbol{w}^{\mathrm{H}}\frac{\partial \boldsymbol{A}}{\partial \rho}\frac{\partial \boldsymbol{w}}{\partial \rho} = \frac{\partial^2 \lambda}{(\partial\rho)^2} + 2\boldsymbol{w}^{\mathrm{H}}\frac{\partial \lambda}{\partial \rho}\frac{\partial \boldsymbol{w}}{\partial \rho} \tag{A65}$$

where the second line follows from multiplying from the left with $\boldsymbol{w}^{\mathrm{H}}$, and inserting $\boldsymbol{w}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{w}^{\mathrm{H}}\lambda$ and $\partial^2\boldsymbol{A}/(\partial\rho)^2 = \boldsymbol{0}$. By multiplying (A61) from the left with $(\partial\boldsymbol{w}/\partial\rho)^{\mathrm{H}}$, we obtain

$$\left(\frac{\partial \boldsymbol{w}}{\partial \rho}\right)^{\mathrm{H}}\left(\frac{\partial \boldsymbol{A}}{\partial \rho} - \frac{\partial \lambda}{\partial \rho}\right)\boldsymbol{w} = \left(\frac{\partial \boldsymbol{w}}{\partial \rho}\right)^{\mathrm{H}}(\lambda\mathbf{I} - \boldsymbol{A})\frac{\partial \boldsymbol{w}}{\partial \rho}. \tag{A66}$$

From (A65) and (A66) it follows that

$$\frac{\partial^2 \lambda}{(\partial\rho)^2} = 2\left(\frac{\partial \boldsymbol{w}}{\partial \rho}\right)^{\mathrm{H}}(\lambda\mathbf{I} - \boldsymbol{A})\frac{\partial \boldsymbol{w}}{\partial \rho} \geq 0 \tag{A67}$$

since $\lambda\mathbf{I} - \boldsymbol{A}$ is positive semi-definite. Consequently

$$\frac{\partial |\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2}{\partial \rho} \geq 0. \tag{A68}$$

We observe that

- an increase in ζ implies an increase in γ_k, which implies a decrease in $u_k'(\gamma_k)$ (due to $u_k''(\gamma_k) < 0$), which implies a decrease in ρ, i. e., $\rho(\zeta)$ is strictly decreasing in ζ;
- due to (A63), λ is non-decreasing in ρ and therefore non-increasing in ζ; furthermore, if $\lambda > 0$, clearly $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2 > 0$, and λ is strictly decreasing in ζ;
- and, due to (A68), $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2$ is non-decreasing in ρ and therefore non-increasing in ζ.

With these monotonicity properties we are able to more closely characterize the solutions to the KKT conditions. For this purpose, we distinguish between the cases where zero-forcing is possible and where zero-forcing is not possible, as the behavior of the eigenvalue λ will turn out to be fundamentally different between these two cases.

A6.1 Zero-Forcing is Possible

The non-trivial zero-forcing solution $\boldsymbol{v}_k^{\mathrm{ZF}}$ fulfills $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k^{\mathrm{ZF}}|^2 > 0$ and $\pi_j^{(\ell)}\boldsymbol{h}_{jk}^{\mathrm{T}}\boldsymbol{v}_k^{\mathrm{ZF}} = 0$ for all $j \neq k$. Clearly such a vector exists if and only if $\boldsymbol{h}_{kk}$ is linearly independent of the space spanned by all $\pi_j^{(\ell)}\boldsymbol{h}_{jk}$ for $j \neq k$. In this case the matrix $\boldsymbol{A}$ has exactly one positive eigenvalue regardless of ρ. Furthermore, $\boldsymbol{A}\boldsymbol{v} = \boldsymbol{0}$ implies that $\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v} = 0$. Therefore, the KKT conditions can be fulfilled either by the null vector (and potentially other equivalent beamformers in the null space of $\boldsymbol{A}$) or by a unit-norm vector $\boldsymbol{v}_k$ that is the eigenvector of the matrix $\boldsymbol{A}(\rho(|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2))$ corresponding to the unique positive eigenvalue. In the latter case, $\boldsymbol{v}_k = \boldsymbol{w}$ and $\zeta = |\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2 = |\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2$.

Since $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2$ is non-increasing in ζ, and both $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2$ and ζ are in the interval $[0, \|\boldsymbol{h}_{kk}\|_2^2]$, it follows that there is exactly one value ζ_1 for which $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2 = \zeta_1$; it also follows that $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2 > \zeta$ if $\zeta < \zeta_1$, and that $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2 < \zeta$ if $\zeta > \zeta_1$. These properties of $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2$ are illustrated in Figure A1. Consequently, ζ_1 can be determined with arbitrary precision by means of bisection: starting with the lower bound zero and the upper bound $\|\boldsymbol{h}_{kk}\|_2^2$, we repeatedly check whether the arithmetic mean of the lower and upper bound is smaller or larger than $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2$ and update either the lower or upper bound accordingly.

It is clear that the zero vector cannot maximize the objective, as the zero-forcing solution is strictly superior to the zero vector. Therefore, the unique non-zero solution to the KKT conditions determined with the line search is the only candidate for the maximizer of the pricing objective and thus the solution to (4.91).

A6.2 Zero-Forcing is Not Possible

If $\boldsymbol{h}_{kk}^*$ lies in the span of the vectors $\pi_j^{(\ell)}\boldsymbol{h}_{jk}^*$ with $j \neq k$, i. e., in the span of the columns of the matrix $\boldsymbol{B}$, there exists a value ρ_2 for which the positive outer product $\rho_2\boldsymbol{h}_{kk}^*\boldsymbol{h}_{kk}^{\mathrm{T}}$ exactly cancels out one negative eigenvalue resulting from $-\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}$. To illustrate this, we use the reduced singular value decomposition $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}$ with $\boldsymbol{\Sigma} \in \mathbb{R}^{T\times T}$, where T is the rank of $\boldsymbol{B}$, as well as $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U} = \mathbf{I}$ and $\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V} = \mathbf{I}$. Since $\boldsymbol{h}_{kk}^*$ lies in the span of $\boldsymbol{B}$, there exists a vector $\boldsymbol{a} \in \mathbb{C}^T$ for which $\boldsymbol{h}_{kk}^* = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{a}$ and we can express the matrix $\boldsymbol{A}$ as

$$\boldsymbol{A} = \rho\boldsymbol{h}_{kk}^*\boldsymbol{h}_{kk}^{\mathrm{T}} - \boldsymbol{B}\boldsymbol{B}^{\mathrm{H}} = \boldsymbol{U}\boldsymbol{\Sigma}(\rho\boldsymbol{a}\boldsymbol{a}^{\mathrm{H}} - \mathbf{I})\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{H}}. \tag{A69}$$

With an arbitrary unitary matrix $\boldsymbol{W}$ of which the first column is collinear to $\boldsymbol{a}$, i. e., with $\boldsymbol{W}^{\mathrm{H}}\boldsymbol{a} = \|\boldsymbol{a}\|_2\boldsymbol{e}_1$, we can write

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{W}(\rho\|\boldsymbol{a}\|_2^2\boldsymbol{e}_1\boldsymbol{e}_1^{\mathrm{T}} - \mathbf{I})\boldsymbol{W}^{\mathrm{H}}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{H}}. \tag{A70}$$

It can now be seen that for $\rho = \rho_2 = 1/\|\boldsymbol{a}\|_2^2$ the first element of the diagonal matrix $\rho\|\boldsymbol{a}\|_2^2\boldsymbol{e}_1\boldsymbol{e}_1^{\mathrm{T}} - \mathbf{I}$ is zero and the rank of $\boldsymbol{A}$ is $T-1$, whereas for $\rho \neq \rho_2$ the rank of $\boldsymbol{A}$ is T. In the special case of $\rho = \rho_2$ the highest eigenvalue is $\lambda = 0$ and the eigenvector $\boldsymbol{w} = \alpha\boldsymbol{U}\boldsymbol{\Sigma}^{-1}\boldsymbol{a}$ lies in the null space of $\boldsymbol{A}$, but not in the null space of $\boldsymbol{h}_{kk}^{\mathrm{T}}$. Consequently $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2 > 0$.

As noted in the discussion of the KKT conditions at the beginning of this appendix, the case $\rho = \rho_2$ may yield a solution to the KKT conditions with $\mu = 0$ that is strictly superior to the zero vector. A candidate beamformer $\boldsymbol{v}_k$ for such a solution must fulfill $\rho(|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v}_k|^2) = \rho_2$ and at the same time $\|\boldsymbol{v}_k\|_2^2 \leq 1$ and $\boldsymbol{v}_k \parallel \boldsymbol{w}$. Beamformers that are in the null space of $\boldsymbol{A}$, but not parallel to $\boldsymbol{w}$, do not have to be considered, as they contain a component in the null space of $\boldsymbol{U}^{\mathrm{H}}$, which can be considered "wasted" power with no effect on the objective function.

With the power gain ζ_2 that fulfills $\rho(\zeta_2) = \rho_2$, existence of such a solution is equivalent to $0 \leq \zeta_2 \leq |\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2 = \left(\rho_2^2 \cdot \|\boldsymbol{\Sigma}^{-2}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{h}_{kk}^*\|_2^2\right)^{-1}$: the vector $\boldsymbol{v}_k$ fulfilling the KKT conditions can be

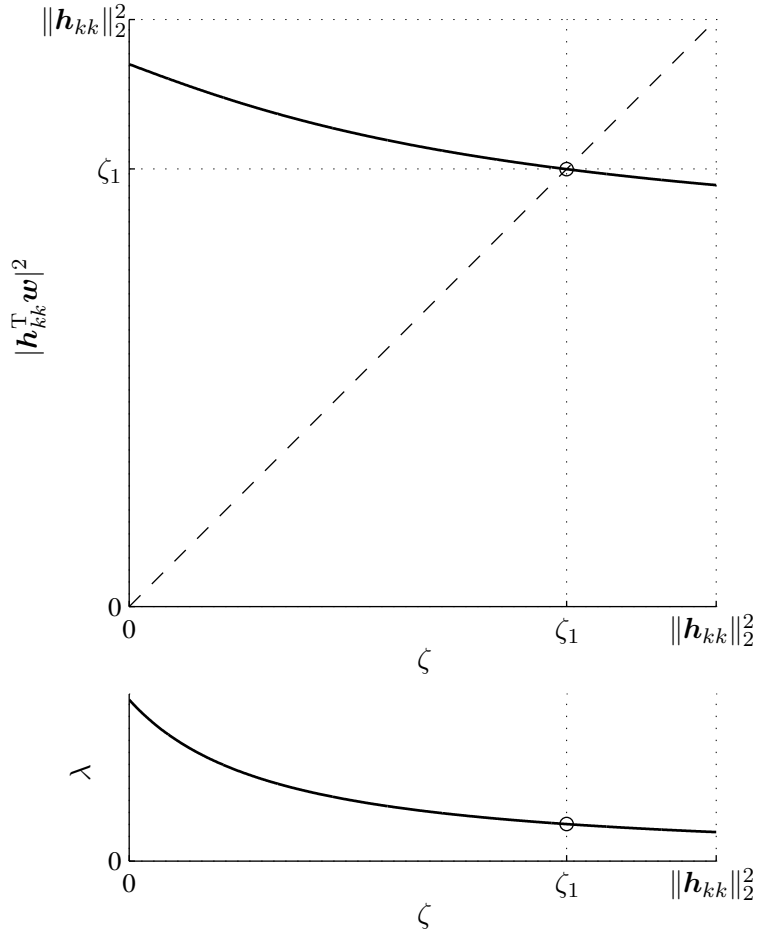


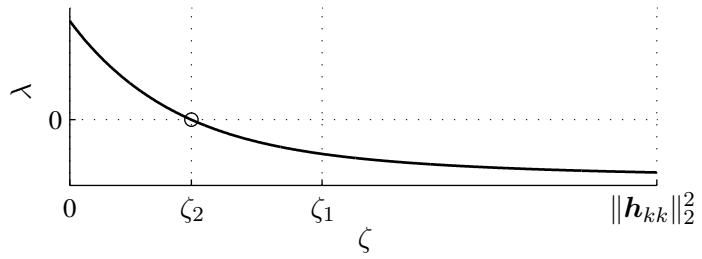
Figure A1: Behavior of $|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2$ and λ over ζ if zero-forcing is possible. The circle marks the value of ζ for which the KKT conditions of (4.91) are fulfilled.

found by weighting $\boldsymbol{w}$ with the scalar factor $\sqrt{\zeta_2/|\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2} \in [0,1]$. Also, $\zeta_2 \leq |\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2$ implies $\zeta_2 \leq \zeta_1$, where ζ_1 is the unique power gain for which the KKT conditions can be fulfilled with $\mu > 0$ (cf. Figure A1); and since λ is non-increasing in ζ and $\lambda = 0$ for $\zeta = \zeta_2$, the highest eigenvalue for ζ_1 cannot be positive. Therefore, $\boldsymbol{v}_k$ is the only non-trivial solution to the KKT conditions.

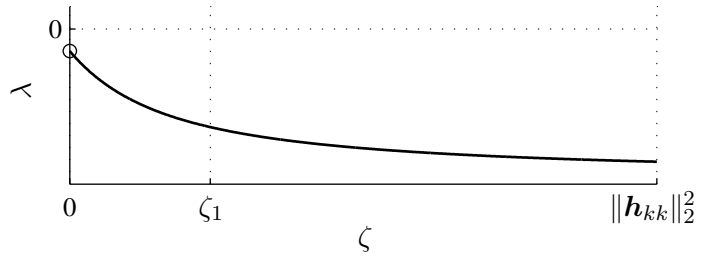
If $\zeta_2 < 0$, on the other hand, ρ_2 is not achievable for any non-negative power gain. Also, $\boldsymbol{A}$ in this case is negative semi-definite regardless of ζ and $\boldsymbol{A}\boldsymbol{v} = \boldsymbol{0}$ implies that $\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{v} = 0$. Here, the null vector (or any equivalent beamformer in the null space of $\boldsymbol{U}^{\mathrm{H}}$) is the only solution to the KKT conditions.

If $\zeta_2 > |\boldsymbol{h}_{kk}^{\mathrm{T}}\boldsymbol{w}|^2$, finally, it is not possible to achieve ρ_2 without violating the power constraint. Due to the previously discussed monotonicity properties, however, it is clear that $\zeta_1 < \zeta_2$ and that for ζ_1 the highest eigenvalue of $\boldsymbol{A}$ is positive. The value of ζ_1 can be found via bisection starting with the lower bound zero and the upper bound $\min\{\zeta_2, \|\boldsymbol{h}_{kk}\|_2^2\}$ and the optimal update is the unit-norm eigenvector $\boldsymbol{w}$ evaluated at ζ_1.

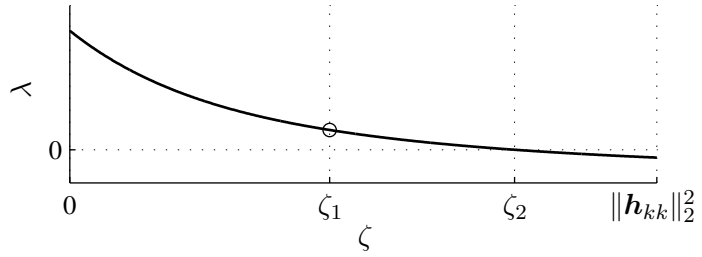
With these observations, we obtain the update procedure described in Section 4.4.5. The behavior of λ over ζ for the three different cases is illustrated in Figure A2.



(a) $0 \leq \zeta_2 \leq \zeta_1$: $\mu = 0$ and power control



(b) $\zeta_2 < 0$: $\mu = 0$ and zero power



(c) $\zeta_2 > \zeta_1$: $\mu = \lambda > 0$ and full power

Figure A2: Three possibilities when zero-forcing is not possible. The circle marks the value of ζ for which the KKT conditions of (4.91) are fulfilled.

Bibliography

[1] N. G. de Bruijn, *Asymptotic Methods in Analysis*. Mineola, NY: Dover Publications, Inc., 1981.

[2] J. Huang, R. A. Berry, and M. L. Honig, "Distributed interference compensation for wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 5, pp. 1074–1084, May 2006.

[3] D. A. Schmidt, A. Gründinger, W. Utschick, and M. L. Honig, "Distributed precoder optimization for interfering MISO channels," in *Proc. Int. ITG Workshop Smart Antennas 2008*, Darmstadt, Germany, Feb. 2008.

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

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, ser. Wiley Series in Telecommunications. New York, NY: John Wiley & Sons, Inc., 1991.

[6] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Stat., Probabil.*, vol. 1, Berkeley, CA, 1961, pp. 611–644.

[7] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *Ann. Probabil.*, vol. 2, no. 5, pp. 805–814, Oct. 1974.

[8] A. B. Carleial, "A case where interference does not reduce capacity," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 5, pp. 569–570, Sep. 1975.

[9] ——, "Interference channels," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 1, pp. 60–70, Jan. 1978.

[10] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 49–60, Jan. 1981.

[11] A. Schrijver, *Theory of Linear and Integer Programming*, ser. Wiley-Interscience Series in Discrete Mathematics and Optimization. Chichester, England: John Wiley & Sons, Ltd, 1986.

[12] I. Sason, "On achievable rate regions for the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1345–1356, Jun. 2004.

[13] H.-F. Chong, M. Motani, H. K. Garg, and H. El Gamal, "On the Han-Kobayashi region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3188–3195, Jul. 2008.

[14] H. Sato, "The capacity of the Gaussian interference channel under strong interference," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 6, pp. 786–788, Nov. 1981.

[15] ——, "Two-user communication channels," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 3, pp. 295–304, May 1977.

[16] A. B. Carleial, "Outer bounds on the capacity of interference channels," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 4, pp. 602–606, Jul. 1983.

[17] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.

[18] G. Kramer, "Outer bounds on the capacity of Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 3, pp. 581–586, Mar. 2004.

[19] X. Shang, G. Kramer, and B. Chen, "A new outer bound and the noisy-interference sum-rate capacity for Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 689–699, Feb. 2009.

[20] A. S. Motahari and A. K. Khandani, "Capacity bounds for the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 620–643, Feb. 2009.

[21] S. Sridharan, A. Jafarian, A. Vishwanath, and S. A. Jafar, "Capacity of symmetric K-user Gaussian very strong interference channels," in *Proc. IEEE Global Commun. Conf. (GLOBECOM) 2008*, New Orleans, LA, Nov. 2008.

[22] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[23] A. S. Motahari, S. Oveis-Gharan, M.-A. Maddah-Ali, and A. K. Khandani, "Real interference alignment: Exploiting the potential of single antenna systems," arXiv.org, Nov. 2009. [Online]. Available: http://arxiv.org/abs/0908.2282v2

[24] S. Sridharan, A. Jafarian, S. Vishwanath, S. A. Jafar, and S. Shamai (Shitz), "A layered lattice coding scheme for a class of three user Gaussian interference channels," in *Proc. 46th Annu. Allerton Conf. Commun., Control, Computing*, Monticello, IL, Sep. 2008, pp. 531–538.

[25] R. Etkin and E. Ordentlich, "On the degrees-of-freedom of the K-user Gaussian interference channel," arXiv.org, Jan. 2009. [Online]. Available: http://arxiv.org/abs/0901.1695v1

[26] S. A. Jafar and S. Vishwanath, "Generalized degrees of freedom of the symmetric Gaussian K user interference channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3297–3303, Jul. 2010.

[27] X. Shang, B. Chen, and M. J. Gans, "On the achievable sum rate for MIMO interference channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4313–4320, Sep. 2006.

[28] E. Telatar and D. Tse, "Bounds on the capacity region of a class of interference channels," in *Proc. Int. Symp. Inform. Theory 2007*, Nice, France, Jun. 2007, pp. 2871–2874.

[29] S. Vishwanath and S. A. Jafar, "On the capacity of vector Gaussian interference channels," in *Proc. Inform. Theory Workshop 2004*, San Antonio, TX, Oct. 2004, pp. 365–369.

[30] X. Shang, B. Chen, G. Kramer, and H. V. Poor, "On the capacity of MIMO interference channels," in *Proc. 46th Annu. Allerton Conf. Commun., Control, Computing*, Monticello, IL, Sep. 2008, pp. 700–707.

[31] ——, "Capacity regions and sum-rate capacities of vector Gaussian interference channels," arXiv.org, Jul. 2009. [Online]. Available: http://arxiv.org/abs/0907.0472v1

[32] S. A. Jafar and M. J. Fakhereddin, "Degrees of freedom for the MIMO interference channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2637–2642, Jul. 2007.

[33] A. Ghasemi, A. S. Motahari, and A. K. Khandani, "Interference alignment for the K user MIMO interference channel," in *Proc. IEEE Int. Symp. Inform. Theory 2010*, Austin, TX, Jun. 2010, pp. 360–364.

[34] C. M. Yetis, T. Gou, S. A. Jafar, and A. H. Kayran, "On feasibility of interference alignment in MIMO interference networks," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4771–4782, Sep. 2010.

[35] S. Stańczak, M. Wiczanowski, and H. Boche, *Fundamentals of Resource Allocation in Wireless Networks*, 2nd ed., ser. Foundations in Signal Processing, Communications and Networking. Springer, 2009.

[36] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY: Cambridge University Press, 2004.

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

[38] A. Mas-Colell, M. D. Whinston, and J. R. Green, *Microeconomic Theory*. New York, NY: Oxford University Press, 1995.

[39] F. Kelly, "Charging and rate control for elastic traffic," *European Trans. Telecommun.*, vol. 8, no. 1, pp. 33–37, Jan. 1997.

[40] J. Brehmer and W. Utschick, "On proportional fairness in nonconvex wireless systems," in *Proc. Int. ITG Workshop Smart Antennas 2009*, Berlin, Germany, Feb. 2009.

[41] J. Brehmer, "Utility maximization in nonconvex wireless systems," Ph.D. dissertation, Technische Universität München, Munich, Germany, 2010.

[42] J. F. Nash, "The bargaining problem," *Econometrica*, vol. 18, no. 2, pp. 155–162, Apr. 1950.

[43] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556–567, Oct. 2000.

[44] C. Touati, H. Kameda, and A. Inoie, "Fairness in non-convex systems," University of Tsukuba, Ibaraki, Japan, Tech. Rep. CS-TR-05-4, Sep. 2005.

[45] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, ser. Computer Science and Applied Mathematics. San Diego, CA: Academic Press, 1979.

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

[47] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 57–73, Feb. 2008.

[48] H. Tuy, "Monotonic optimization: Problems and solution approaches," *SIAM J. Optimization*, vol. 11, no. 2, pp. 464–494, 2000.

[49] L. P. Qian, Y. J. Zhang, and J. Huang, "MAPEL: Achieving global optimality for a non-convex wireless power control problem," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1553–1563, Mar. 2009.

[50] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "Approaching the capacity of wireless networks through distributed interference alignment," arXiv.org, Mar. 2008. [Online]. Available: http://arxiv.org/abs/0803.3816v1

[51] J. Zander, "Distributed cochannel interference control in cellular radio systems," *IEEE Trans. Veh. Technol.*, vol. 41, no. 3, pp. 305–311, Aug. 1992.

[52] G. J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 641–646, Nov. 1993.

[53] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Pricing and power control in a multicell wireless data network," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 1883–1892, Oct. 2001.

[54] M. Xiao, N. B. Shroff, and E. K. P. Chong, "A utility-based power-control scheme in wireless cellular systems," *IEEE/ACM Trans. Netw.*, vol. 11, no. 2, pp. 210–221, Apr. 2003.

[55] C. Shi, R. A. Berry, and M. L. Honig, "Monotonic convergence of distributed interference pricing in wireless networks," in *Proc. IEEE Int. Symp. Inform. Theory 2009*, Seoul, Korea, Jun. 2009, pp. 1619–1623.

[56] X. Shang, B. Chen, and H. V. Poor, "Multi-user MISO interference channels with single-user detection: Optimality of beamforming and the achievable rate region," arXiv.org, Jul. 2009. [Online]. Available: http://arxiv.org/abs/0907.0505v1

[57] R. Zhang and S. Cui, "Cooperative interference management with MISO beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5450–5458, Oct. 2010.

[58] R. Mochaourab and E. Jorswieck, "Optimal beamforming in interference networks with perfect local channel information," arXiv.org, Nov. 2010. [Online]. Available: http://arxiv.org/abs/1004.4492v2

[59] E. A. Jorswieck, E. G. Larsson, and D. Danev, "Complete characterization of the Pareto boundary for the MISO interference channel," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5292–5296, Oct. 2008.

[60] J. Qiu, R. Zhang, Z.-Q. Luo, and S. Cui, "Optimal distributed beamforming for MISO interference channels," arXiv.org, Nov. 2010. [Online]. Available: http://arxiv.org/abs/1011.3890v2

[61] M. Bengtsson and B. Ottersten, "Optimal downlink beamforming using semidefinite optimization," in *Proc. 37th Annu. Allerton Conf. Commun., Control, Computing*, Monticello, IL, Sep. 1999, pp. 987–996.

[62] A. Wiesel, Y. C. Eldar, and S. Shamai (Shitz), "Linear precoding via conic optimization for fixed MIMO receivers," vol. 54, no. 1, pp. 161–176, Jan. 2006.

[63] H. D. Mittelmann, "An independent benchmarking of SDP and SOCP solvers," *Math. Programming*, vol. 95, no. 2, pp. 407–430, 2003.

[64] H. Behnke and F. Sommer, *Theorie der analytischen Funktionen einer komplexen Veränderlichen*, 3rd ed. Berlin, Germany: Springer-Verlag, 1965.

[65] Y.-F. Liu, Y.-H. Dai, and Z.-Q. Luo, "Coordinated beamforming for MISO interference channel: Complexity analysis and efficient algorithms," IEEE Trans. Signal Process., to be published.

[66] E. A. Jorswieck and E. G. Larsson, "Monotonic optimization framework for the two-user MISO interference channel," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 2159–2168, Jul. 2010.

[67] R. Zakhour and D. Gesbert, "Coordination on the MISO interference channel using the virtual SINR framework," in *Proc. Int. ITG Workshop Smart Antennas 2009*, Berlin, Germany, Feb. 2009.

[68] S. W. Peters and R. W. Heath, Jr., "Cooperative algorithms for MIMO interference channels," *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, pp. 206–218, Jan. 2011.

[69] D. A. Schmidt, C. Shi, R. A. Berry, M. L. Honig, and W. Utschick, "Minimum mean squared error interference alignment," in *Proc. 43rd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2009.

[70] H. Shen, B. Li, M. Tao, and X. Wang, "MSE-based transceiver designs for the MIMO interference channel," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3480–3489, Nov. 2010.

[71] S. S. Christensen, R. Agarwal, E. de Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[72] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA J. Numer. Anal.*, vol. 8, no. 1, pp. 141–148, 1988.

[73] J. Nocedal and S. J. Wright, *Numerical Optimization*, ser. Springer Series in Operations Research. New York, NY: Springer Science+Business Media, 1999.

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

[75] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo, "Linear transceiver design for interference alignment: Complexity and computation," arXiv.org, Sep. 2010. [Online]. Available: http://arxiv.org/abs/1009.3481v1

[76] D. A. Cox, J. Little, and D. O'Shea, *Using Algebraic Geometry*, 2nd ed., ser. Graduate Texts in Mathematics. New York, NY: Springer Science+Business Media, 2005.

[77] I. Z. Emiris and J. F. Canny, "Efficient incremental algorithms for the sparse resultant and the mixed volume," *J. Symbolic Comput.*, vol. 20, no. 2, pp. 117–149, Aug. 1995.

[78] O. González and I. Santamaría, "Interference alignment in single-beam MIMO networks via homotopy continuation," in *Proc. 36th Int. Conf. Acoust., Speech, Signal Process.*, Prague, Czech Republic, May 2011.

[79] C. W. Wampler, "Bezout number calculations for multi-homogeneous polynomial systems," *Applied Math. Comput.*, vol. 51, no. 2-3, pp. 143–157, Oct. 1992.

[80] D. A. Schmidt, W. Utschick, and M. L. Honig, "Large system performance of interference alignment in single-beam MIMO networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM) 2010*, Miami, FL, Dec. 2010.

[81] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York, NY: McGraw-Hill, 2002.

[82] H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd ed., ser. Wiley Series in Probability and Statistics. Hoboken, NJ: John Wiley & Sons, Inc., 2003.

[83] K. Johannson, "Shape fluctuations and random matrices," *Commun. Math. Phys.*, vol. 209, no. 2, pp. 437–476, 2000.

[84] I. M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *Ann. Stat.*, vol. 29, no. 2, pp. 295–327, 2001.

[85] C. A. Tracy and H. Widom, "Level-spacing distributions and the Airy kernel," *Commun. Math. Phys.*, vol. 159, no. 1, pp. 151–174, 1994.

[86] V. A. Marchenko and L. A. Pastur, "Distribution of eigenvalues in certain sets of random matrices," *Matematicheskii Sbornik*, vol. 72 (114), no. 4, pp. 507–536, 1967, (in Russian).

[87] M. F. Demirkol and M. A. Ingram, "Power-controlled capacity for interfering MIMO links," in *Proc. 54th IEEE Veh. Technol. Conf. Fall 2001*, vol. 1, Atlantic City, NJ, Oct. 2001, pp. 187–191.

[88] ——, "Stream control in networks with interfering MIMO links," in *Proc. IEEE Wireless Commun., Netw. Conf. 2003*, New Orleans, LA, Mar. 2003.

[89] S. Ye and R. S. Blum, "Optimized signaling for MIMO interference systems with feedback," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2839–2848, Nov. 2003.

[90] G. Arslan, M. F. Demirkol, and Y. Song, "Equilibrium efficiency improvement in MIMO interference systems: A decentralized stream control approach," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2984–2993, Aug. 2007.

[91] G. Scutari, D. P. Palomar, and S. Barbarossa, "The MIMO iterative waterfilling algorithm," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1917–1935, May 2009.

[92] J. Park, Y. Sung, and H. V. Poor, "On beamformer design for multiuser MIMO interference channels," arXiv.org, Nov. 2010. [Online]. Available: http://arxiv.org/abs/1011.6121v1

[93] D. A. Schmidt, W. Utschick, and M. L. Honig, "Beamforming techniques for single-beam MIMO interference networks," in *Proc. 48th Annu. Allerton Conf. Commun., Control, Computing*, Monticello, IL, Sep. 2010.

[94] I. Santamaria, O. Gonzales, R. W. Heath Jr., and S. W. Peters, "Maximum sum-rate interference alignment algorithms for MIMO channels," in *Proc. IEEE Global Commun. Conf. (GLOBECOM) 2010*, Miami, FL, Dec. 2010.

[95] H. Sung, S.-H. Park, K.-J. Lee, and I. Lee, "Linear precoder designs for K-user interference channels," *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, pp. 291–301, Jan. 2010.

[96] K. R. Kumar and F. Xue, "An iterative algorithm for joint signal and interference alignment," in *Proc. IEEE Int. Symp. Inform. Theory 2010*, Austin, TX, Jun. 2010, pp. 2293–2297.

[97] S.-H. Park, H. Park, Y.-D. Kim, and I. Lee, "Regularized interference alignment based on weighted sum-MSE criterion for MIMO interference channels," in *Proc. IEEE Int. Conf. Commun. 2010*, Cape Town, South Africa, May 2010.