

## Research Article

# Joint Utility-Based Power Control and Receive Beamforming in Decentralized Wireless Networks

Angela Feistel,<sup>1,2</sup> Sławomir Stańczak,<sup>1,2</sup> and Daniel Tomiczek<sup>2</sup>

<sup>1</sup>Heinrich-Hertz Chair, Faculty of EECS, Technical University of Berlin, Einsteinufer 25, 10587 Berlin, Germany

<sup>2</sup>Fraunhofer German-Sino Lab for Mobile Communications, Einsteinufer 37, 10587 Berlin, Germany

Correspondence should be addressed to Angela Feistel, [angela.feistel@hhi.fraunhofer.de](mailto:angela.feistel@hhi.fraunhofer.de)

Received 1 November 2009; Accepted 19 April 2010

Academic Editor: Jinhua Jiang

Copyright © 2010 Angela Feistel et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper addresses the problem of joint resource allocation in general wireless networks and its practical implementation aspects. The objective is to allocate transmit powers and receive beamformers to the users in order to maximize a network-wide utility that represents the attained QoS and is a function of the signal-to-interference ratios. This problem is much more intricate than the corresponding QoS-based power control problem. In particular, it is not known which class of utility functions allows for a convex formulation of this problem. In case of perfect synchronization, the joint power and receiver control problem can be reformulated as a power control problem under optimal receivers. Standard gradient projection methods can be applied to solve this problem. However, these algorithms are not applicable in decentralized wireless networks. Therefore, we decompose the problem and propose a convergent alternate optimization that is amenable to distributed implementation. In addition, in real-world networks noisy measurements and estimations occur. Thus, the proposed algorithm has to be investigated in the framework of stochastic approximation. We discuss practical implementation aspects of the proposed stochastic algorithm and investigate its convergence properties by simulations.

## 1. Introduction

Two central mechanisms for resource allocation and interference management in wireless networks are power control and beamforming. In order to ensure a high utilization of wireless resources, transmit powers and beamformers should be optimized jointly to exploit interdependencies between them. As is widely known the overall network can be optimized with respect to different optimization goals. In general, there exist two main approaches that are typically used. The classical QoS-based approach aims at satisfying a certain quality-of-service (QoS) requirement with minimum power. To circumvent the feasibility problem a related approach is to solve the so-called max-min SIR-balancing problem. In contrast to this stands the utility-based resource allocation problem where the network operator aims at optimizing a weighted aggregate utility so as to maximize the overall network performance. By appropriately choosing the utility function one can trade overall system efficiency against fairness. Widely known are the  $\alpha$ -fair strictly concave utility functions introduced by [1].

In decentralized wireless networks, however, in addition to efficiently managing wireless resources the two challenging tasks are to distributedly assign these resources and to apply stochastic algorithms that deal with noisy measurements and estimations. Thus in this paper we focus on the following problem: maximizing an aggregate utility jointly over powers and receive beamformers in real-world decentralized wireless networks.

*1.1. Related Work.* Classical QoS-based power control has been studied extensively (e.g., [2–4]). It aims at allocating transmit powers to the users such that each user meets its SIR target. Provided that the SIR requirements are feasible there exist iterative distributed algorithms that attain the target SIR [3, 5, 6]. Note, that a closely related approach to the classical approach is to maximize the minimum SIR [7–10]. In contrast to the classical QoS-based power control, the objective of utility-based power control is to optimize the overall network performance with respect to some aggregate utility function [11–18]. Recently distributed

utility-based power control algorithms have been developed by [14, 16–18]. In [17] the problem of joint power control and end-to-end congestion control is addressed where the power control part is a special case of the power control problem in [14]. The approach of [16] is a game-theoretic one. References [16, 17] apply a flooding protocol to pass locally available quantities to other nodes. The authors of [18] interpret the utility-based power control problem as a joint optimization of powers and SIR assignment over the feasibility region. They proposed a distributed power control and SIR assignment algorithm for the uplink in a multicell wireless network. In contrast [14] proposed a distributed utility-based power control algorithm for general wireless networks applying the notion of the adjoint network and thus avoiding to use a relatively expensive flooding protocol. In addition, the authors touched the problem of stochastic approximation and show how to deal with it in practice.

Independently from and simultaneously to our work the authors of [19] have proposed a distributed utility-based joint power control and receive beamforming algorithm for cellular uplinks applying the scheme of [18]. Apart from that, so far most work on joint power control and beamforming has focused on the QoS-based resource allocation, especially on the so-called max-min SIR-balancing and its related problem. For example in [20–22] the duality between uplink and downlink channels is exploited. Another strategy was proposed by [23, 24] showing that the problem can be embedded in semidefinite and conic optimization programs. The work of [25] extended the publication [22] to solve the max-min SIR-balancing problem under general power constraints.

However, apart from [15] the above set of publications considered only the deterministic case. First works incorporating imprecise knowledge of received waveforms can be found in [26–28]. Recently, stochastic algorithms for joint QoS-based power control and receive beamforming and their convergence analysis have been proposed by [29, 30].

*1.2. Summary of Main Results and Paper Organization.* In the following we consider the problem of joint power control and receive beamforming in order to maximize a certain aggregate utility function that represents the QoS attained and is a function of the SIR. However, in contrast to the pure power control problem [14], it is not known which class of utility functions allows a convex formulation of this joint optimization problem and thus enables an efficient global solution in distributed wireless networks. In particular, in case of the logarithmic function, the aggregate utility function appears to have relatively many local maxima.

Now, in this paper, under the assumption of perfect synchronization we first reformulate the joint power and receiver control problem as a pure power control problem under optimal receivers. This follows from the fact that an optimal receiver can be obtained in closed-form solution for any power vector. However, an efficient implementation of the equivalent gradient projection algorithm is notoriously difficult to achieve in decentralized wireless networks. Thus, we decompose the problem into two coupled subproblems

and propose an alternating algorithm that converges to a stationary point. If we confine our attention to utility functions whose relative concavity is larger than that of the logarithmic function, numerical experiments suggest that then the proposed algorithm may converge to a global maximum for a large set of initial SIRs.

In contrast to [15] which touches the problem briefly this paper provides a more detailed analysis. In addition it is devoted to practical implementation aspects that are completely missing so far. As already mentioned, in real-world networks noisy measurements and estimations occur. We embed the proposed alternating algorithm into the framework of stochastic approximation. In particular, we discuss in detail the imprecise knowledge of received waveforms and the influence of step size control on the convergence properties. Finally we provide extensive simulations on the convergence behavior as well as performance comparisons with pure power control schemes.

Potential applications of the resource allocation scheme presented in this paper are envisaged for example in wireless mesh networks to control transmit powers and beamformers of base stations (mesh routers). These base stations create a wireless backbone via multihop ad hoc networking and have practically unlimited energy supply.

## 2. System Model and Problem Statement

*2.1. System Model.* We consider a general multiple-antenna wireless network with an established network topology, in which all links share a common wireless spectrum. All users are equipped with  $M \geq 2$  antennas. Let  $K \geq 2$  users compete for access to the wireless links and let  $\mathcal{K} = \{1, \dots, K\}$  denote the index set of all users. Assume that  $k \in \mathcal{K}$  is arbitrary but fixed and define  $\mathbf{u}_l := \mathbf{u}_l^{(k)} = (u_{1,l}^{(k)}, \dots, u_{M,l}^{(k)}) \in \mathbb{C}^M$  to be the effective transmit vector of transmitter  $l \in \mathcal{K}$  associated with receiver  $k$ . The effective transmit vector  $\mathbf{u}_l$  is the product of the channel matrix between transmitter  $l \in \mathcal{K}$  and receiver  $k$  and its transmit beamformer. It determines the “direction” of the transmit signal. The effective transmit vector is assumed to be *arbitrary but fixed*, which implies that the channels and transmit beamformers are fixed. In contrast, the receive beamformers acting as linear receivers should be jointly optimized with transmit powers of the users. We use  $\mathbf{v}_k \in \mathbb{C}^M$  and  $p_k \geq 0$  to denote the receive beamformer and transmit power of user  $k$ , respectively. The receive beamformers of all users are collected in the receive beamforming matrix  $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_K) \in \mathbb{C}^{M \times K}$  and their transmit powers in the power vector  $\mathbf{p} = (p_1, \dots, p_K) \in \mathbb{R}_+^K$  (In what follows,  $\mathbb{R}_+$  and  $\mathbb{R}_{++}$  denote the set of nonnegative reals and positive reals, resp.). The transmit powers of the users are subject to individual power constraints  $P_1, \dots, P_K > 0$  so that  $\mathbf{p} \in \mathbf{P}$  must hold, where  $\mathbf{P} = \{\mathbf{p} \in \mathbb{R}_+^K : \forall_k p_k \leq P_k\} \subset \mathbb{R}_+^K$ . Furthermore, since the signal-to-interference ratio (SIR) is independent of the norm of the receive beamformers, we can assume that  $\|\mathbf{v}_k\|_2 = 1$  for each  $1 \leq k \leq K$ , and hence  $\mathbf{V} = \{\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_K) \in \mathbb{C}^{M \times K} : \forall_k \|\mathbf{v}_k\|_2 = 1\}$  denotes the set of all beamforming matrices. Note that both  $\mathbf{P}$  and  $\mathbf{V}$  are compact sets, so is also their Cartesian product  $\mathbf{P} \times \mathbf{V}$ .

Finally, we define  $P_+ = P \cap \mathbb{R}_{++}^K$ . In words,  $P_+$  is the set of *positive* power vectors satisfying the power constraints.

The main figure of merit is the SIR at the output of each receiver. Using the above notation and considering the fact that all users are perfectly synchronized, the SIR of user  $k$  is given by

$$\text{SIR}_k(\mathbf{p}, \mathbf{v}_k) := \frac{p_k G_{k,k}}{\sum_{l=1, l \neq k}^K p_l G_{k,l} + \sigma^2}, \quad (1)$$

where  $\sigma^2 > 0$  is the variance of independent zero-mean additive Gaussian noise, and  $G_{k,l} = |\langle \mathbf{v}_k, \mathbf{u}_l \rangle|^2$  is the attenuation of the power from transmitter of user  $l$  to receiver of user  $k$  where  $\langle \mathbf{v}_k, \mathbf{u}_l \rangle$  denotes the inner product of the vectors  $\mathbf{v}_k, \mathbf{u}_l$ . Note that the SIR of user  $k$  depends only on the  $k$ th receive beamformer  $\mathbf{v}_k$ .

**2.2. Problem Statement.** One way to control both transmit powers of the users and their receive beamformers is to apply a utility-based framework. Let  $U(\text{SIR}_k(\mathbf{p}, \mathbf{v}_k))$  be the utility of user  $k$  that represents the QoS level of user  $k$  under power vector  $\mathbf{p}$  and receive beamformer  $\mathbf{v}_k$ . Other typical interpretations include the degree of user satisfaction with the received SIR or the revenue of the network operator. Throughout the paper we assume that  $U$  satisfies the following conditions.

(A.1)  $U : \mathbb{R}_{++} \rightarrow \mathbb{Q}$  is a twice continuously differentiable, strictly *increasing* and strictly concave function of the SIR where  $\mathbb{Q}$  is an open interval on the real line such that  $U^{-1} : \mathbb{Q} \rightarrow \mathbb{R}_{++}$ .

(A.2)  $\lim_{x \rightarrow 0} U(x) := +\infty \Rightarrow \lim_{x \rightarrow 0} U'(x) = -\infty$ .

(A.3)  $U_e(x) := U(e^x)$  is convex on  $\mathbb{R}$ .

Since  $\mathbb{R}_{++}$  is an open set, all these assumptions imply that the first derivative  $U'(x)$  is positive on  $\mathbb{R}_{++}$ , that is, there are no isolated points  $x > 0$  such that  $U'(x) = 0$ .

The joint utility-based power control and receive beamforming problem can be written as follows. Given any weight vector  $\mathbf{w} = (w_1, \dots, w_K) > 0$ , we search for a power vector  $\mathbf{p}^* \in P$  and a beamforming matrix  $\mathbf{V}^* = (\mathbf{v}_1^*, \dots, \mathbf{v}_K^*)$  such that

$$(\mathbf{p}^*, \mathbf{V}^*) = \arg \max_{(\mathbf{p}, \mathbf{V}) \in P \times V} G(\mathbf{p}, \mathbf{V}), \quad (2)$$

where

$$G(\mathbf{p}, \mathbf{V}) = \sum_{k \in K} w_k U(\text{SIR}_k(\mathbf{p}, \mathbf{v}_k)). \quad (3)$$

Since the noise variance is strict positive, standard arguments can be used to show that with our choice of the utility functions, the maximum exists. The convexity discussion of this problem, the development of a distributed algorithm, and its implementation in real-world environments with noisy measurements together with the performance evaluation by simulations will be the main tasks of this paper.

### 3. A Class of Utility Functions

Suppose that the utility function  $U : \mathbb{R}_{++} \rightarrow \mathbb{Q}$  is further confined to satisfy

$$r(x, U) = -\frac{xU''(x)}{U'(x)} \geq 1, \quad (4)$$

where  $U'(x) > 0$  and  $U''(x) < 0$ ,  $x > 0$  denote the first and second derivatives of  $U$ , respectively. Then, we know that [15], for any fixed  $\mathbf{V} \in V$ ,  $G_e(\mathbf{s}, \mathbf{V}) := G(e^{\mathbf{s}}, \mathbf{V})$  is concave in the logarithmic power vector  $\mathbf{s} := \log \mathbf{p} \in S$  with  $\mathbf{p} \in P_+$  and

$$S := \{\mathbf{s} = \log \mathbf{p} : \mathbf{p} \in P_+\} \subset \mathbb{R}^K. \quad (5)$$

Here and hereafter,  $\log \mathbf{p}$ ,  $\mathbf{p} \in \mathbb{R}_{++}^K$ , and  $e^{\mathbf{s}}$ ,  $\mathbf{s} \in \mathbb{R}^K$ , are defined component-wise. Since the logarithm is a bijection from  $\mathbb{R}_{++}$  onto  $\mathbb{R}$  and  $\mathbf{p}^* > 0$ , there is a one-to-one relationship  $\mathbf{s}^* = \log \mathbf{p}^*$  between optimal power vectors  $\mathbf{p}^*$  and optimal logarithmic power vectors  $\mathbf{s}^*$ . The motivation beyond this substitution is the following fact [15].

**Lemma 1.**  $r(x, U) \geq 1$  if and only if  $U(e^x)$  is concave on  $\mathbb{R}$ .

Since  $U(x)$ ,  $x > 0$ , is twice continuously differentiable,  $U(e^x)$ ,  $x \in \mathbb{R}$ , is concave if and only if  $e^x(U''(e^x)e^x + U'(e^x)) \leq 0$ ,  $x \in \mathbb{R}$ . This in turn holds if and only if  $U''(x)x + U'(x) \leq 0$ , which is equivalent to (11) since  $U'(x) > 0$  for all  $x > 0$ .

Therefore, for  $G_e(\mathbf{s}, \mathbf{V})$  to be concave in  $\mathbf{s}$ , it is sufficient that  $\text{SIR}_k(e^{\mathbf{s}})$  is a log-concave function of  $\mathbf{s} \in \mathbb{R}^K$ . In economics, the quantity  $r(x, U) \geq 0$  is known as the coefficient of relative risk aversion [31] and is used to measure the relative concavity of  $U(x)$ . The larger the value of  $r(x, U) \geq 0$  is, the larger is the relative concavity of  $U(x)$  at  $x > 0$ , and therefore a better fairness performance (at the cost of the throughput performance) can be expected. A prominent example of a function that satisfies (4) is the logarithmic function  $U(x) = \log(x)$ ,  $x > 0$ , in which case  $r(x, U) = 1$ ,  $x > 0$ .

Now the question is what happens if we use this class of utility functions in the joint power control and receive beamforming problem (2). First note that this problem can be written as a power control problem because

$$\begin{aligned} \max_{(\mathbf{p}, \mathbf{V}) \in P \times V} G(\mathbf{p}, \mathbf{V}) &= \max_{\mathbf{p} \in P} G(\mathbf{p}, \mathbf{V}^*(\mathbf{p})) \\ &= \max_{\mathbf{p} \in P_+} G(\mathbf{p}, \mathbf{V}^*(\mathbf{p})), \end{aligned} \quad (6)$$

where  $\mathbf{V}^*(\mathbf{p}) = (\mathbf{v}_1^*(\mathbf{p}), \dots, \mathbf{v}_K^*(\mathbf{p}))$  is used to denote an optimal receive beamforming matrix for a given  $\mathbf{p} > 0$  and the last step follows from (A.2). Obviously, since the SIR of user  $k$  depends only on the  $k$ th receive beamformer, one obtains

$$\begin{aligned} \max_{\mathbf{V} \in V} G(\mathbf{p}, \mathbf{V}) &= \sum_{k \in K} \max_{\|\mathbf{v}_k\|_2=1} w_k U(\text{SIR}_k(\mathbf{p}, \mathbf{v}_k)) \\ &= \sum_{k \in K} w_k U(\text{SIR}_k^*(\mathbf{p})), \end{aligned} \quad (7)$$

where

$$\text{SIR}_k^*(\mathbf{p}) := \max_{\|\mathbf{v}_k\|_2=1} \text{SIR}_k(\mathbf{p}, \mathbf{v}_k) \quad (8)$$

is the SIR under an optimal receive beamforming (see Section 4) and the last step follows since  $U(x), x > 0$ , is a strictly increasing function and  $\mathbf{w}$  is a positive vector. Now, using the substitution  $\mathbf{s} = \log \mathbf{p}$ ,  $\mathbf{p} \in \mathbb{P}_+$ , (in accordance with the power control problem of [15]) and

$$F(\mathbf{s}) := \sum_{k \in \mathbb{K}} w_k U(\text{SIR}_k^*(e^{\mathbf{s}})), \quad (9)$$

it follows that a solution to (2) is any pair  $(\mathbf{p}^*, \mathbf{V}^*) \in \mathbb{P}_+ \times \mathbb{V}$  given by  $\mathbf{p}^* = e^{\mathbf{s}^*}$  and  $\mathbf{V}^* = \mathbf{V}^*(\mathbf{p}^*)$  where

$$\mathbf{s}^* = \arg \max_{\mathbf{s} \in \mathbb{S}} F(\mathbf{s}). \quad (10)$$

In words, as in [15], the problem reduces to a power control problem except that now each SIR is assumed to attain its maximum overall receive beamformers.

Unfortunately, the condition (4) is not sufficient for  $F(\mathbf{s})$  defined by (9) to be a concave function of  $\mathbf{s} \in \mathbb{S}$ . A simple counter example is constructed in the appendix for  $U(x) = \log(x)$ ,  $x > 0$ . Numerical experiments show that if  $U(x) = \log(x)$ ,  $x > 0$ , then the gradient projection algorithm is not globally convergent, that is, it in general converges to a local maximum which is not global. Given  $\mathbb{V}$  and  $\mathbb{P}$ , the aggregate utility function seems to have relatively many local maxima.

A simple idea is to further restrict the class of utility functions by requiring larger values of  $r(x, U)$  for all  $x > 0$ . For instance, we could demand that

$$r(x, U) > 1. \quad (11)$$

This excludes the logarithmic function and implies that  $U(e^x)$  is strictly concave. A class of utility functions that satisfies (11) are the following functions

$$U^{(\alpha)}(x) := \frac{x^{1-\alpha}}{1-\alpha}, \quad \alpha \geq 2, \alpha \in \mathbb{N}. \quad (12)$$

Indeed, it may be easily verified that  $r(x, U^{(\alpha)}) = \alpha$ , and hence (11) holds for all  $\alpha \geq 2, \alpha \in \mathbb{N}$ . Another example is

$$U(x) = \log \left[ \frac{x}{(1+x)} \right], \quad x > 0, \quad (13)$$

in which case  $r(x, U) = (1+2x)/(1+x) \in (1, 2)$ . So, at low values of  $x > 0$ , the function in (13) behaves like the logarithmic function. In contrast, as  $x$  increases, it is similar to the negative inverse function. Numerical experiments with the utility function (13) suggest that in this case, the gradient projection algorithm (see Section 4.1) converges to a global maximum for a relatively large set of initial SIR values. When compared with the logarithmic utility function, convergence to a local point was observed in significantly fewer cases. However, we can show that  $F(\mathbf{s})$  with (13) is not concave in general and the standard gradient projection algorithms are not globally convergent for all initial SIR levels.

An interesting problem is whether a global convergence (if not for all starting points, then at least for most of them) of the gradient projection algorithm can be achieved by requiring that  $r(x, U) \geq c$ ,  $x > 0$ , for some sufficiently large constant  $c \geq 2$ . Increasing the constant  $c$  leads to utility functions with larger relative concavities. In particular, as shown below, if there is a utility function for which each addend in (9) is concave on  $\mathbb{R}^K$ , then  $F(\mathbf{s})$  is concave for all utility functions with a larger coefficient  $r(x, U)$ .

*Observation 1.* Let  $g : \mathbb{R}_{++} \rightarrow \mathbb{Q}_1$  be any utility function for which (4) holds, and suppose that each addend in  $F(\mathbf{s})$  with  $U(x) = g(x)$ ,  $x > 0$ , is concave on  $\mathbb{R}^K$ . Then,  $F(\mathbf{s})$  with  $U(x) = f(x)$ ,  $x > 0$ , is concave for any utility function  $f : \mathbb{R}_{++} \rightarrow \mathbb{Q}_2$  such that  $r(x, g) \leq r(x, f)$  for all  $x > 0$ .

Since  $g$  and  $f$  are bijective utility functions, there is a twice continuously differentiable and strictly increasing function  $h : \mathbb{Q}_1 \rightarrow \mathbb{Q}_2$  such that  $f(x) = h(g(x))$ ,  $x > 0$ . So, the observation follows if  $h(x)$  is concave. Considering the fact that  $f'(x) = h'(g(x))g'(x)$  with  $g'(x) > 0$  for all  $x > 0$ , the second derivative of  $f(x)$  yields  $f''(x) = h''(g(x))(g'(x))^2 + f'(x)g''(x)/g'(x)$ ,  $x > 0$ . By  $r(x, g) \leq r(x, f)$ , we have  $f''(x)g'(x) \leq f'(x)g''(x)$ ,  $x > 0$ , so that  $f''(x) \geq h''(g(x))(g'(x))^2 + f''(x)$ ,  $x > 0$ . This implies that  $h''(g(x)) \leq 0$ ,  $x > 0$ , and hence one obtains  $h''(x) \leq 0$ ,  $x \in \mathbb{Q}_1$  due to the bijectivity of  $g$ .

Applying this observation to the class in (12) reveals that if there was some  $\alpha \geq 2$  such that  $U^{(\alpha)}(\text{SIR}_k^*(e^{\mathbf{s}}))$  is concave on  $\mathbb{R}^K$  for each  $k \in \mathbb{K}$ , then the problem (10) would be a convex problem for all  $U^{(\alpha')}(x)$  with  $\alpha \leq \alpha'$ . Then, as discussed in the following section, we would be able to efficiently and arbitrarily close approximate the max-min fair rate allocation for any power constraints.

*3.1. An Arbitrarily Close Approximation of the Max-Min Fair Allocation.* Reference [1] introduced the class of utility functions in (12) to obtain different tradeoffs between throughput and fairness performance in wireline communications networks. In particular, it was shown that if each source is assigned the utility function  $U^{(\alpha)}(x)$ ,  $x > 0$ , then the corresponding rate allocation tends to the max-min rate allocation as  $\alpha \rightarrow \infty$ . For a large family of modulations determining the relationship between data rates attainable on wireless links and the SIR at the receiver output, this result carries over to our setting. To be precise, assume that  $\Phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is a one-to-one continuously differentiable function that maps the SIR values onto the data rates. A common assumption is that  $\Phi(x) = \log(1+x)$ ,  $x \geq 0$ . By this model, the set of all simultaneously achievable data rates is

$$\mathbb{C} = \{\mathbf{x} \geq 0 : \forall_k x_k \leq \Phi(\text{SIR}_k^*(\mathbf{p})), \mathbf{p} \in \mathbb{P}\} \quad (14)$$

which is a (connected) compact set since  $\Phi(\text{SIR}_k(\mathbf{p}, \mathbf{v}_k))$  is continuous on the compact set  $\mathbb{P} \times \mathbb{S}^{M-1}$ , where  $\mathbb{S}^{M-1}$  is the unit sphere in  $\mathbb{C}^M$ . This yields the following observation (see [15] and [1, Lemma 3]).

*Observation 2.* Let  $\mathbf{w} > 0$  be arbitrary, and let  $U = U^{(\alpha)}$  be given by (12). Then,  $(\Phi(\text{SIR}_1^*(\mathbf{p}^*)), \dots, \Phi(\text{SIR}_K^*(\mathbf{p}^*)))$  converges to a max-min fair rate allocation as  $\alpha \rightarrow \infty$ .

By the observation,  $\mathbf{p}^* = e^{\mathbf{s}^*}$  with  $U(x) = U^{(\alpha)}(x)$  converges to the max-min power allocation as  $\alpha$  tends to infinity [22]. Moreover, for every  $\alpha$ ,  $\nabla F(\mathbf{s})$  exists and is continuous on  $\mathbb{R}^K$  so that efficient gradient projection algorithms could be used to approximate the max-min power allocation for any power constraints if the algorithms were global convergent for some sufficiently large  $\alpha$  (as discussed before).

#### 4. Utility-Based Power Control and Beamforming Algorithm

In this section, we derive a gradient projection algorithm for the problem (19) and prove its convergence. To this end, let us first identify optimal receive beamformers. By (8), an optimal receive beamformer of user  $k$  is exactly that beamformer for which the  $k$ th SIR attains its maximum. Hence,

$$\mathbf{v}_k^*(\mathbf{p}) = \arg \max_{\|\mathbf{v}_k\|_2=1} \frac{p_k \mathbf{v}_k^H \mathbf{u}_k \mathbf{u}_k^H \mathbf{v}_k}{\mathbf{v}_k^H \mathbf{Z}_k(\mathbf{p}) \mathbf{v}_k}, \quad (15)$$

where

$$\mathbf{Z}_k(\mathbf{p}) = \sum_{l \neq k} p_l \mathbf{u}_l \mathbf{u}_l^H + \sigma^2 \mathbf{I} \quad (16)$$

is positive definite since  $\sigma^2$  is positive. As a consequence, the inverse matrix of  $\mathbf{Z}_k(\mathbf{p})$  exists regardless of the choice of the effective transmit vectors  $\mathbf{u}_k$  and  $\mathbf{p} \in \mathcal{P}$ . Note that the SIR can be written in this compact form due to the assumption of perfect synchronization. An optimal receive beamformer  $\mathbf{v}_k^*(\mathbf{p})$  can be easily found when the  $\text{SIR}_k$  is rewritten as a Rayleigh quotient to obtain [32]

$$\mathbf{v}_k^*(\mathbf{p}) = c_k \mathbf{Z}_k^{-1}(\mathbf{p}) \mathbf{u}_k, \quad (17)$$

where  $c_k > 0$  is a constant chosen such that  $\|\mathbf{v}_k^*(\mathbf{p})\|_2 = 1$ . Consequently, with an optimal beamformer, the SIR of user  $k$  is equal to

$$\text{SIR}_k^*(\mathbf{p}) = p_k \mathbf{u}_k^H \mathbf{Z}_k^{-1}(\mathbf{p}) \mathbf{u}_k. \quad (18)$$

From this, it follows that

$$\begin{aligned} \mathbf{s}^* &= \arg \max_{\mathbf{s} \in \mathcal{S}} F(\mathbf{s}) \\ &= \arg \max_{\mathbf{s} \in \mathcal{S}} \sum_{k \in \mathcal{K}} w_k U\left(e^{s_k} \mathbf{u}_k^H \mathbf{Z}_k^{-1}(e^{\mathbf{s}}) \mathbf{u}_k\right), \end{aligned} \quad (19)$$

and  $(\mathbf{s}^* = \log \mathbf{p}^*)$ ,

$$\mathbf{V}^* = (c_1 \mathbf{Z}_1^{-1}(\mathbf{p}^*) \mathbf{u}_1, \dots, c_K \mathbf{Z}_K^{-1}(\mathbf{p}^*) \mathbf{u}_K) \in \mathcal{V}, \quad (20)$$

with appropriately chosen constants  $c_1, \dots, c_K > 0$ .

If we assume the utility function (13) or the functions (12), then  $F(\mathbf{s})$  in (19) can be written using the inverse of

$$\mathbf{Z}(\mathbf{p}) = \mathbf{Z}_k(\mathbf{p}) + p_k \mathbf{u}_k \mathbf{u}_k^H \quad (21)$$

which is independent of the index  $k$ . Indeed, by the Sherman-Morrison formula [33], it follows that

$$\mathbf{Z}_k^{-1}(\mathbf{p}) = \mathbf{Z}^{-1}(\mathbf{p}) + \frac{\mathbf{Z}^{-1}(\mathbf{p}) \mathbf{u}_k \mathbf{u}_k^H \mathbf{Z}^{-1}(\mathbf{p}) p_k}{1 - p_k \mathbf{u}_k^H \mathbf{Z}^{-1}(\mathbf{p}) \mathbf{u}_k}, \quad (22)$$

and hence

$$\frac{\text{SIR}_k^*(\mathbf{p})}{1 + \text{SIR}_k^*(\mathbf{p})} = p_k \mathbf{u}_k^H \mathbf{Z}^{-1}(\mathbf{p}) \mathbf{u}_k. \quad (23)$$

So, if  $U(x) = \log(x/(1+x))$ ,  $x > 0$ , the aggregate utility function in (9) yields

$$\begin{aligned} F(\mathbf{s}) &= \sum_{k \in \mathcal{K}} w_k \log\left(e^{s_k} \mathbf{u}_k^H \mathbf{Z}^{-1}(e^{\mathbf{s}}) \mathbf{u}_k\right) \\ &= \sum_{k \in \mathcal{K}} w_k \log\left(e^{s_k} \mathbf{u}_k^H \text{adj}(\mathbf{Z}(e^{\mathbf{s}})) \mathbf{u}_k\right) \\ &\quad - \|\mathbf{w}\|_1 \log \det(\mathbf{Z}(e^{\mathbf{s}})). \end{aligned} \quad (24)$$

Choosing  $U(x) = U^{(\alpha)}(x)$  given by (12) gives

$$F(\mathbf{s}) = \frac{1}{1 - \alpha} \sum_{k \in \mathcal{K}} w_k \left( \frac{1 - e^{s_k} \mathbf{u}_k^H \mathbf{Z}^{-1}(e^{\mathbf{s}}) \mathbf{u}_k}{e^{s_k} \mathbf{u}_k^H \mathbf{Z}^{-1}(e^{\mathbf{s}}) \mathbf{u}_k} \right)^{\alpha-1}, \quad (25)$$

where  $\alpha \geq 2$  and the constant  $1/(1-\alpha)$  can be neglected as it has no impact on the maximizer.

*4.1. Gradient Projection Algorithm.* All partial derivatives of  $\text{SIR}_k^*(e^{\mathbf{s}})$  with  $\text{SIR}_k^*(\mathbf{p})$  given by (18) exist and are continuous functions on  $\mathbb{R}^K$  because the inverse matrix  $\mathbf{Z}_k^{-1}(e^{\mathbf{s}})$  exists for all  $\mathbf{s} \in \mathbb{R}^K$ , regardless of the choice of the effective transmit vectors, and the entries in  $\mathbf{Z}_k^{-1}(e^{\mathbf{s}})$  vary continuously with the entries in  $\mathbf{Z}_k(e^{\mathbf{s}})$ . Hence, we can consider a gradient projection algorithm with a constant step size  $\delta > 0$  (sufficiently small)

$$\mathbf{s}(n+1) = \Pi_{\mathcal{S}}[\mathbf{s}(n) + \delta \nabla F(\mathbf{s}(n))], \quad \mathbf{s}(0) \in \mathbb{R}^K, \quad (26)$$

where  $\Pi_{\mathcal{S}}(\mathbf{x})$  is the projection of  $\mathbf{x} \in \mathbb{R}^K$  on the closed convex set  $\mathcal{S}$  [34, 35] and the  $k$ th partial derivative  $\nabla_k F(\mathbf{s}) = (\partial F / \partial s_k)(\mathbf{s})$  yields

$$\begin{aligned} \nabla_k F(\mathbf{s}) &= w_k U'(\text{SIR}_k^*(e^{\mathbf{s}})) \text{SIR}_k^*(e^{\mathbf{s}}) \\ &\quad - e^{s_k} \sum_{l \neq k} w_l e^{s_l} U'(\text{SIR}_l^*(e^{\mathbf{s}})) \left| \mathbf{u}_l^H \mathbf{Z}_l^{-1}(e^{\mathbf{s}}) \mathbf{u}_k \right|^2, \end{aligned} \quad (27)$$

where the following identity was used. For an invertible and differentiable matrix function  $\mathbf{A}(x)$ , there holds

$$\frac{d\mathbf{A}^{-1}(x)}{dx} = -\mathbf{A}^{-1}(x) \frac{d\mathbf{A}(x)}{dx} \mathbf{A}^{-1}(x). \quad (28)$$

Hence, due to the individual power constraints on each user  $\forall_k s_k \leq \log P_k$ , the algorithm (26) takes the form

$$s_k(n+1) = \min \left\{ \log P_k, s_k(n) + \delta \left[ w_k U'(\text{SIR}_k^*(e^s)) \text{SIR}_k^*(e^s) - e^{s_k} \sum_{l \neq k} w_l e^{s_l} U'(\text{SIR}_l^*(e^s)) \left| \mathbf{u}_l^H \mathbf{Z}_l^{-1}(e^s) \mathbf{u}_k \right|^2 \right] \right\}, \quad (29)$$

where  $\text{SIR}_k^*(e^s)$  is defined by (18).

**Lemma 2.** *For a sufficiently small step size  $\delta > 0$ , the sequence  $\{\mathbf{p}(n)\}$  generated by the algorithm (29) with  $\mathbf{p}(n) = e^{s(n)}$  converges to a local stationary point.*

*Proof.* By standard results [34, 35], the gradient projection algorithm converges to a stationary point for sufficiently small values of  $\delta > 0$  if  $F(\mathbf{s})$  is bounded above, continuously differentiable on  $S$ , and the gradient  $\nabla F(\mathbf{s})$  is Lipschitz continuous on any bounded subset of  $S$ . The first condition is clearly satisfied due to the power constraints. The second condition holds as well since, by assumption, the utility function  $U(x)$  is twice continuously differentiable. Hence, the Hessian of  $F(\mathbf{s})$  is bounded in the matrix 2-norm on any bounded subset of  $S$ . This implies that  $\nabla F(\mathbf{s})$  is Lipschitz continuous on any bounded subset of  $S$  [36, page 70].  $\square$

Note that the maximum feasible step size in the algorithm may depend on the choice of the starting point  $\mathbf{s}(0)$ .

## 5. Distributed Implementation

The computation of the gradient in (29) might be too expensive to be implemented in a distributed environment. In this section, we slightly modify the algorithm so that it can be implemented in a distributed manner. The basic idea is to increase the value of the function  $G(\mathbf{p}, \mathbf{V})$  in the following alternating fashion. For some given receive beamforming matrix  $\mathbf{V}(n)$  and power vector  $\mathbf{p}(n)$ , a new power vector  $\mathbf{p}(n+1)$  is chosen such that  $G(\mathbf{p}(n), \mathbf{V}(n)) \leq G(\mathbf{p}(n+1), \mathbf{V}(n))$ . Then, the beamforming matrix is updated by  $\mathbf{V}(n+1)$  such that  $G(\mathbf{p}(n+1), \mathbf{V}(n)) \leq G(\mathbf{p}(n+1), \mathbf{V}(n+1))$ . This alternating process is repeated until convergence.

Let us first consider the power vector update. To this end, let  $\mathbf{V}$  be fixed and define  $F_{\mathbf{V}}(\mathbf{s}) := G(e^{\mathbf{s}}, \mathbf{V})$ . Then, the power vector can be updated according to the following algorithm

$$s_k(t+1) = \min \left[ \log P_k, s_k(t) + \delta_p(t) \nabla_k \hat{F}_{\mathbf{V}(t)}(\mathbf{s}(t)) \right], \quad t \in \mathbb{N}_0 \quad (30)$$

for some  $\mathbf{s}(0) \in \mathbb{R}^K$ , where, with some abuse of notation,  $\nabla \hat{F}_{\mathbf{V}(t)}(\mathbf{s}(t))$  is used to denote a noisy estimation of the gradient vector  $\nabla F_{\mathbf{V}(t)}(\mathbf{s}(t))$  and  $\{\delta_p(t)\}$  with  $\delta_p(t) > 0$  is an appropriately chosen sequence of diminishing step sizes [37]. If  $\{s_k(t)\}_{t=1}^L$  is a sequence generated by (30) for some  $L \geq 1$ , then we put  $\mathbf{s}(n+1) = (s_1(L), \dots, s_K(L))$ . Note that the estimate  $\nabla_k \hat{F}_{\mathbf{V}(t)}(\mathbf{s}(t))$ ,  $k \in \mathbb{K}$  can be computed

in a distributed manner using the adjoint network of [15]. This scheme enables each transmitter to estimate its current update direction from the received signal power. This mitigates the problem of global coordination of the transmitters when carrying out gradient-projection algorithms in distributed wireless networks. More precisely, instead of each node sending its message separately as in case of classical flooding protocols, nodes transmit simultaneously (only coarse synchronization is required) over the adjoint network such that each node can estimate its gradient component from the received power. The price for this are possible estimation errors that usually can be dealt with a diminishing step size [37] as is shortly discussed in the following Section 5.1.

Now assume that  $\mathbf{s} = \log \mathbf{p}$  is fixed. Distributed algorithms for computing optimal receive beamformers defined by (17) are widely established. These algorithms are based either on blind or pilot-based estimation methods [38]. In the latter case, if  $X_k$  is a pilot symbol of user  $k$  with zero mean and  $E[|X_k|^2] = e^{s_k}$ , and  $\mathbf{r}_k \in \mathbb{C}^M$  represents the observations at receiver  $k$ , then  $\mathbf{v}_k^*$  given by (17) minimizes the mean square error  $\theta_k(\mathbf{v}_k) = E[|\kappa X_k - \langle \mathbf{v}_k, \mathbf{r}_k \rangle|^2]$  over  $\mathbb{C}^M$ , where  $\kappa > 0$  is a normalizing constant chosen such that, in the minimum,  $\|\mathbf{v}_k\|_2 = 1$ . For practical implementation, we can assume  $\kappa = 1$ , and then normalize the beamformers so that their  $l^2$ -norms are equal to one. Besides note that the expectation is taken with respect to  $\mathbf{r}^{(k)} := (\mathbf{r}_k, X_k)$ , which depends on the logarithmic power vector  $\mathbf{s} \in \mathbb{R}^K$ . Now if the convex function  $\theta_k(\mathbf{v}_k)$  was explicitly known, then the algorithm (with the complex gradient operator  $\nabla$  which gives the direction of steepest ascent of  $\theta_k : \mathbb{C}^M \rightarrow \mathbb{R}_+$ )

$$\mathbf{v}_k(t+1) = \mathbf{v}_k(t) - \delta_k \nabla \theta_k(\mathbf{v}_k(t)), \quad k \in \mathbb{K}, t \in \mathbb{N}_0 \quad (31)$$

would converge to  $\mathbf{v}_k^*$  defined by (17) for a sufficiently small step size  $\delta_k > 0$ . The problem is that the function  $\theta_k$  is usually not known since the distribution of  $\mathbf{r}_k$  is not known [38]. Therefore,  $\nabla \theta_k(\mathbf{v}_k(n))$  cannot be computed and the algorithm must be modified using the framework of stochastic approximation [37]. The idea is to consider the functions  $\theta_{\mathbf{r}^{(k)}}(\mathbf{v}_k) = |\kappa X_k - \langle \mathbf{v}_k, \mathbf{r}_k \rangle|^2$  for all  $\mathbf{r}^{(k)}$  as noisy estimations of  $\theta_k(\mathbf{v}_k)$ . Then, under some conditions on the estimation error and for any  $\mathbf{v}_k(0) \in \mathbb{C}^M$ , the algorithm

$$\mathbf{v}_k(t+1) = \mathbf{v}_k(t) - \delta_k(t) \nabla \theta_{\mathbf{r}^{(k)}(t)}(\mathbf{v}_k(t)), \quad t \in \mathbb{N}_0 \quad (32)$$

converges to  $\mathbf{v}_k^*$  (in some probabilistic sense), provided that the step size  $\delta_k(t) > 0$  with  $\lim_{t \rightarrow \infty} \delta_k(t) = 0$  and  $\sum_{t=0}^{\infty} \delta_k(t) = +\infty$  is chosen suitably [26].

Now combining these two ingredients leads to the following joint power control and receive beamforming algorithm. At the beginning of every frame,  $\mathbf{s}(0)$  and  $\mathbf{V}(0)$  are set to be equal to the current transmit powers and receive beamformers. Then, all users concurrently execute  $N \geq 1$  updates of their transmit powers and receive beamformers. The  $n$ th update consists of the following intermediate steps.

- (i) For fixed  $\mathbf{V}(n)$  and some  $L \geq 1$ , each user  $k \in \mathbb{K}$  generates a sequence  $\{s_k^{(l)}\}_{l=1}^L$  by carrying out (30) and defines  $s_k(n+1) = s_k^{(L)}$ .

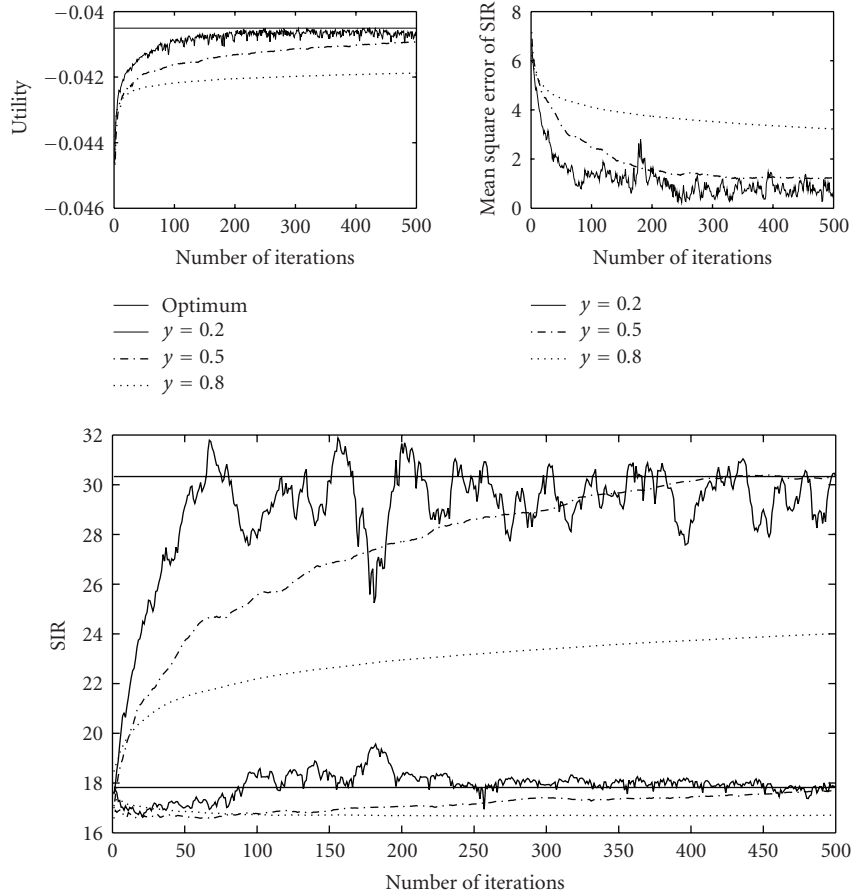


FIGURE 1: Convergence behavior of the distributed algorithm in case of noisy measurements for different values of  $\gamma$ ,  $c_1 = 30$  and  $c_2 = 0.5$ .

- (ii) For some  $L \geq 1$  and with  $E[|X_k|^2] = e^{s_k(n+1)}$ , each user  $k \in \mathcal{K}$  executes  $L$  iterations of the algorithm (32) to obtain the sequence  $\{\mathbf{v}_k^{(l)}\}_{l=1}^L$ . It defines  $\mathbf{v}_k(n+1) = \mathbf{v}_k^{(L)}$ .

The convergence of the algorithm (in some probabilistic sense) strongly depends on the choice of the step sizes in (30) and (32) as well as on the properties of the estimation errors in (30) and (32). However, we point out that the algorithm is motivated by the following observation. If the estimates in (30) are known perfectly meaning that we can use  $\delta_p(t) = \delta_p$  for sufficiently small  $\delta_p > 0$  and (31) is used instead of (32), then the sequence  $\{\mathbf{s}(n), \mathbf{V}(n)\}$  generated by the resulting algorithm converges to a stationary point. This is because, under this assumption, (30) and (31) are both monotonic, and hence we have (for all  $n \in \mathbb{N}_0$ )

$$\begin{aligned} G(\mathbf{p}(n), \mathbf{V}(n)) &\leq G(\mathbf{p}(n+1), \mathbf{V}(n)) \\ &\leq G(\mathbf{p}(n+1), \mathbf{V}(n+1)). \end{aligned} \quad (33)$$

This implies that the sequence  $\{G(\mathbf{p}(n), \mathbf{V}(n))\}$  is monotonically increasing, provided that the step sizes are sufficiently small. Moreover, it is bounded since  $G(\mathbf{p}(n), \mathbf{V}(n)) \leq G(\mathbf{p}^*, \mathbf{V}^*)$  for all  $n \in \mathbb{N}_0$ . Therefore, the algorithm converges to a stationary point. In addition, verifying the second order

sufficiency conditions would show that this stationary point is also a local maximizer for the problem (2).

Due to scarce resources in wireless networks, it is reasonable to choose the number of updates  $N = 1$  in every frame. In addition, instead of transmitting pilot signals in the intermediate step (ii), the optimal receive beamformers can be estimated during the data transmission using some blind estimation method (see [38] and references therein). So, at the beginning of every frame, the step (i) is executed only once. Then, the resulting transmit powers are used for data transmission. During this time, the receive beamformers are updated online after each transmitted symbol. However, numerical experiments suggest that the scheme should not exclusively rely on blind methods to estimate the optimal receivers with a sufficient accuracy.

**5.1. Stochastic Approximation View.** As already mentioned, in real-world networks estimation errors and other distorting factors as quantization noise occur. Now the interesting question is, what is the impact of these stochastic noisy measurements on the convergence properties. Does the proposed algorithm still converge and under what conditions? In the case of such uncertainties, the proposed algorithm has to be analyzed in the context of stochastic approximation theory. In the following we give several interesting insights. However,

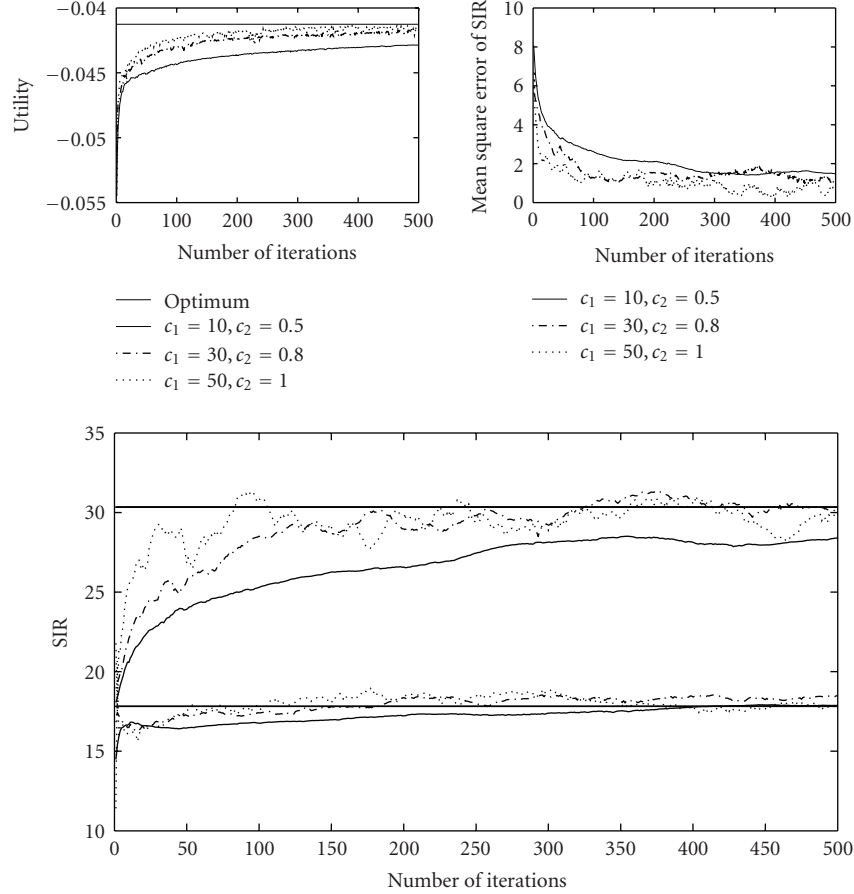


FIGURE 2: Convergence behavior of the distributed algorithm in case of noisy measurements for different values of the step sizes  $c_1, c_2$  and  $y = 0.5$ .

the topic is too broad to be discussed in all details. We also refer to [37] as a comprehensive reference.

We assume that the estimated gradient component  $\nabla \hat{F}_{\mathbf{V}(t)}(\mathbf{s}(t))$  is a random variable of the form

$$\nabla_k \hat{F}_{\mathbf{V}(t)}(\mathbf{s}(t)) = \nabla_k F_{\mathbf{V}(t)}(\mathbf{s}(t)) + M_k(t), \quad (34)$$

where  $M_k(t)$ ,  $k \in \mathcal{K}$  is the estimation noise process that fulfills the following conditions:

- (A.4) The estimation noise process depends on the receiver noise process which is assumed to be a martingale difference that is uncorrelated with transmit symbols and has a finite variance.
- (A.5) The estimation noise is zero mean and exogeneous, in the sense that  $M_k(t)$ ,  $k \in \mathcal{K}$  is independent of the iterate value.

Assuming these two conditions one can deal with the estimation noise applying a diminishing step size sequence that satisfies  $\delta_p(t) > 0$  with  $\lim_{t \rightarrow \infty} \delta_p(t) = 0$  and  $\sum_{t=0}^{\infty} \delta_p(t) = +\infty$ . A typical choice for a step size sequence is for instance  $\delta_p(t) = ct^\gamma$  for some  $\gamma \in (0, 1]$ . The choice of the step size is central to the effectiveness of the algorithm as is shown by simulations in the next section.

In the previous algorithm the powers and beamformers are updated in parallel, meaning that the power control algorithm does not wait for the convergence of the receive beamformers and vice versa. Thus the convergence of this practical stochastic algorithm is only verified by simulations presented in the following section. In addition, note that condition (A.5) is not necessarily fulfilled by the distributed power control algorithm. Thus the estimates  $\nabla_k \hat{F}_{\mathbf{V}(t)}(\mathbf{s}(t))$  may be biased by some  $b_k(t)$  meaning that  $\nabla_k \hat{F}_{\mathbf{V}(t)}(\mathbf{s}(t)) = \nabla_k F_{\mathbf{V}(t)}(\mathbf{s}(t)) + M_k(t) + b_k(t)$ . Simulation results indicate that the algorithm still converges to a contraction region around the optimal point provided that the bias is bounded by a scaled version of the true gradients.

## 6. Numerical Results

**6.1. Influence of Step Size Control.** In the following, we show exemplarily the convergence behavior of the proposed scheme for  $U(x) = -1/x$ ,  $x > 0$  and a random channel realization. We consider a wireless system with  $M = 2$  transmit and receive antennas, and  $K = 4$  users operating at a SNR level of 30 dB. The weight vector is  $\mathbf{w} = \mathbf{1}$ . The noisy measurements of the gradient are assumed to be  $\nabla_k \hat{F}_{\mathbf{V}(t)}(\mathbf{s}(t)) = \nabla_k F_{\mathbf{V}(t)}(\mathbf{s}(t)) + z$ ,  $k \in \mathcal{K}$ , where  $z$  is an



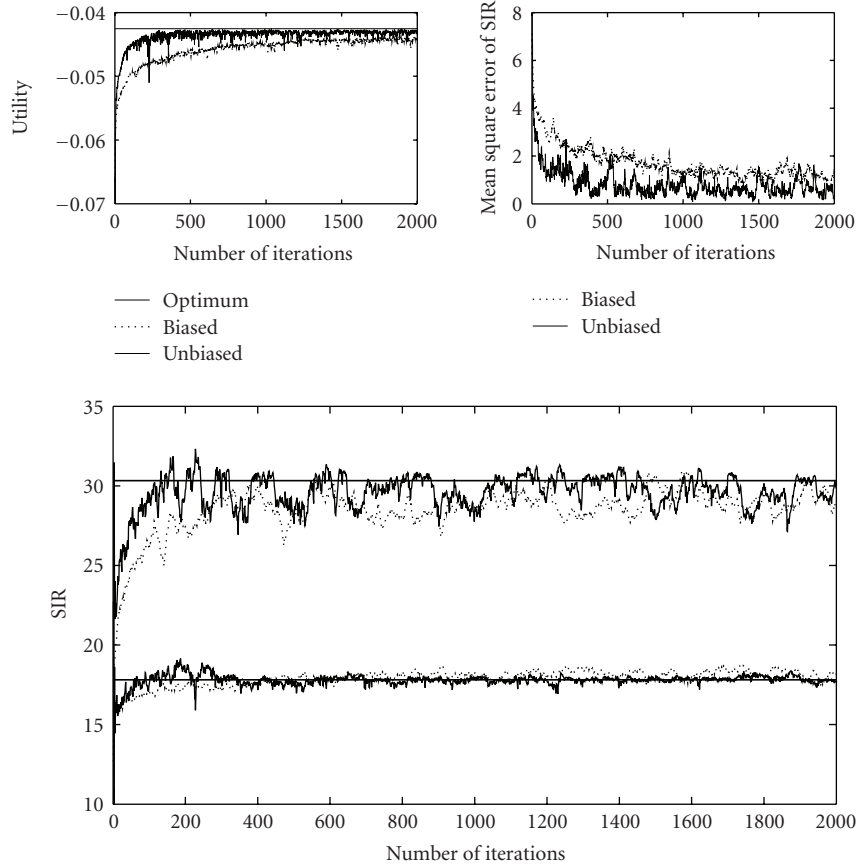


FIGURE 3: Convergence behavior of the distributed algorithm in case of noisy measurements  $\gamma = 0.4$ ,  $c_1 = 40$ ,  $c_2 = 1$  for the biased and unbiased case.

independent zero-mean Gaussian random variable (and thus fulfils the conditions of a martingale difference noise) whose variance  $\sigma_z^2(t)$  depends on  $t$  and is 10 percent of the absolute gradient value. We have  $L = 1$  and  $L = 16$  steps in (i) and (ii), respectively. Hence, during each iteration step  $n$ , the algorithm performs 1 power control step and estimates the beamformers using 16 pilot symbols. The diminishing step sizes for the intermediate steps are  $\delta_p(t) = c_1/(t+1)^\gamma$  and  $\delta_k(t) = c_2/(t+1)^\gamma$  for some positive constants  $c_1, c_2$  and some exponent  $\gamma \in (0, 1]$ . Figure 1 depicts the aggregate utility, the mean square error of the SIR, and the SIR for two users over the number of iterations  $n$  for different values of  $\gamma$  to show the influence of the diminishing step size. As can be easily seen, if the step size vanishes fast the algorithm converges much slower than with a slowly decreasing step size. However, the behavior is very smooth causing nearly no oscillations in contrast to a slowly decreasing step size. Figure 2 depicts the aggregate utility, the mean square error of the SIR, and the SIR for two users over the number of iterations  $n$  for different values of  $c_1$  and  $c_2$  and a fixed  $\gamma = 0.5$  to show the influence of the start step size values. Here a higher (but sufficiently small) start step size leads to a faster but oscillating convergence compared to lower start step sizes with a slow but smooth convergence behavior.

Summarizing, we state the following. It is important, that the step sizes  $c_1$  and  $c_2$  are sufficiently small to ensure that the algorithm does not diverge. Besides the decrease of the step sizes adjusted by exponent  $\gamma$  should be not too small to avoid a very slow convergence speed. In case of a dynamic environment where the channel changes over time,  $\gamma$  should be chosen to be able to follow the channel changes. This is paid with a more oscillating behavior. Finally note that the length of the pilot sequences also depends on the number of users because the link-specific pilot sequence is typically a pseudo-noise sequence with good autocorrelation properties.

**6.2. Influence of Biased Gradient Estimators.** In Figure 3 a convergence example is depicted for the case that the estimates  $\nabla_k \hat{F}_{V(t)}(\mathbf{s}(t))$  are biased by some  $b_k(t)$ . Further independent simulations suggest that the proposed algorithm converges to a contraction region around the optimal point if the bias is small enough. Otherwise the algorithm may diverge. However, the conditions on the bias to ensure convergence to a contraction region remain an open question.

**6.3. Comparison with Pure Utility-Based Power Control.** In this last section, we compare utility-based power control with

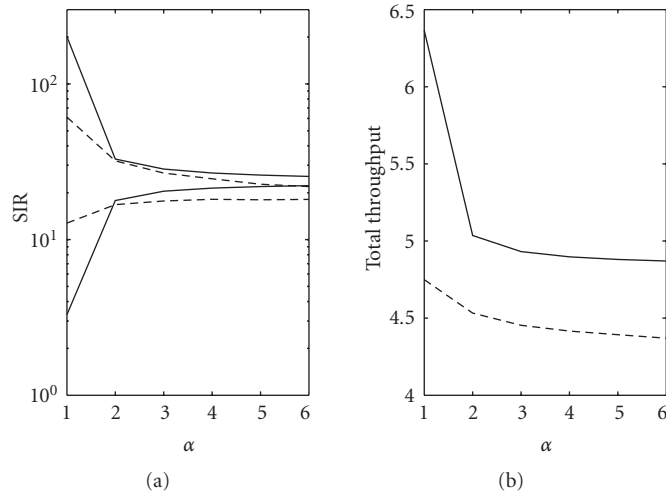


FIGURE 4: Maximum and minimum SIR (a), throughput (b) over  $\alpha$  for utility-based power control (dashed lines) and utility-based joint power and receiver control (solid lines).

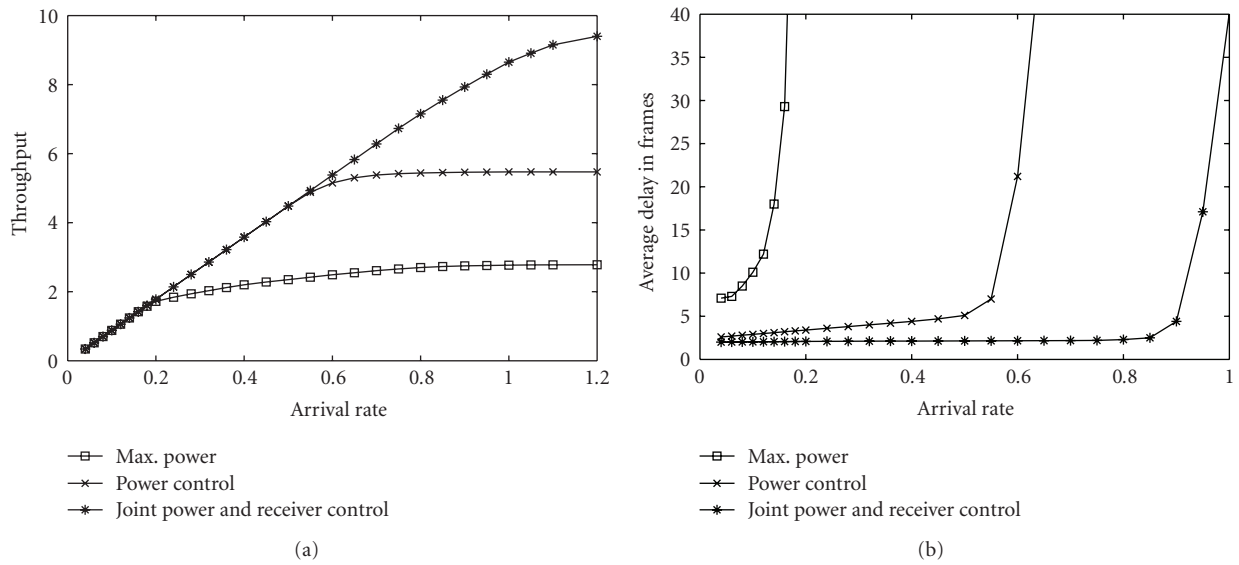


FIGURE 5: Throughput (a) and delay (b) performance as a function of the arrival rate for different resource allocation schemes and a given mesh network (c).

joint utility-based power control and receive beamforming. In Figure 4 the maximum and minimum SIRs are depicted as a function of  $\alpha$  representing the concavity of the utility function chosen. The figures show that a significant performance gain can be achieved by a joint optimization. Note that in this simulation example only a total throughput of 4.3 can be supported if the users transmit with maximum power and receive with a filter that is matched to the channel. In addition, the simulations confirm that with increasing concavity ( $\alpha$ ) the utility-based resource allocation strategy achieves fairness at the expense of a decreasing throughput performance. For  $\alpha \rightarrow \infty$  max-min fairness is achieved.

Finally, in Figure 5 we show the performance gains that can be achieved in an exemplary wireless mesh network (Downlink) where the base stations are connected wirelessly with an Access Point (AP). More precisely, the total network throughput and the delay performance are depicted over the arrival rate. For a fixed routing and a fixed scheduling strategy we compare the static resource allocation, that adapts the beamformers to the channel and transmits with maximum available transmit powers, with the utility-based power control and with the joint utility-based power control and receive beamforming for  $U(x) = \log(x)$ . The weights are chosen to represent the queue differences in order to support low delays. As can be easily seen the joint resource allocation outperforms the utility-based power control.

## 7. Conclusions

We proposed a framework for joint power control and receive beamforming in wireless networks, with the goal to maximize some aggregate utility function of the SIRs. The paper is a step to better understand the problem of utility-based power control and receive beamforming. We especially give insights into practical implementation issues and exemplarily show the effects of noisy estimations (unbiased and biased) as well as the influence of step size control on the convergence properties. However, the interesting theoretical issue of global convergence seems to be further unresolved.

## Appendix

We show that  $F(\mathbf{s})$  given by (9) is not concave in general when  $U(x) = \log(x)$ ,  $x > 0$ . To this end, consider the 2-user case with  $\mathbf{w} = \mathbf{1}$  so that  $F(\mathbf{s}) = \sum_{k=1}^2 \log(e^{s_k} \mathbf{u}_k^H \mathbf{Z}_k^{-1} \mathbf{u}_k)$  where  $\mathbf{Z}_1 = e^{s_2} \mathbf{u}_2 \mathbf{u}_2^H + \sigma^2 \mathbf{I}$  and  $\mathbf{Z}_2 = e^{s_1} \mathbf{u}_1 \mathbf{u}_1^H + \sigma^2 \mathbf{I}$ . Define  $\|\mathbf{u}_1\|_2^2 = c_{11} > 0$ ,  $\|\mathbf{u}_2\|_2^2 = c_{22} > 0$  and  $|\langle \mathbf{u}_1, \mathbf{u}_2 \rangle|^2 = c_{12} \geq 0$ . Without loss of generality, assume  $\sigma^2 = 1$ . Using the Sherman-Morrison formula [33] yields  $\mathbf{Z}_1^{-1} = (\mathbf{I} + e^{s_2} \mathbf{u}_2 \mathbf{u}_2^H)^{-1} = \mathbf{I} - (e^{s_2} / (1 + e^{s_2} c_{22})) \mathbf{u}_2 \mathbf{u}_2^H$  and  $\mathbf{Z}_2^{-1}$  is obtained in an equivalent way. Therefore,

$$F(\mathbf{s}) = s_1 + s_2 + \log \frac{c_{11} + e^{s_2} \rho}{1 + e^{s_2} c_{22}} + \log \frac{c_{22} + e^{s_1} \rho}{1 + e^{s_1} c_{11}}, \quad (\text{A.1})$$

where  $\rho = c_{11} c_{22} - c_{12}$ . Assume that  $\mathbf{u}_1 \neq \mathbf{u}_2$  and  $c_{12} > 0$ , in which case  $\rho > 0$ . Now taking the second derivative either with respect to  $s_1$  or  $s_2$  shows that  $F$  is not concave in general.

## Acknowledgments

This paper was supported in part by the German Ministry for Education and Research (BMBF) under Grant 01BU680 and in part by the Deutsche Forschungsgemeinschaft (DFG) under Grant STA 8653-1. The paper was partly presented at the 2007 Conference on Information Sciences and Systems (CISS), Washington, USA.

## References

- [1] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 556–567, 2000.
- [2] Q. Wu, "Optimum transmitter power control in cellular systems with heterogeneous SIR thresholds," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 4, pp. 1424–1429, 2000.
- [3] N. Bambos, S. C. Chen, and G. J. Pottie, "Channel access algorithms with active link protection for wireless communication networks with power control," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 583–597, 2000.
- [4] A. Feiten and R. Mathar, "Optimal power control for multiuser CDMA channels," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '05)*, pp. 1903–1907, Adelaide, Australia, September 2005.
- [5] G. J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Transactions on Vehicular Technology*, vol. 42, no. 4, pp. 641–646, 1993.
- [6] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1341–1347, 1995.
- [7] J. Zander, "Distributed cochannel interference control in cellular radio systems," *IEEE Transactions on Vehicular Technology*, vol. 41, no. 3, pp. 305–311, 1992.
- [8] J. Zander, "Performance of optimum transmitter power control in cellular radio systems," *IEEE Transactions on Vehicular Technology*, vol. 41, no. 1, pp. 57–62, 1992.
- [9] N. Bambos, "Toward power-sensitive network architectures in wireless communications: concepts, issues, and design aspects," *IEEE Personal Communications*, vol. 5, no. 3, pp. 50–59, 1998.
- [10] S. Koskie and Z. Gajic, "A Nash game algorithm for SIR-based power control in 3G wireless CDMA networks," *IEEE/ACM Transactions on Networking*, vol. 13, no. 5, pp. 1017–1026, 2005.
- [11] D. Goodman and N. Mandayam, "Power control for wireless data," *IEEE Personal Communications*, vol. 7, no. 2, pp. 48–54, 2000.
- [12] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Efficient power control via pricing in wireless data networks," *IEEE Transactions on Communications*, vol. 50, no. 2, pp. 291–303, 2002.
- [13] M. Xiao, N. B. Shroff, and E. K. P. Chong, "A utility-based power-control scheme in wireless cellular systems," *IEEE/ACM Transactions on Networking*, vol. 11, no. 2, pp. 210–221, 2003.
- [14] S. Stańczak, M. Wiczanowski, and H. Boche, "Distributed utility-based power control: objectives and algorithms," *IEEE Transactions on Signal Processing*, vol. 55, no. 10, pp. 5058–5068, 2007.
- [15] S. Stańczak, M. Wiczanowski, and H. Boche, *Fundamentals of Resource Allocation in Wireless Networks*, vol. 3, Springer, Berlin, Germany, 2009.

- [16] J. Huang, R. A. Berry, and M. L. Honig, "Distributed interference compensation for wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 5, pp. 1074–1084, 2006.
- [17] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: jointly optimal congestion control and power control," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 104–116, 2005.
- [18] P. Hande, S. Rangan, M. Chiang, and X. Wu, "Distributed uplink power control for optimal SIR assignment in cellular data networks," *IEEE/ACM Transactions on Networking*, vol. 16, no. 6, pp. 1420–1433, 2008.
- [19] T. Lan, P. Hande, and M. Chiang, "Joint beamforming and power control for optimal SIR assignment in cellular uplinks," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '07)*, pp. 2436–2440, Nice, France, June 2007.
- [20] F. Rashid-Farrokhi, L. Tassiulas, and K. J. R. Liu, "Joint optimal power control and beamforming in wireless networks using antenna arrays," *IEEE Transactions on Communications*, vol. 46, no. 10, pp. 1313–1324, 1998.
- [21] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 1, pp. 18–28, 2004.
- [22] M. Schubert and H. Boche, "QoS-based resource allocation and transceiver optimization," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 6, pp. 383–529, 2006.
- [23] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," in *Handbook of Antennas in Wireless Communications*, CRC Press, Boca Raton, Fla, USA, 2001.
- [24] A. Wiesel, Y. C. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 161–176, 2006.
- [25] A. Feistel, S. Stanczak, and M. Kaliszan, "Resource allocation in multi-antenna systems under general power constraints," in *Proceedings of the ITG/IEEE International Workshop on Smart Antennas (WSA '09)*, Berlin, Germany, February 2009.
- [26] M. Honig, U. Madhow, and S. Verdu, "Blind adaptive multiuser detection," *IEEE Transactions on Information Theory*, vol. 41, no. 4, pp. 944–960, 1995.
- [27] S. Ulukus and R. D. Yates, "Stochastic power control for cellular radio systems," *IEEE Transactions on Communications*, vol. 46, no. 6, pp. 784–798, 1998.
- [28] M. K. Varanasi, "Nonlinear multiuser receivers with distributed power control in cellular radio networks," in *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing*, pp. 820–830, Monticello, Ill, USA, September 1999.
- [29] J. Luo, S. Ulukus, and A. Ephremides, "Standard and quasi-standard stochastic power control algorithms," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2612–2624, 2005.
- [30] D. Das and M. K. Varanasi, "Distributed algorithms for joint optimization of multiuser receivers and power control," *IEEE Transactions on Wireless Communications*, vol. 6, no. 9, pp. 3374–3383, 2007.
- [31] A. Mas-Colell, M. D. Whinston, and J. R. Green, *Microeconomic Theory*, Oxford University Press, Oxford, UK, 1995.
- [32] P. Viswanath, V. Anantharam, and D. N. C. Tse, "Optimal sequences, power control, and user capacity of synchronous CDMA systems with linear MMSE multiuser receivers," *IEEE Transactions on Information Theory*, vol. 45, no. 6, pp. 1968–1983, 1999.
- [33] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia, Pa, USA, 2000.
- [34] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, Mass, USA, 1995.
- [35] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation*, Prentice Hall, Englewood Cliffs, NJ, USA, 1989.
- [36] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Classics in Applied Mathematics 30, SIAM, Philadelphia, Pa, USA, 2000.
- [37] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, Springer, Berlin, Germany, 2003.
- [38] S. Verdu, *Multiuser Detection*, Cambridge University Press, Cambridge, UK, 1998.