Springer Optimization and Engineering manuscript No.
(will be inserted by the editor)

Distributed Robust Optimization (DRO) Part II: Wireless Power Control

Yihong Wu · Kai Yang · Jianwei Huang · Xiaodong Wang · Mung Chiang

the date of receipt and acceptance should be inserted later

Abstract Optimization formulations and distributed algorithms have long been used for resource allocation problems in wireless networks including power control. However, the often assumed constant parameters in these formulations are in fact time-varying, unknown, or based on inaccurate estimates in real systems. Taking into account these factors, is it still possible to keep the algorithms distributed while making the solutions robust to inevitable errors in problem formulations? We answer this question through the general framework of Distributed Robust Optimization, as discussed in Part I of this two-part paper. In Part II we focus on proposing new distributed power control algorithms and analyzing their convergences under general row-wise uncertainty sets of channel matrices. We also quantify the tradeoff between robustness, measured by the deviation from the nominal problem, and distributiveness, measured by the amount of message passing in the algorithm. These results are then particularized to practical robustness models against channel fluctuations, signal to interference ratio (SIR) measurement errors, and user dynamics, described through ellipsoidal, polyhedral, and D-norm uncertain sets, respectively. Tradeoffs of robustness versus energy expenditure are also studied.

Yihong Wu
Department of Electrical Engineering, Princeton University, Princeton, NJ, USA
E-mail: yihongwu@princeton.edu

Kai Yang
Bell Labs, Alcatel Lucent, Murray Hill, NJ, USA
E-mail: kai.yang@alcatel-lucent.com

Jianwei Huang
Department of Information Engineering, The Chinese University of Hong Kong, HKSAR, China
E-mail: jwhuang@ie.cuhk.edu.hk

Xiaodong Wang
Department of Electrical Engineering, Columbia University, New York, NY, USA
E-mail: wangx@ee.columbia.edu

Mung Chiang
Department of Electrical Engineering, Princeton University, Princeton, NJ, USA
E-mail: chiangm@princeton.edu

Keywords Wireless power control · distributed robust optimization · uncertain systems · communication networks · network analysis and control.

1 Introduction

Optimization models have been widely used in communications and networking problems. There is often a need to provide both distributed and robust solutions to these problems, yet existing literature either treats robust optimization in a centralized way or distributed optimization without systematically modeling robustness. Distributed robust optimization (DRO) studies distributiveness-preserving robust formulation of optimization problems. Each robust optimization problem is characterized by a triple: a *nominal formulation*, the original problem with unperturbed constants, a *definition of robustness*, such as worst-case and probabilistic setup formulations, and an *uncertainty set*, which is application-specific and approximates uncertainties in practice. We are interested in seeking distributed robust solutions and quantifying tradeoffs between distributiveness, robustness and performance of the algorithm. Distributiveness is often measured by the communication overhead required to reach a prescribed level of optimality gap. In Part I [1] of the paper, we developed the foundation of DRO and illustrated it through an example of robust rate control.

In this paper we apply the DRO framework extensively to the wireless power control problem (see, e.g., the survey in [2]). Departing from a worst-case robustness definition, we first study the problem under general row-wise uncertainty sets of the channel matrix and propose a fixed-point algorithm. The uncertainty set structure is applicable to model independent variation of each user's channel coefficients. Utilizing nonnegative matrix theory and the standard interference function framework [3], we show that the common *necessary and sufficient* condition of feasibility, convergence, and optimality of the algorithm is that the maximal spectral radius of normalized channel matrices is bounded away from one. This generalizes the well-known results in [4] to a robust scenario.

Furthermore, we introduce the notion of *infrequent feedback* where globally coupled parameters are updated less frequently than user-end iterations. Frequency of feedback is one way to quantify the amount of message passing, thus the degree of distributedness, in a solution algorithm. Based on this technique, we propose a family of iterative algorithms parametrized by the frequency of message passing. We show that the algorithm converges exponentially fast in the number of iterations to the optimal power allocation under the same feasibility condition. The convergence speed decreases as the uncertainty set grows or the frequency of message passing decreases; this characterizes a fundamental tradeoff between distributiveness, robustness and performance. A quantitative analysis of this tradeoff provides insights for network designers to choose the best operating point of robust power control protocols. The above general results and algorithms are then applied to specific uncertainty models to produce practical distributive solutions.

The rest of the paper is organized as follows. Section 2 provides a brief review on the nominal wireless power control problem and the DRO framework. Section 3 studies the robust formulation under general row-wise uncertainty sets and proposes an iterative solution. The main result is a spectral characterization of feasibility, convergence and optimality as well as the convergence rate. Sections 4 – 6 are devoted to modeling channel fluctuations, signal to interference ratio (SIR) measurement inaccuracy, and user dynamics, using ellipsoids, polyhedrons, and D-norm uncertainty sets, respectively. We choose ellipsoidal uncertainty sets as the primary example. Section 7 concludes the paper. Several technical proofs are relegated to the appendix.

2 Background

2.1 The Nominal Problem and DPC (Distributed Power Control) Algorithm

Consider the following system model as in the seminal work by Foschini and Miljanic [4]. There exists a set of L users in the network. Each user consists of a transmitter node and a receiver node. If all users are distinct, this models a wireless ad hoc network. If all users share the same transmitter node (or receiver node), this models the downlink (or uplink) transmission in a cellular network. The Signal to Interference Ratio (SIR) on the link of user i is

$$\mathsf{SIR}_i = \frac{G_{ii} p_i}{\sum_{j \neq i} G_{ij} p_j + n_i}, \tag{1}$$

where G_{ij} is the channel gain from user j's transmitter to user i's receiver, and n_i is the AWGN noise power for user i's receiver. We want to optimize the users' transmission power $\boldsymbol{p} = [p_1, \ldots, p_L]$ to achieve a target SIR $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_L] \succ 0$[1], such that the total transmission power is minimized:

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^{\mathrm{T}} \boldsymbol{p} \\ \text{subject to} & \mathsf{SIR}_i(\boldsymbol{p}) \geq \gamma_i, \quad \forall i = 1, \ldots, L, \\ \text{variables} & \boldsymbol{p} \succeq 0, \end{array} \tag{2}$$

where $\mathbf{1}$ denotes the all-one vector. We further define $\boldsymbol{v} = \left[\frac{\gamma_1 n_1}{G_{11}}, \ldots, \frac{\gamma_L n_L}{G_{LL}}\right]$ and the normalized channel matrix $\boldsymbol{F} = [F_{ij}]$ with

$$F_{ij} = \begin{cases} 0 & i = j, \\ \frac{G_{ij}\gamma_i}{G_{ii}} & i \neq j, \end{cases} \tag{3}$$

then

$$\mathsf{SIR}_i = \frac{p_i \gamma_i}{\sum_{j \neq i} F_{ij} p_j + v_i}, \tag{4}$$

and Problem (2) can be written equivalently as:

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^{\mathrm{T}} \boldsymbol{p} \\ \text{subject to} & (\boldsymbol{I} - \boldsymbol{F})\, \boldsymbol{p} \succeq \boldsymbol{v}, \\ \text{variables} & \boldsymbol{p} \succeq 0, \end{array} \tag{5}$$

where $\boldsymbol{I}$ denote the $L \times L$ identity matrix.

It has been proved in [4] that Problem (2) is feasible and has the following unique, globally optimal solution

$$\bar{\boldsymbol{p}} = (\boldsymbol{I} - \boldsymbol{F})^{-1} \boldsymbol{v}, \tag{6}$$

[1] Throughout the paper, $\boldsymbol{x} \succeq \boldsymbol{y}$ (resp. $\boldsymbol{x} \succ \boldsymbol{y}$) denotes that $\boldsymbol{x} - \boldsymbol{y}$ is componentwise nonnegative (resp. positive).

if and only if $\rho(\boldsymbol{F}) < 1$, where $\rho(\boldsymbol{F})$ denotes the spectral radius of $\boldsymbol{F}$. Furthermore, if each user i locally measures its SIR value $\mathsf{SIR}_i(k)$ at each time slot k, and updates its transmission power according to

$$p_i(k+1) = \frac{\gamma_i}{\mathsf{SIR}_i(k)} p_i(k), \quad i = 1, \ldots, L, \tag{7}$$

the system will globally converge to the optimal solution in (6). Researchers in the wireless networking community refer to this fully distributed power control algorithm in (7) as the *DPC (Distributed Power Control) algorithm*. In the rest of this section, we will consider the robust optimization problems under either uncertainties in channel coefficients $\boldsymbol{F} = [F_{ij}]$, SIR estimate errors, or randomness in terms of users entering and leaving the network.

Robust power control has been studied in limited contexts before. Reference [5] initiated the study on how to reduce the impact of new users entering the system to the SIR of the existing links, by gradual power-up of incoming links and adding a protection margin to the target SIR of existing links. Here we address the issue from an alternative perspective of D-norm robust optimization and against a range of uncertainties: channel fluctuation, SIR measurement errors, and users entering and leaving the systems. In [6] the tradeoff between robustness and extra power consumption is studied with a penalty-defined formulation, while the key focus of this paper is to study the tradeoff between robustness and distributiveness as shown in Section 4.4, though we also study the tradeoff of extra power allocation versus robustness in Section 4.5. Moreover, the algorithm in [6] is primal-dual based and involves centralized computation by the base station, while Algorithm 2 we proposed has less complexity and only requires global message passing of a single parameter.

2.2 Distributed Robust Optimization

As in the Distributed Robust Optimization framework developed [1], we will focus on a class of optimization problems with the following nominal form: maximization of a *concave* objective function over a given data set characterized by *linear* constraints,

$$\begin{aligned} \text{maximize} \quad & f_0(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b}, \\ \text{variables} \quad & \boldsymbol{x}, \end{aligned} \tag{8}$$

where $\boldsymbol{A}$ is an $M \times N$ matrix, $\boldsymbol{x}$ is an $N \times 1$ vector, and $\boldsymbol{b}$ is an $M \times 1$ vector. This class of problems can model a wide range of engineering systems (e.g., [4, 7–9]) including the power control problems here.

Denote the j^{th} row of $\boldsymbol{A}$ by $\boldsymbol{a}_j^T$, which lies in a compact uncertainty set $\mathcal{A}_j$. Then the *robust* optimization problem in this paper can be written in the following form:

$$\begin{aligned} \text{maximize} \quad & f_0(\boldsymbol{x}), \\ \text{subject to} \quad & \boldsymbol{a}_j^T \boldsymbol{x} \leq b_j, \ \forall \boldsymbol{a}_j \in \mathcal{A}_j, \ \forall 1 \leq j \leq M, \\ \text{variables} \quad & \boldsymbol{x}. \end{aligned} \tag{9}$$

We can show that the robust optimization problem (9) can be equivalently written in a form represented by *protection functions* instead of uncertainty sets. Denote the nominal counterpart of problem (9) with a coefficient matrix $\bar{\boldsymbol{A}}$, i.e., the values when there is no uncertainty, with the j^{th} row's coefficient $\bar{\boldsymbol{a}}_j \in \mathcal{A}_j$. As proved in [1], we have

Proposition 1 *Problem (9) is equivalent to the following* convex *optimization problem:*

$$\begin{array}{lll} \text{maximize} & f_0(\boldsymbol{x}), & \\ \text{subject to} & \bar{\boldsymbol{a}}_j^T \boldsymbol{x} + \boldsymbol{g}_j(\boldsymbol{x}) & \leq b_j, \ \forall 1 \leq j \leq M. \\ \text{variables} & \boldsymbol{x}, & \end{array} \tag{10}$$

where

$$\boldsymbol{g}_j(\boldsymbol{x}) = \sup_{\boldsymbol{a}_j \in \mathcal{A}_j} (\boldsymbol{a}_j - \bar{\boldsymbol{a}}_j)^T \boldsymbol{x} \tag{11}$$

is the protection function for the j^{th} constraint, which depends on the uncertainty set $\mathcal{A}_j$ and the nominal row $\bar{\boldsymbol{a}}_j$. Each $\boldsymbol{g}_j$ is a convex function.

Different forms of $\mathcal{A}_j$ will lead to different protection functions $\boldsymbol{g}_j$, which result in different robustness and performance tradeoffs of the formulation. Next we consider several different approaches in terms of modeling $\mathcal{A}_j$ and the corresponding protection function $\boldsymbol{g}_j$.

3 Robust Power Control with General Uncertainty Sets

In this subsection we consider the robust formulation of Problem (2) under general row-wise uncertainty sets in the normalized channel matrix $\boldsymbol{F}$. In Sections 4 to 6, we see that the general fixed point algorithm proposed in Theorem 1 can be distributively implemented by exploiting the structures of the uncertainty sets.

Denote the normalized channel gain between user j's transmitter and user i's receiver as $F_{ij} = \bar{F}_{ij} + \Delta F_{ij}$, where $\bar{F}_{ij}$ is the nominal value (e.g., long term average value), and ΔF_{ij} is the perturbation associated with $\bar{F}_{ij}$. Further denote the i^{th} row of $\bar{\boldsymbol{F}}$ as $\bar{\boldsymbol{F}}_i$ and the corresponding perturbation vector as $\Delta \boldsymbol{F}_i$. Let $\mathcal{F}_i$ be the uncertainty set of $\boldsymbol{F}_i$, which captures the variations of interfering channel gains relative to the main channel gain of user i. The specific shape of the uncertainty set depends on the underlying channel model and sources of uncertainty, as shown in later subsections. The uncertainty set of $\boldsymbol{F}$ is

$$\mathcal{F} = \{\boldsymbol{F} : \boldsymbol{F}_i \in \mathcal{F}_i, \ i = 1, \ldots, L\}, \tag{12}$$

and $\bar{\boldsymbol{F}} \in \mathcal{F}$. The row-wise structure of $\mathcal{F}$ models the channel uncertainty well, since the channel gain vector at each user's receiver varies independently.

By Proposition 1, the robust version of the nominal power control problem (2) is given by

$$\begin{array}{rl} \text{minimize} & \mathbf{1}^{\mathrm{T}} \boldsymbol{p} \\ \text{subject to} & \boldsymbol{p} \succeq \bar{\boldsymbol{F}} \boldsymbol{p} + \boldsymbol{v} + \boldsymbol{g}(\boldsymbol{p}), \\ \text{variables} & \boldsymbol{p} \succeq 0. \end{array} \tag{13}$$

where $g_i(\boldsymbol{p})$ is the protection function for the i^{th} row as in (11):

$$g_i(\boldsymbol{p}) = \sup_{\boldsymbol{F}_i \in \mathcal{F}_i} (\boldsymbol{F}_i - \bar{\boldsymbol{F}}_i)^{\mathrm{T}} \boldsymbol{p}. \tag{14}$$

Next we state our main theorem: the necessary and sufficient condition for the feasibility of general robust power control problems, as well as the convergence and optimality of the fixed point iteration in (17), is that the supremum of the spectral radius of $\boldsymbol{F}$ in $\mathcal{F}$ is strictly less than 1. This is a generalization of the convergence condition of DPC algorithm in [4], which is a special case of Theorem 1 when $\mathcal{F}$ is a singleton set containing only the nominal value. Furthermore, the optimal power allocation in the robust case admits a simple characterization: it is the componentwise supremum of the optimal power (6) in DPC algorithm, when $\boldsymbol{F}$ ranges over the whole uncertainty set. This conclusion agrees with the worst-case robustness definition, in which case the optimal solution is given by conservatively assigning the largest power to each user under the worst channel condition.

Theorem 1 *The following three statements are equivalent:*

1) Problem (13) is feasible, i.e., there exists a $\hat{\boldsymbol{p}} \succeq 0$ *such that for any* $\boldsymbol{F} \in \mathcal{F}$*,*

$$\hat{\boldsymbol{p}} \succeq \boldsymbol{F}\hat{\boldsymbol{p}} + \boldsymbol{v}. \tag{15}$$

2) $\mathcal{F}$ *is bounded and*

$$\sup_{\boldsymbol{F}\in\mathcal{F}} \rho(\boldsymbol{F}) < 1. \tag{16}$$

3) Starting from any initial point, the following iteration

$$\boldsymbol{p}(k+1) = \bar{\boldsymbol{F}}\boldsymbol{p}(k) + \boldsymbol{v} + \boldsymbol{g}(\boldsymbol{p}(k)) \tag{17}$$

converges to the optimal solution of (13), given by the unique solution of

$$\boldsymbol{p}^* = \bar{\boldsymbol{F}}\boldsymbol{p}^* + \boldsymbol{g}(\boldsymbol{p}^*) + \boldsymbol{v}. \tag{18}$$

Moreover, the optimal power allocation admits the following characterization:

$$\boldsymbol{p}^* = \sup_{\boldsymbol{F}\in\mathcal{F}} (\boldsymbol{I} - \boldsymbol{F})^{-1}\boldsymbol{v}, \tag{19}$$

where the supremum is taken componentwise. If additionally the protection function $\boldsymbol{g}$ *is Fréchet-differentiable at* $\boldsymbol{p}^*$*, the convergence is exponential and the rate is given by*

$$\lim_{k\to\infty} \log \frac{1}{\|\boldsymbol{p}(k) - \boldsymbol{p}^*\|} = \log \frac{1}{\sup_{\boldsymbol{F}\in\mathcal{F}} \rho(\boldsymbol{F})}, \tag{20}$$

where $\|\cdot\|$ *is an arbitrary vector norm.*

Proof We show that 3) $\Leftrightarrow$ 1) $\Leftrightarrow$ 2).

3) $\Rightarrow$ 1) is obvious.

1) $\Rightarrow$ 3): We make use of Yates' approach of standard interference function [3] by defining the interference function as

$$\boldsymbol{I}(\boldsymbol{p}) = \bar{\boldsymbol{F}}\boldsymbol{p} + \boldsymbol{v} + \boldsymbol{g}(\boldsymbol{p}), \tag{21}$$

with $\boldsymbol{g}(\boldsymbol{p})$ as in (14). Then (15) implies that $\hat{\boldsymbol{p}} \succeq \boldsymbol{I}(\hat{\boldsymbol{p}})$, equivalent to the feasibility of Problem (13).

Next we verify that $\boldsymbol{I}(\boldsymbol{p})$ is a *standard* interference function [3, p. 1343]:

1. Positivity: $\boldsymbol{I}(\boldsymbol{p}) \succ 0$ since $\boldsymbol{v} \succ 0$.

2. Monotonicity: Note that $\boldsymbol{I}_j(\boldsymbol{p}) = \boldsymbol{v}_j + \sup_{\boldsymbol{F}_j \in \mathcal{A}_j} \boldsymbol{F}_j^{\mathrm{T}} \boldsymbol{p}$. Since every $\boldsymbol{F}$ in $\mathcal{F}$ has non-negative entries, for any $\boldsymbol{p}' \succeq \boldsymbol{p}$, we have $\boldsymbol{F}\boldsymbol{p}' \succeq \boldsymbol{F}\boldsymbol{p}$ and $\boldsymbol{I}(\boldsymbol{p}') \succeq \boldsymbol{I}(\boldsymbol{p})$.
3. Scalability: $\forall \alpha > 1$, $\forall j$,

$$\alpha \boldsymbol{I}_j(\boldsymbol{p}) = \alpha \boldsymbol{v}_j + \alpha \sup_{\boldsymbol{F}_j \in \mathcal{A}_j} \boldsymbol{F}_j^{\mathrm{T}} \boldsymbol{p} > \boldsymbol{I}_j(\alpha \boldsymbol{p}).$$

Hence (17) converges to the optimal solution of (13) both synchronously and asynchronously, by [3, Theorem 2 and 4] respectively. The uniqueness of the solution follows from [3, Theorem 1].

1) $\Leftrightarrow$ 2): This is the most involved part of the proof. See Appendix A, which also includes the proof of (19) and (20).

Remark 1 There exists a *worst* normalized channel matrix $\boldsymbol{F}^*$ in the closure of $\mathcal{F}$ (with respect to the Frobenius norm) that governs the feasibility of the robust power control problem and the convergence behavior of the iterative algorithm (17). From the proof of Theorem 1, we see that $\boldsymbol{F}^*$ admits the following characterizations:

1. $\rho(\boldsymbol{F}^*) = \sup_{\boldsymbol{F} \in \mathcal{F}} \rho(\boldsymbol{F})$.
2. The optimal robust power allocation is given by $\boldsymbol{p}^* = (\boldsymbol{I} - \boldsymbol{F}^*)^{-1}\boldsymbol{v}$. Hence $\boldsymbol{F}^*$ achieves the componentwise supremum in (19) simultaneously. Had we known the worst channel matrix $\boldsymbol{F}^*$, running the original DPC algorithm using $\boldsymbol{F}^*$ would converge to the optimal robust allocation.
3. $\boldsymbol{F}^*$ achieves the componentwise supremum $\sup_{\boldsymbol{F} \in \mathcal{F}} \boldsymbol{F}\boldsymbol{p}^*$ simultaneously.
4. A sufficient condition for the Fréchet-differentiability of $\boldsymbol{g}$ is $\mathcal{F}$ being strictly convex with non-empty interior [10, p. 107], e.g., ellipsoids. Then $\boldsymbol{F}^*$ is uniquely determined by the Jacobian matrix of $\boldsymbol{g}$ at $\boldsymbol{p}^*$:

$$\boldsymbol{F}^* = D_{\boldsymbol{p}}\boldsymbol{g}(\boldsymbol{p}^*) + \bar{\boldsymbol{F}}. \tag{22}$$

Theorem 1 implies that if the robust power control problem is feasible, the fixed-point iteration algorithm (17) will drive each user's power to the globally optimal allocation. However, one caveat is that the evaluation of the protection function of each user usually requires the power of others, leading to tight coupling among users. Nonetheless, under certain regular uncertainty sets (e.g., ellipsoid, polyhedron, or D-norm), (17) can be implemented in a distributive fashion as manifested later (in (34), (59) and (65), respectively), where only one global parameter needs to be updated.

Moreover, the technique of infrequent feedback can be used to further reduce global message passing and balance the *tradeoff* between distributiveness and robustness. More specifically, each user updates its power iteratively according to (17) but only updates the value of the protection function every M iterations, and this will not affect the convergence or optimality. In this manner we can reduce the number of message passing and hence increase the distributiveness of the algorithm. This idea is made precise by the following result.

Theorem 2 *Let $M \in \mathbb{N}$ denote the update interval. Represent any time index k as (s, l), where $k = sM + l$, $s \in \mathbb{Z}_+$ and $0 \le l \le M - 1$. Suppose the feasibility condition (15) in Theorem 1 is satisfied. Then starting from any initial point, the following iteration*

$$\boldsymbol{p}(k+1) = \bar{\boldsymbol{F}}\boldsymbol{p}(k) + \boldsymbol{v} + \boldsymbol{g}(\boldsymbol{p}(s,0)) \tag{23}$$

converges exponentially fast to the optimal solution $\boldsymbol{p}^*$ *given by (18), and the convergence rate is given by*

$$E(M,\mathcal{F}) \triangleq \lim_{k\to\infty} \log \frac{1}{\|\boldsymbol{p}(k)-\boldsymbol{p}^*\|} \tag{24}$$

$$= \frac{1}{M} \log \frac{1}{\rho\left((\boldsymbol{I}-\bar{\boldsymbol{F}})^{-1}(\boldsymbol{I}-\bar{\boldsymbol{F}}^M)(\boldsymbol{F}^*-\bar{\boldsymbol{F}})+\bar{\boldsymbol{F}}^M\right)}, \tag{25}$$

where $\boldsymbol{F}^*$ *is the worst normalized channel matrix defined in Remark 1.*

Proof See Appendix B.

Given a fixed nominal value $\bar{\boldsymbol{F}}$, the convergence rate $E(M,\mathcal{F})$ of iteration (23) is uniquely determined by the update interval M and the uncertainty set $\mathcal{F}$. We conjecture that the exponent $E(M,\mathcal{F})$ is monotonically decreasing in M as well as in $\mathcal{F}$, that is, $E(M,\mathcal{F}) \le E(M,\mathcal{G})$ if $\mathcal{G}\subset\mathcal{F}$, which is obviously true at $M=1$. It is also straightforward to verify the conjecture in the nominal case, i.e. $\mathcal{F}=\{\bar{\boldsymbol{F}}\}$, noting that $\rho(\bar{\boldsymbol{F}}^M)=\rho(\bar{\boldsymbol{F}})^M$. This conjecture agrees with various numerical evidence (see Section 4.4) and the intuition that reduced message passing and increased robust consideration necessarily incur a penalty on the convergence speed. Indeed, when $M>1$, the iterative algorithm is being conservative in the sense that the *de facto* protection function is based on the latest updated power allocation. With the most frequent feedback ($M=1$), $E(1,\mathcal{F})$ reduces to (20), and the optimality gap decays fastest according to $\rho(\boldsymbol{F}^*)^k$.

Equation (25) characterizes the tradeoff among robustness, distributiveness, and performance. To compute the function $E(M,\mathcal{F})$, we first need to find the worst channel matrix $\boldsymbol{F}^*$ using (22). To evaluate the Jacobian matrix of the protection function, however, we need to compute the optimal power allocation $\boldsymbol{p}^*$, which in general lacks closed-form expressions. On the other hand, (25) does not provide a finite-step estimate on the $\boldsymbol{p}(k)-\boldsymbol{p}^*$. Indeed, the best estimate [11, p. 306] from (25) is $\mathcal{O}\left(k^{L-1}\exp[-kE(M,\mathcal{F})]\right)$, which is still of an asymptotic nature. Therefore, it is desirable to obtain a uniform bound on $E(M,\mathcal{F})$ as well as a finite-step error estimate involving only the "size" of the uncertainty set measured by matrix norm. This is given by the following result.

Theorem 3 *Let* $\|\boldsymbol{F}\|$ *denote the induced matrix norm of* $\boldsymbol{F}$*, i.e.,*

$$\|\boldsymbol{F}\| = \max_{\|\boldsymbol{p}\|=1} \|\boldsymbol{F}\boldsymbol{p}\| \tag{26}$$

for some vector norm $\|\cdot\|$. *Let* $\sigma_{\mathcal{F}} = \sup_{F\in\mathcal{F}} \left\|\boldsymbol{F}-\bar{\boldsymbol{F}}\right\|$. *If*

$$\sigma_{\mathcal{F}} + \|\bar{\boldsymbol{F}}\| < 1, \tag{27}$$

then the convergence rate can be lower bounded as

$$E(M,\mathcal{F}) \ge \underline{E}(M,\sigma_{\mathcal{F}}) \triangleq \frac{1}{M}\log\frac{1}{\frac{1-\|\bar{\boldsymbol{F}}|^M}{1-\|\bar{\boldsymbol{F}}\|}\sigma_{\mathcal{F}}+\|\bar{\boldsymbol{F}}\|^M}. \tag{28}$$

Moreover, we have the finite-step error control:

$$\left\|\boldsymbol{p}(k)-\boldsymbol{p}^*\right\| \le \left\|\boldsymbol{p}(0)-\boldsymbol{p}^*\right\| \exp[-k\underline{E}(M,\sigma_{\mathcal{F}})]. \tag{29}$$

Proof See Appendix B.

It should be pointed out that, as a lower bound of $E(M, \mathcal{F})$, $\underline{E}(M, \sigma_{\mathcal{F}})$ is not tight. For instance, in the nominal problem when $\mathcal{F} = \{\|\bar{\boldsymbol{F}}\|\}$, $\underline{E}(M, 0) = \log \frac{1}{\|\bar{\boldsymbol{F}}\|}$, while the exact convergence rate is $E(M, \{\bar{\boldsymbol{F}}\}) = \log \frac{1}{\rho(\bar{\boldsymbol{F}})}$. Nevertheless, the following proposition shows that $\underline{E}(M, \sigma_{\mathcal{F}})$ preserves the desired behavior of $E(M, \mathcal{F})$, that is, monotonicity in distributiveness and robustness:

Proposition 2 *Under the condition (27), the lower bound of the convergence rate satisfies the following,*

1. $\underline{E}(M, \sigma_{\mathcal{F}}) > 0$*, for* $M > 0$ *and* $0 \le \sigma_{\mathcal{F}} < 1 - \|\bar{\boldsymbol{F}}\|$.
2. $\underline{E}(M, \sigma_{\mathcal{F}})$ *is strictly decreasing in* M. $\underline{E}(1, \sigma_{\mathcal{F}}) = \log \frac{1}{\sigma_{\mathcal{F}} + \|\bar{\boldsymbol{F}}\|}$ *and*

$$\lim_{M \to \infty} \underline{E}(M, \sigma_{\mathcal{F}}) = 0.$$

3. $\underline{E}(M, \sigma_{\mathcal{F}})$ *is strictly decreasing in* $\sigma_{\mathcal{F}}$. $E(M, 0) = \log \frac{1}{\|\bar{\boldsymbol{F}}\|}$ *and*

$$\lim_{\sigma_{\mathcal{F}} \to 1 - \|\bar{\boldsymbol{F}}\|} \underline{E}(M, \sigma_{\mathcal{F}}) = 0.$$

Proof See Appendix C.

In the sequel, we analyze the necessary and sufficient conditions for the feasibility of the robust power control problem under different practically-driven uncertainty models, as well as various tradeoff among robustness, distributiveness, performance, and the total power consumption.

4 Modeling Channel Uncertainty by Ellipsoids

In this subsection, the uncertainty in normalized channel matrix $\boldsymbol{F}$ due to fluctuation of the channels is modeled by ellipsoids. Let $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_L]^{\mathrm{T}} \succeq 0$. The uncertainty set $\mathcal{F}_i$ for $\boldsymbol{F}_i$ under ellipsoid approximation can be represented by

$$\mathcal{F}_i = \Big\{ \bar{\boldsymbol{F}}_i + \Delta \boldsymbol{F}_i : \sum_{j \ne i} |\Delta F_{ij}|^2 \le \epsilon_i^2 \Big\}, \tag{30}$$

where ϵ_i denotes the maximal deviation of each entry in $\boldsymbol{F}_i$.

As derived in [1, Section II.C], the protection function associated with (30) is

$$g_i^{\mathrm{ell}}(\boldsymbol{p}) = \epsilon_i \sqrt{\sum_{j \ne i} p_j^2} = \epsilon_i \sqrt{\|\boldsymbol{p}\|_2^2 - p_i^2}, \tag{31}$$

and the robust version of the power control problem (5) is

$$\begin{array}{rl} \text{minimize} & \mathbf{1}^{\mathrm{T}} \boldsymbol{p} \\ \text{subject to} & (\boldsymbol{I} - \bar{\boldsymbol{F}})\boldsymbol{p} \succeq \boldsymbol{v} + \boldsymbol{g}^{\mathrm{ell}}(\boldsymbol{p}), \\ \text{variables} & \boldsymbol{p} \succeq 0. \end{array} \tag{32}$$

This is the key problem that we will solve in this subsection. We recognize it as an SOCP (Second-Order Cone Programming) problem [12], which can be rewritten in the standard form as

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^{\mathrm{T}}\boldsymbol{p} \\ \text{subject to} & \epsilon_i\sqrt{\boldsymbol{p}^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{e_i}\boldsymbol{e_i}^{\mathrm{T}})\boldsymbol{p}} \leq (\boldsymbol{e_i}-\bar{\boldsymbol{F}}_i)^{\mathrm{T}}\boldsymbol{p} - v_i, \\ & i=1,\ldots,L \\ \text{variables} & \boldsymbol{p} \succeq 0, \end{array} \tag{33}$$

where $\boldsymbol{e}_i$ denotes the i^{th} standard basis vector.

4.1 Optimality Conditions and Distributed Algorithms

Plugging (31) into (17), we obtain the following distributive algorithm:

Algorithm 1 *For $i=1,\ldots,L$, at each time slot k, user i updates its transmission power according to*

$$p_i(k+1) = \frac{\gamma_i}{\mathsf{SIR}_i}p_i(k) + \epsilon_i\sqrt{Q^2(k) - p_i^2(k)} \tag{34}$$

where

$$Q(k) = \|\boldsymbol{p}(k)\|_2\,. \tag{35}$$

Note that (34) is a generalization of the classical DPC algorithm in (7) with a protection function $g_i^{\mathrm{ell}}(\boldsymbol{p}(k)) = \epsilon_i\sqrt{Q^2(k)-p_i^2(k)}$ added for each user i. It reduces to the DPC algorithm when $\boldsymbol{\epsilon}=0$, i.e., no robustness is taken into consideration. Similarly to the DPC algorithm, it is a primal-based algorithm involving no diminishing step size or dual variables. The only coupled parameter is $Q(k)$, the ℓ_2-norm of the user power vector $\boldsymbol{p}(k)$, which needs to be updated at every time slot k.

By Theorem 1, the problem of feasibility under ellipsoid uncertainty set as well as convergence and optimality for Algorithm 1 boils down to computing the maximal spectral radius in the uncertainty set, which is not easily obtainable in closed form. Instead we give a upper bound which leads to the following sufficient condition:

Theorem 4 *Problem (32) is feasible, or equivalently, Algorithm 1 converges to the optimal solution of Problem (32), denoted by $\boldsymbol{p}^*$, if*

$$\min\left\{\left\|\bar{\boldsymbol{F}}\right\|_2, \frac{\rho(\bar{\boldsymbol{F}}+\bar{\boldsymbol{F}}^{\mathrm{T}})}{2}\right\} + \|\boldsymbol{\epsilon}\|_2 < 1, \tag{36}$$

where $\left\|\bar{\boldsymbol{F}}\right\|_2$ is the induced ℓ_2-norm of $\bar{\boldsymbol{F}}$. If $\bar{\boldsymbol{F}}$ is symmetric, (36) reduces to

$$\|\boldsymbol{\epsilon}\|_2 + \rho(\bar{\boldsymbol{F}}) < 1. \tag{37}$$

Proof See Appendix D.

Remark 2 From Levinger's inequality [13],

$$\frac{\rho(\bar{\boldsymbol{F}} + \bar{\boldsymbol{F}}^{\mathrm{T}})}{2} \geq \rho(\bar{\boldsymbol{F}}), \tag{38}$$

we see that $\rho(\bar{\boldsymbol{F}})$ is a lower bound to the sufficient condition in Theorem 4. In a way it characterizes the maximal robustness the system can provide against perturbations. This is in accordance with the notion of stability in matrix theory. When $\rho(\bar{\boldsymbol{F}})$ approaches one, the iterative algorithm converges more slowly and the system becomes less resilient to noise or perturbation in the channel matrix.

4.2 Infrequent Feedback

According to Theorem 2, we can allow Q in Algorithm 1 to be updated at a slower pace without affecting the convergence or optimality, hence the following algorithm:

Algorithm 2 *Let $M \geq 1$ be the number of time slots between two adjacent updates of Q. For $i = 1, \ldots, L$, at each time slot $k = (s, l)$, user i updates its transmission power according to*

$$p_i(k+1) = \frac{\gamma_i}{\mathsf{SIR}_i} p_i(k) + \epsilon_i \sqrt{Q^2(s,0) - p_i^2(s,0)}. \tag{39}$$

In Algorithm 2, user i broadcasts its power p_i and computes Q using (35) every M iterations. Between the updates of Q, each user updates its power based on the most recently computed protection function. For the downlink transmission in a single cell network, Q can be simply broadcast by the base station.

Due to the smoothness of $\boldsymbol{g}^{\mathrm{ell}}$, Algorithm 2 converges exponentially fast to the optimal solution of Problem (32) under the feasibility condition. Since the ellipsoid $\mathcal{F}$ is completely determined by $\boldsymbol{\epsilon}$, we denote the convergence rate by

$$E(M, \boldsymbol{\epsilon}) = \frac{1}{M} \log \frac{1}{\rho\left((\boldsymbol{I} - \bar{\boldsymbol{F}})^{-1}(\boldsymbol{I} - \bar{\boldsymbol{F}}^{M}) D_{\boldsymbol{p}} \boldsymbol{g}^{\mathrm{ell}}(\boldsymbol{p}^*) + \bar{\boldsymbol{F}}^{M}\right)}, \tag{40}$$

where

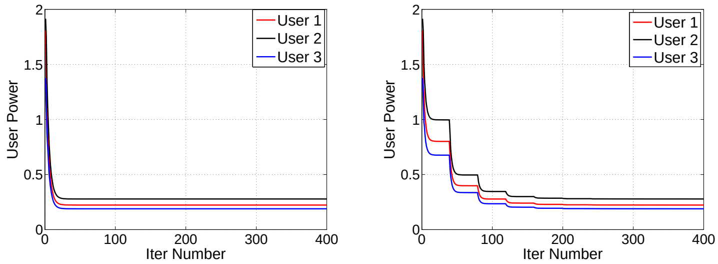
$$\left[D_{\boldsymbol{p}} \boldsymbol{g}^{\mathrm{ell}}(\boldsymbol{p})\right]_{ij} = \frac{\partial g_i^{\mathrm{ell}}}{\partial p_j}(\boldsymbol{p}) = \begin{cases} 0 & i = j, \\ \frac{\epsilon_i p_j}{\sqrt{\|\boldsymbol{p}\|_2^2 - p_i^2}} & i \neq j. \end{cases} \tag{41}$$

In particular, the convergence rate of Algorithm 1 is given by $E(1, \boldsymbol{\epsilon}) = \log \frac{1}{\rho(\boldsymbol{F}^*)}$ where the worst channel matrix is given by $\boldsymbol{F}^* = \bar{\boldsymbol{F}} + D_{\boldsymbol{p}} \boldsymbol{g}^{\mathrm{ell}}(\boldsymbol{p}^*)$.

In view of the upper bound

$$\sigma_{\mathcal{F}} = \sup_{F \in \mathcal{F}} \left\|\boldsymbol{F} - \bar{\boldsymbol{F}}\right\|_2 \leq \sup_{F \in \mathcal{F}} \left\|\boldsymbol{F} - \bar{\boldsymbol{F}}\right\|_{\mathrm{F}} = \|\boldsymbol{\epsilon}\|_2, \tag{42}$$

particularizing Theorem 3 to the ellipsoid uncertainty sets gives the following result, which controls the transient behavior of Algorithm 2:



(a) Without feedback delay ($M = 1$) (b) With feedback delay ($M = 40$).

Fig. 1 Convergence of Algorithm 2.

Theorem 5 *Assume*

$$\|\boldsymbol{\epsilon}\|_2 + \left\|\bar{\boldsymbol{F}}\right\|_2 < 1. \tag{43}$$

Then Algorithm 2 satisfies

$$\left\|\boldsymbol{p}(k) - \boldsymbol{p}^*\right\|_2 \leq \left\|\boldsymbol{p}(0) - \boldsymbol{p}^*\right\|_2 \exp[-k\underline{E}(M, \|\boldsymbol{\epsilon}\|_2)]. \tag{44}$$

where

$$\underline{E}(M, \|\boldsymbol{\epsilon}\|_2) = -\frac{1}{M} \log \left(\left\|\bar{\boldsymbol{F}}\right\|_2^M + \frac{\|\boldsymbol{\epsilon}\|_2}{1 - \left\|\bar{\boldsymbol{F}}\right\|_2} (1 - \left\|\bar{\boldsymbol{F}}\right\|_2^M) \right) \tag{45}$$

is a lower bound on the convergence rate $E(M, \boldsymbol{\epsilon})$.

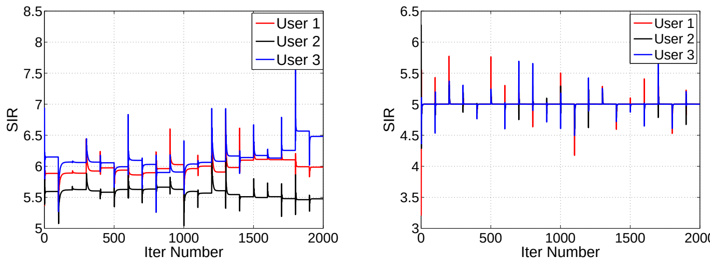
4.3 Numerical Results and Performance Comparison

We simulate the performance of Algorithm 2 for 3 users with Rayleigh fading channel. The channel uncertainty parameter $\epsilon = 5\%$, and the common target SIR $\gamma = 5.0$. Fig. 1 shows the results without ($M = 1$) and with infrequent feedback ($M = 40$). In both cases, the algorithm converges to the optimal solution (verified by a centralized algorithm using the MOSEK toolbox [14]) exponentially fast.

We also compare the performance of Algorithm 2 and the original DPC algorithm in terms of the immunity against channel fluctuation. The simulation setup is the same as in Fig. 1, where the channel matrix changes randomly for twenty times. We define a channel outage whenever a user's received SIR drops below the target SIR γ. As shown in Fig. 2, Algorithm 2 avoids channel outage since it is designed for the worst channel in the uncertainty set, while the original DPC algorithm leads to frequent channel outages.

4.4 Robustness-Distributiveness Tradeoff

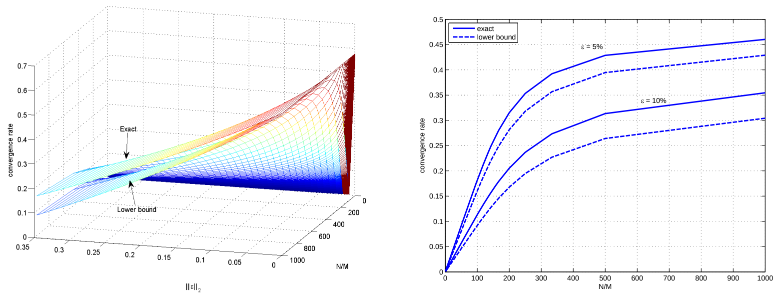
If we fix the total number of iterations as N and the desired optimality gap $\|\boldsymbol{p}(N) - \boldsymbol{p}^*\|_2 = \delta$, then there exists an interesting tradeoff between robustness and distributiveness. In particular, if more robustness is desired (i.e., a lager $\|\boldsymbol{\epsilon}\|_2$), we will have more frequent message passing and less distributiveness (i.e., a smaller number of



(a) Algorithm 2 avoids channel outage.

(b) DPC algorithm leads to frequent channel outage.

Fig. 2 User received SIRs under random channel fluctuation (only one channel realization is shown here).



(a) Exact exponent $E(M, \boldsymbol{\epsilon})$ versus lower bound $\underline{E}(M, \|\boldsymbol{\epsilon}\|_2)$.

(b) $\epsilon = 5\%$ and 10%.

Fig. 3 Convergence rate of Algorithm 2 against robustness $\|\boldsymbol{\epsilon}\|_2$ and number of message passing N/M, where $\boldsymbol{\epsilon} = \epsilon\mathbf{1}$ and $N = 1000$.

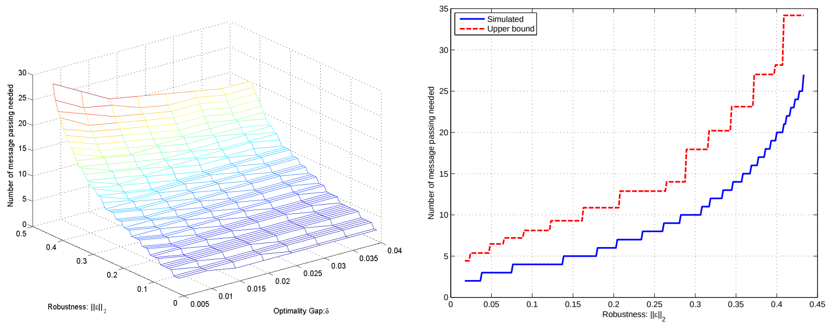
update interval M). Here one message passing corresponds to each user announcing his power level once, or the base station evaluating the current Q in (35) and broadcasting it to the users.

The convergence rate $E(M, \boldsymbol{\epsilon})$ and lower bound $\underline{E}(M, \|\boldsymbol{\epsilon}\|_2)$ characterize the tradeoff between distributiveness $\frac{N}{M}$ and robustness $\|\boldsymbol{\epsilon}\|_2$, as plotted in Fig. 3. For notational simplicity assume that $\boldsymbol{\epsilon} = \epsilon\mathbf{1}$. For a given robustness requirement ϵ, Algorithm 2 converges slower as less messages are passed. In the limit when M tends to infinity, the convergence rate vanishes. For a fixed number of message passing $\frac{N}{M}$, i.e., fixed M, the speed of convergence of Algorithm 2 decreases as ϵ increases, that is, more robustness are taken into account in the power allocation.

By (44), a sufficient condition to achieve the desired optimality gap δ is

$$\underline{E}(M, \|\boldsymbol{\epsilon}\|_2) \geq \frac{1}{N} \log\left(\frac{\|\boldsymbol{p}(0) - \boldsymbol{p}^*\|_2}{\delta}\right). \tag{46}$$

By Proposition 2, $\underline{E}(M, \|\boldsymbol{\epsilon}\|_2)$ is strictly monotonic in M, and its inverse function is denoted by $\underline{E}^{-1}$. We can solve (46) and obtain a lower bound on the largest allowed



(a) Three-dimensional tradeoff surface between robustness $\|\boldsymbol{\epsilon}\|_2$, optimality gap δ and the number of message passing.

(b) Robustness-distributiveness tradeoff curve when $\delta = 1\%$.

Fig. 4 Robustness-distributive-performance tradeoff under ellipsoid uncertainty sets.

value of update interval M:

$$M \geq \underline{E}^{-1}\left[\frac{1}{N}\log\left(\frac{\|\boldsymbol{p}(0)-\boldsymbol{p}^*\|_2}{\delta}\right)\right], \tag{47}$$

which leads to the following upper bound on the total number of message passing for reaching a prescribed optimality gap δ with an uncertainty ellipsoid of parameter ϵ and a total of N iterations

$$\frac{N}{\underline{E}^{-1}\left[\frac{1}{N}\log\left(\frac{\|\boldsymbol{p}(0)-\boldsymbol{p}^*\|_2}{\delta}\right)\right]}. \tag{48}$$

This upper bound is plotted in Fig. 4(b), together with the simulated result, which accurately portrays the growth of the number of message passing dictated by increasing robustness considerations. We also see a clear tradeoff between the robustness and the distributiveness. As the power allocation becomes more robust, more message passing among users is necessary. For example, for an error threshold of $\delta = 1\%$, only 4 global message passing is needed for $\epsilon = 5\%$, while 15 messages must be passed in order to achieve the robustness $\epsilon = 20\%$. The simulated three-dimensional tradeoff among robustness ϵ, optimality gap δ, and the number of message passing is given in Fig. 4(a).

4.5 Robustness-Power Tradeoff

The tradeoff between increased power consumption and the robustness are shown in Fig. 5, with comparison with the total power allocation of DPC algorithm. Note here the allocated power are the solution of Problem (32) as well as the steady state of Algorithm 2. The power-robustness tradeoff can be analyzed as follows. Let $\bar{\boldsymbol{p}}$ and $\boldsymbol{p}^*$ denote the solution to the nominal problem (2) and the robust problem (34) respectively. Then

$$\bar{\boldsymbol{p}} = \bar{\boldsymbol{F}}\bar{\boldsymbol{p}} + \boldsymbol{v}, \tag{49}$$

$$\boldsymbol{p}^* = \bar{\boldsymbol{F}}\boldsymbol{p}^* + \boldsymbol{g}^{\text{ell}}(\boldsymbol{p}^*) + \boldsymbol{v}. \tag{50}$$

where $\boldsymbol{g}^{\mathrm{ell}}$ is in (31). Let $\boldsymbol{p}^* = \bar{\boldsymbol{p}} + \Delta\boldsymbol{p}$, and $\mathbf{1}^{\mathrm{T}}\Delta\boldsymbol{p}$ is the extra power consumption. From (49) and (50) we have $\bar{\boldsymbol{p}} + \Delta\boldsymbol{p} = \bar{\boldsymbol{F}}\bar{\boldsymbol{p}} + \bar{\boldsymbol{F}}\Delta\boldsymbol{p} + \boldsymbol{g}^{\mathrm{ell}}(\bar{\boldsymbol{p}} + \Delta\boldsymbol{p}) + \boldsymbol{v}$, therefore $(\boldsymbol{I} - \bar{\boldsymbol{F}})\Delta\boldsymbol{p} = \boldsymbol{g}^{\mathrm{ell}}(\bar{\boldsymbol{p}} + \Delta\boldsymbol{p}) = \boldsymbol{g}^{\mathrm{ell}}(\bar{\boldsymbol{p}}) + D_{\boldsymbol{p}}\boldsymbol{g}^{\mathrm{ell}}(\bar{\boldsymbol{p}}) + \mathcal{O}\left(|\Delta\boldsymbol{p}|^2\right)$, where $D_{\boldsymbol{p}}\boldsymbol{g}^{\mathrm{ell}}(\bar{\boldsymbol{p}})$ is given by (41). Hence a good estimate of the total power allocation is

$$\mathbf{1}^{\mathrm{T}}\Delta\boldsymbol{p} \approx \mathbf{1}^{\mathrm{T}}(\boldsymbol{I} - \bar{\boldsymbol{F}} - D_{\boldsymbol{p}}\boldsymbol{g}^{\mathrm{ell}}(\bar{\boldsymbol{p}}))^{-1}\boldsymbol{g}^{\mathrm{ell}}(\bar{\boldsymbol{p}}). \tag{51}$$

Assume (43) holds. Then $\rho(\bar{\boldsymbol{F}} + D_{\boldsymbol{p}}\boldsymbol{g}^{\mathrm{ell}}(\bar{\boldsymbol{p}})) \leq \left\|\bar{\boldsymbol{F}} + D_{\boldsymbol{p}}\boldsymbol{g}^{\mathrm{ell}}(\bar{\boldsymbol{p}})\right\|_2 < 1$. Hence

$$\begin{aligned}\mathbf{1}^{\mathrm{T}}\Delta\boldsymbol{p} &\approx \mathbf{1}^{\mathrm{T}}(\boldsymbol{I} - \bar{\boldsymbol{F}} - D_{\boldsymbol{p}}\boldsymbol{g}^{\mathrm{ell}}(\bar{\boldsymbol{p}}))^{-1}\boldsymbol{g}^{\mathrm{ell}}(\bar{\boldsymbol{p}}) = \mathbf{1}^{\mathrm{T}}\sum_{k=0}^{\infty}(\bar{\boldsymbol{F}} + D_{\boldsymbol{p}}\boldsymbol{g}^{\mathrm{ell}}(\bar{\boldsymbol{p}}))^k\boldsymbol{g}^{\mathrm{ell}}(\bar{\boldsymbol{p}}) \\ &= \mathbf{1}^{\mathrm{T}}\left(\sum_{k=0}^{\infty}\bar{\boldsymbol{F}}^k\right)\boldsymbol{g}^{\mathrm{ell}}(\bar{\boldsymbol{p}}) + \mathcal{O}\left(\|\boldsymbol{\epsilon}\|_2^2\right) = \mathbf{1}^{\mathrm{T}}(\boldsymbol{I} - \bar{\boldsymbol{F}})^{-1}\boldsymbol{g}^{\mathrm{ell}}(\bar{\boldsymbol{p}}) + \mathcal{O}\left(\|\boldsymbol{\epsilon}\|_2^2\right).\end{aligned}$$

In view of (31), we notice that the first term dominates when $\|\boldsymbol{\epsilon}\|_2$ is small, and the extra power consumption scales with robustness as $\mathcal{O}\left(\|\boldsymbol{\epsilon}\|_2\right)$. The approximation of total power consumption based on (51) is also plotted in Fig. 5 along with the power-robustness tradeoff curve. The transient behavior is shown in Fig. 6. We see that the close-form approximation in (51) fits well with the numerical calculation.

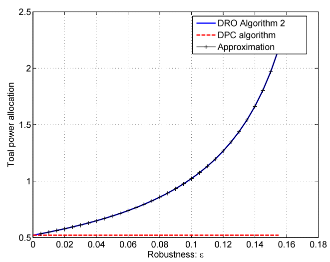
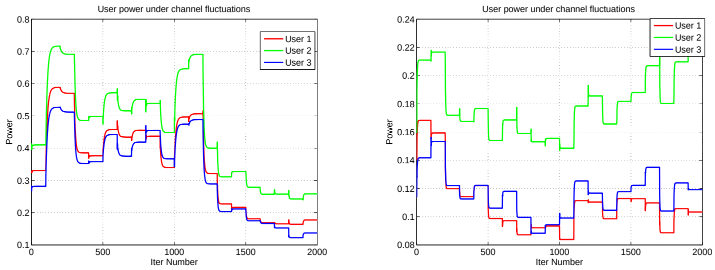


Fig. 5 Robustness-power tradeoff curve under ellipsoid uncertainty set, with approximation calculated based on (51).

Remark 3 It has been shown that the original distributed power control algorithm in [4] is essentially a Jacobi method which aims to solve a linear system [15]. Similarly, the distributed robust power control algorithm can be viewed as a nonlinear Jacobi method. Assume $[p_1(k), p_2(k), ..., p_L(k)]^{\mathrm{T}}$ is a feasible solution to (33), then algorithm (34) actually solves a one-dimensional optimization problem with respect to the power of user i, assuming the power levels of other users are fixed. If all users update power levels in a synchronous manner, the resulting algorithm is a Jaboci method. If we allow users to update their power levels asynchronously, the obtained algorithm can be viewed as a Gauss-Seidel-type method.



(a) Robust algorithm.

(b) DPC algorithm.

Fig. 6 Users' transmission power when the channels undergo random fluctuation in the uncertainty sets for 20 times ($\epsilon = 10\%$).

5 Modeling SIR Measurement Inaccuracy using Polyhedrons

Now we turn to another type of robustness desirable in power control. To model SIR measurement errors, consider a row-wise uncertainty polyhedron

$$\mathcal{F}_i = \left\{ \bar{\boldsymbol{F}}_i + \Delta \boldsymbol{F}_i : \left| \sum_{j \neq i} \frac{\Delta F_{ij}}{t_{ij}} \right| \leq 1 \right\}, \tag{52}$$

where $t_{ij} > 0$ parametrizes the maximal deviation of F_{ij} from its nominal value. It plays the same role of controlling the size of the uncertainty set as $\boldsymbol{\epsilon}$ does in the ellipsoidal case. This uncertainty setup will be very useful in modeling inaccurate SIR measurements, in that the relative error can be equivalently characterized by a uncertainty polyhedron in terms of the normalized channel matrix $\boldsymbol{F}$. Assume that user i measures the SIR at its receiver with relative error Δ_i, that is,

$$\frac{\mathsf{SIR}_i}{\overline{\mathsf{SIR}}_i} \in [1 - \Delta_i, 1 + \Delta_i]. \tag{53}$$

where $\overline{\mathsf{SIR}}_i$ is the nominal value corresponding to $\bar{\boldsymbol{F}}$. By (4), (53) is equivalent to

$$\frac{\sum_{j \neq i} F_{ij} p_j + v_i}{\sum_{j \neq i} \bar{F}_{ij} p_j + v_i} \in [1 - \Delta_i, 1 + \Delta_i], \tag{54}$$

that is,

$$\left| \sum_{j \neq i} \Delta F_{ij} p_j \right| \leq \Delta_i \Big(\sum_{j \neq i} \bar{F}_{ij} p_j + v_i \Big) = \frac{\Delta_i p_i \gamma_i}{\overline{\mathsf{SIR}}_i}. \tag{55}$$

Hence we see that (53) is a polyhedron in (52) with coefficients

$$t_{ij} = \frac{\Delta_i p_i \gamma_i}{\overline{\mathsf{SIR}}_i p_j}. \tag{56}$$

Next we analyze the corresponding protection function of (52) and the distributed solutions. Note that (52) is similar to (30) with quadratic terms replaced by linear terms. Since

$$\sum_{j \neq i} \Delta F_{ij} p_j = \sum_{j \neq i} (\Delta F_{ij} / t_{ij}) t_{ij} p_j \leq \max_{j \neq i} t_{ij} p_j,$$

with equality achieved by

$$\Delta F_{ij} = \begin{cases} t_{ij} & j = \arg\max_{j \neq i} t_{ij} p_j, \\ 0 & \text{else,} \end{cases}$$

the protection function corresponding to the uncertainty sets in (52) is

$$g_i^{\text{poly}}(\boldsymbol{p}) = \max_{j \neq i} t_{ij} p_j, \quad i = 1, \ldots, L, \tag{57}$$

and the robust version of Problem (5) is given by

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^{\mathrm{T}} \boldsymbol{p} \\ \text{subject to} & \boldsymbol{p} \succeq \bar{\boldsymbol{F}} \boldsymbol{p} + \boldsymbol{g}^{\text{poly}}(\boldsymbol{p}) + \boldsymbol{v} \\ \text{variables} & \boldsymbol{p} \succeq 0. \end{array} \tag{58}$$

A distributed algorithm with limited message passing can be derived similar to Algorithm 2.

Algorithm 3 *User i updates its power at time k accordingly to*

$$p_i(k+1) = \frac{\gamma_i}{\mathsf{SIR}_i} p_i(k) + \max_{j \neq i} t_{ij} p_j(k). \tag{59}$$

When each user allocates the same uncertainty for other user's channel gain, i.e., $t_{ij} = t_i$ for each i, (59) simplies to

$$p_i(k+1) = \frac{\gamma_i}{\mathsf{SIR}_i} p_i(k) + t_i \max_{j \neq i} p_j(k). \tag{60}$$

Therefore, only the largest two value of user's weighted power need to be communicated globally, through broadcasting of the users. However, in the context of modeling inaccurate SIR measurements, Algorithm 3 can be implemented in a full distributed fashion. Plugging (56) into (59), we have

$$\begin{aligned} p_i(k+1) &= \frac{\gamma_i}{\mathsf{SIR}_i} p_i(k) + \max_{j \neq i} \frac{\Delta_i p_i(k) \gamma_i}{\mathsf{SIR}_i} \\ &= \frac{\gamma_i}{\mathsf{SIR}_i}(1 + \Delta_i) p_i(k) \approx \frac{\gamma_i}{\mathsf{SIR}_i(1 - \Delta_i)} p_i(k), \end{aligned} \tag{61}$$

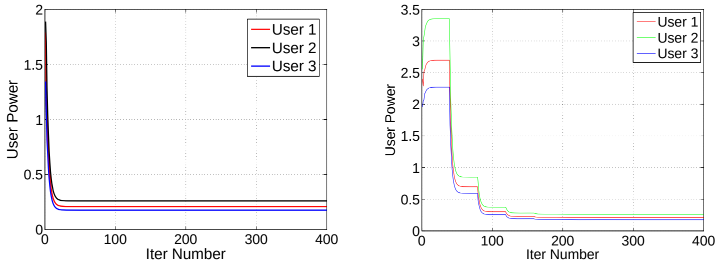
where the last approximation are made when Δ_i is small. Note that this corresponds to each user running DPC algorithm in a conservative way, that is, assuming the largest measurement error and use the worst SIR value $\mathsf{SIR}_i(1 - \Delta_i)$ in the power update, which agrees with our intuition about the worst-case robust formulation.

Similar to Theorem 4, we derive the following sufficient condition for feasibility under polyhedron uncertainty set and convergence of Algorithm 3 to the optimality:

Theorem 6 *Problem (58) is feasible, or equivalently, Algorithm 3 converges to the optimal solution of Problem (58) if*

$$\left\| \bar{\boldsymbol{F}} \right\|_1 + \max_j \sum_i t_{ij} < 1, \tag{62}$$

where $\left\| \bar{\boldsymbol{F}} \right\|_1$ is the induced ℓ_1-norm of $\bar{\boldsymbol{F}}$.



(a) Convergence of Algorithm 3 without feedback delay.

(b) Convergence of Algorithm 3 where messages are passed every 40 iterations.

Fig. 7 Convergence of Algorithm 3 with and without feedback delay.

Proof See Appendix E.

Simulation results of Algorithm 3 are shown in Fig. 7, with a similar setup as in Fig. 1. The robustness-distributiveness tradeoff curve is plotted in Fig. 8. For computational convenience, we choose $t_{ij} = \epsilon_i$ and $\epsilon = \max \epsilon_i$. The optimality gap $\delta = 0.5$ is chosen to be the same as in Fig. 4(b). The tradeoff curve is less smooth compared to Fig. 4(b), due to linear programming nature of Problem (58), whereas Problem (32) is an SOCP. Also observe that the algorithm for polyhedrons converges faster and needs less message passing than that for ellipsoids.

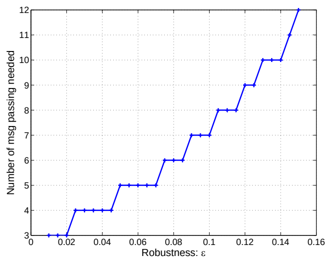


Fig. 8 Robustness-distributiveness tradeoff under polyhedron uncertainty set, where $\epsilon = \max t_{ij}$.

6 Modeling User Dynamics Using D-norm

It is also possible to use D-norm [16] to model the uncertainty due to both channel fluctuations and users randomly entering the system. D-norm is a norm on $\mathbb{R}^L$ defined as $\langle \boldsymbol{y} \rangle_k = \max\{\sum_{i \in S} |y_i| : S \subset \{1, \ldots, L\}, |S| \leq k\}$, where $k \in \{1, \ldots, L\}$, i.e., the sum of the k largest components in $|\boldsymbol{y}|$. It is a generalization of the ℓ_∞ norm, which corresponds to $k = 1$.

Let L and V be the total number of active and possible *virtual* users (i.e., users who are not active but might turn active) in the system respectively. Let $h_{iv} \in [0, \hat{h}_{iv}]$ denote the relative channel gain (normalized by G_{ii}) of virtual user v's transmitter to active user i's receiver. $\boldsymbol{\Delta F}$ denotes the variation of $\boldsymbol{F}$. Also denote by $\bar{p}_v^{\max}$ the upper bound of the transmission power from the v^{th} virtual user. Consider the following protection function for the i^{th} constraint with uncertainty parameter Γ_i:

$$g_i^D(\Gamma_i, \boldsymbol{p}) = \left\langle [\Delta F_{i1} p_1, \ldots, \Delta F_{iL} p_L, \hat{h}_{i1} \bar{p}_1^{\max}, \ldots, \hat{h}_{iV} \bar{p}_V^{\max}]^{\mathrm{T}} \right\rangle_{\Gamma_i}, \tag{63}$$

which models the total effect of the Γ_i largest impacts from emerging users and channel variations. Note that the above maximization can be easily solved due to its special structure. For any given $\boldsymbol{p}$, we only need to sort $p_j |\Delta F_{ij}|$ and $\hat{h}_{iv} \bar{p}_v^{\max}$ for all j and v in the descending order, and sum over the first Γ_i elements to obtain $g_i^D(\Gamma_i, \boldsymbol{p})$. This can be done at the base station and sent to each mobile station. The robust power control problem under both channel fluctuations and user uncertainties as follows:

$$\begin{array}{rl} \text{minimize} & \mathbf{1}^{\mathrm{T}} \boldsymbol{p} \\ \text{subject to} & \boldsymbol{p} \succeq \bar{\boldsymbol{F}} \boldsymbol{p} + \boldsymbol{g}^D(\boldsymbol{\Gamma}, \boldsymbol{p}) + \boldsymbol{v} \\ \text{variables} & \boldsymbol{p} \succeq 0. \end{array} \tag{64}$$

which can be solved by the following distributed algorithm.

Algorithm 4 *User i updates its transmission power in time slot k as*

$$p_i(k+1) = \frac{\gamma_i}{\mathsf{SIR}_i} p_i(k) + g_i(\Gamma_i, \boldsymbol{p}(k)), \tag{65}$$

where $g_i(\Gamma_i, \boldsymbol{p})$ is computed by the basestation and broadcast to users via the downlink channel.

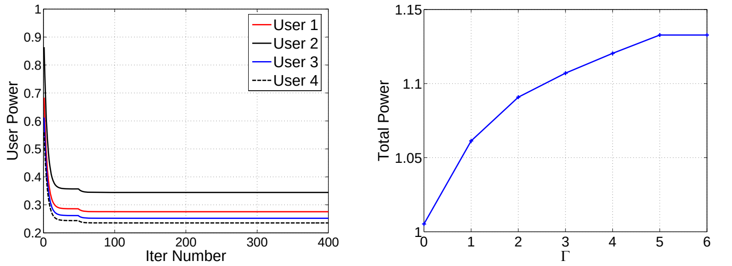
Theorem 7 *Algorithm 4 converges to the optimal solution of Problem (64) if*

$$\|\bar{\boldsymbol{F}}\|_2 + \sqrt{\|\boldsymbol{\Gamma}\|_\infty} \, \|\boldsymbol{\Delta F}\|_{\mathrm{F}} < 1, \tag{66}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm.

Proof See Appendix F.

Note that the convergence of Algorithm 4 does not depend on the number or parameter of virtual users. As an example, we simulate a network with four active users and two virtual users. The convergence of the four active users' power levels are shown in Fig. 9, together with the energy-robustness tradeoff with different values of Γ (for notational simplicity, we let $\Gamma_i = \Gamma$ for all i). Note that the system becomes more robust as Γ increases, but incurs a penalty in the total power consumption.



(a) Convergence of Algorithm 4, where $\Gamma = 3$ and feedback frequency 1/50 iterations.

(b) Performance-robustness tradeoff curve, with $\Gamma = 0$ corresponding to the nominal case.

Fig. 9 Numerical results for Algorithm 4.

7 Concluding remarks

Distributed Robust Optimization (DRO) is an approach that combines robustness and distributiveness into optimization models. In this paper we studied the robust power control problem under the DRO framework and general row-wise uncertainty sets, then particularized our results to uncertainty models induced by channel variations, SIR measurement errors, and user dynamics. Since the channel gain of each user's receiver varies independently, the uncertainty in the channel matrix is well modeled by a row-wise uncertainty set. This structure of the channel matrix is crucial to the derivation of our main results and algorithms. It guarantees that the robust formulation falls gracefully into the standard interference function framework, hence distributiveness-preserving.

Under the general setup, the dependence of convergence rate $E(M, \mathcal{F})$ of fixed-point iterative algorithms on the message passing frequency and uncertainty set is completely characterized, which represents the fundamental tradeoff among robustness, distributiveness, and performance. It is of particular interest to prove (or disprove) the conjecture in Section 3 about monotonicity of $E(M, \mathcal{F})$ in M and $\mathcal{F}$. In addition, we also developed upper bounds for the maximal spectral radius under various robust models, hence sufficient for optimality. It is an interesting matrix theoretical problem to exactly evaluate the maximal spectral radius for typical uncertainty sets.

Apart from the worst-case methodology adopted in this paper, chance-constrained formulations are also possible by incorporating stochastic channel model, where robustness can be measured by the outage probability, i.e., the probability of SIR dropping below the threshold. The distribution of the spectral radius of the channel matrix is closely related to the chance of convergence and optimality of the DRO algorithm, leading to tractable analysis of robustness-distributive tradeoffs.

A Proof of Theorem 1

Proof We complete the proof of Theorem 1 by showing 1) ⇔ 2) and (19).

1) ⇒ 2): Suppose $\exists \hat{\boldsymbol{p}} \in \mathbb{R}_+$, such that $\forall \boldsymbol{F} \in \mathcal{F}$,

$$\hat{\boldsymbol{p}} \succeq \boldsymbol{F}\hat{\boldsymbol{p}} + \boldsymbol{v}. \tag{67}$$

First we show $\mathcal{F}$ is bounded in the Frobenius norm. Suppose otherwise. Then there exists a sequence $\{\boldsymbol{F}^k\}$ in $\mathcal{F}$ such that $\left\|\boldsymbol{F}^k\right\|_{\mathrm{F}} \to \infty$, then there exists some index (i,j) such that $\boldsymbol{F}^k_{i,j} \to \infty$. Since $\hat{\boldsymbol{p}} \succeq \boldsymbol{v} \succ 0$, we have $[\boldsymbol{F}^k \hat{\boldsymbol{p}}]_j \to \infty$, hence $\hat{p}_j = \infty$, contradiction.

Next we show

$$\sup_{\boldsymbol{F}\in\mathcal{F}} \rho(\boldsymbol{F}) < 1. \tag{68}$$

Note that (67) implies that $\max_j \frac{(\boldsymbol{F}\hat{\boldsymbol{p}})_j}{\hat{p}_j} \le 1 - \min_j \frac{v_j}{\hat{p}_j}$. Let $c = \min_j \frac{v_j}{\hat{p}_j}$. Then $0 < c \le 1$ since $0 \prec \boldsymbol{v} \preceq \hat{\boldsymbol{p}}$. On the other hand, the spectral radius of nonnegative matrices admits the following minimax characterization [17, Exercise 2.3.2]:

$$\rho(\boldsymbol{F}) = \min_{\boldsymbol{p}\succeq 0} \max_j \frac{(\boldsymbol{F}\boldsymbol{p})_j}{p_j}. \tag{69}$$

Hence $\rho(\boldsymbol{F}) < 1 - c$, $\forall \boldsymbol{F} \in \mathcal{F}$. Therefore $\sup_{\boldsymbol{F}\in\mathcal{F}} \rho(\boldsymbol{F}) \le 1 - c < 1$.

2) $\Rightarrow$ 1): To show feasibility, it is sufficient to show (19), that is, the optimal power allocation, i.e., the solution of

$$\boldsymbol{p} = \bar{\boldsymbol{F}}\boldsymbol{p} + \boldsymbol{g}(\boldsymbol{p}) + \boldsymbol{v} = \sup_{\boldsymbol{F}\in\mathcal{F}} \boldsymbol{F}\boldsymbol{p} + \boldsymbol{v} \tag{70}$$

is given by

$$\boldsymbol{p}^* = \sup_{\boldsymbol{F}\in\mathcal{F}} (\boldsymbol{I} - \boldsymbol{F})^{-1}\boldsymbol{v}. \tag{71}$$

where the supremum above is taken componentwise.

Denote $\overline{\mathcal{F}}$ the closure of $\mathcal{F}$ (in the topology on all $L \times L$ matrices induced by the Frobenius norm $\|\cdot\|_{\mathrm{F}}$). By the row-wise structure of $\mathcal{F}$ in (12), we have

$$\overline{\mathcal{F}} = \{\boldsymbol{F} : \boldsymbol{F}_i \in \overline{\mathcal{F}}_i,\ i = 1, \ldots, L\}, \tag{72}$$

where $\overline{\mathcal{F}}_i$ denotes the closure of $\mathcal{F}_i$ in the Euclidean space $\mathbb{R}^L$. Since $\mathcal{F}$ is bounded, $\overline{\mathcal{F}}$ is *compact*. Since $\rho(\boldsymbol{F})$, $\boldsymbol{F}\boldsymbol{p}$ and $(\boldsymbol{I} - \boldsymbol{F})^{-1}\boldsymbol{v}$ are all continuous functions in $\boldsymbol{F}$, (68), (70) and (71) are equivalent to

$$\max_{\boldsymbol{F}\in\overline{\mathcal{F}}} \rho(\boldsymbol{F}) < 1 \tag{73}$$

$$\boldsymbol{p} = \max_{\boldsymbol{F}\in\overline{\mathcal{F}}} \boldsymbol{F}\boldsymbol{p} + \boldsymbol{v} \tag{74}$$

$$\boldsymbol{p}^* = \max_{\boldsymbol{F}\in\overline{\mathcal{F}}} (\boldsymbol{I} - \boldsymbol{F})^{-1}\boldsymbol{v}. \tag{75}$$

Let $\boldsymbol{F}^{(i)} \in \overline{\mathcal{F}}$ denote a maximizer of the i^{th} component in (75), and $\tilde{\boldsymbol{p}}_i = (\boldsymbol{I} - \boldsymbol{F}^{(i)})^{-1}\boldsymbol{v}$. By definition, $\boldsymbol{p}^*$ is the componentwise maximum of $\{\tilde{\boldsymbol{p}}_1, \ldots, \tilde{\boldsymbol{p}}_L\}$. Since

$$\tilde{\boldsymbol{p}}_i = \boldsymbol{F}^{(i)}\tilde{\boldsymbol{p}}_i + \boldsymbol{v} \preceq \max_{\boldsymbol{F}\in\overline{\mathcal{F}}} \boldsymbol{F}\tilde{\boldsymbol{p}}_i + \boldsymbol{v} \preceq \max_{\boldsymbol{F}\in\overline{\mathcal{F}}} \boldsymbol{F}\boldsymbol{p}^* + \boldsymbol{v},$$

we have

$$\boldsymbol{p}^* \preceq \max_{\boldsymbol{F}\in\overline{\mathcal{F}}} \boldsymbol{F}\boldsymbol{p}^* + \boldsymbol{v}. \tag{76}$$

On the other hand, let $\boldsymbol{F}^*$ achieves the componentwise maximum in (74), that is, its i^{th} row satisfies $\boldsymbol{F}^*_i \in \arg\max_{\boldsymbol{F}_i\in\overline{\mathcal{F}}_i} \boldsymbol{F}_i^{\mathrm{T}}\boldsymbol{p}^*$. By the row-wise structure of $\overline{\mathcal{F}}$, $\boldsymbol{F}^* \in \overline{\mathcal{F}}$. Then (73) and (76) imply that $\boldsymbol{p}^* \preceq \boldsymbol{F}^*\boldsymbol{p}^* + \boldsymbol{v}$ and $\rho(\boldsymbol{F}^*) < 1$. Hence $(\boldsymbol{I} - \boldsymbol{F}^*)^{-1} = \sum_{k\ge 0}(\boldsymbol{F}^*)^k$ has nonnegative entries. Therefore $\boldsymbol{p}^* \preceq (\boldsymbol{I} - \boldsymbol{F}^*)^{-1}\boldsymbol{v}$. In view of (75), we conclude that

$$\boldsymbol{p}^* = (\boldsymbol{I} - \boldsymbol{F}^*)^{-1}\boldsymbol{v}, \tag{77}$$

In view of the definition of $\boldsymbol{F}^*$ and (74), we have shown that $\boldsymbol{p}^* \in \mathbb{R}^L_+$ is a solution to (70). By the uniqueness shown in Theorem 1 shown before, the proof is complete. We also see that the componentwise maximum of $(\boldsymbol{I}-\boldsymbol{F})^{-1}\boldsymbol{v}$ in (75) is simultaneously maximized by the maximizer $\boldsymbol{F}^*$ of $\boldsymbol{F}\boldsymbol{p}^*$ in (74).

Now we prove the convergence rate is given by (20). Suppose $\boldsymbol{g}$ is Fréchet-differentiable at $\boldsymbol{p}^*$. We show that the Jacobian matrix of the interference function $\boldsymbol{I}$ defined in (21) is given by $\boldsymbol{F}^*$. To see this, note that $\boldsymbol{I}_i(\boldsymbol{p}) = \boldsymbol{v} + \max_{\boldsymbol{F}_i \in \overline{\mathcal{F}_i}} \boldsymbol{F}_i^{\mathrm{T}}\boldsymbol{p}$, which is a convex function. Since $\boldsymbol{I}_i(\boldsymbol{p}^*) = \boldsymbol{v} + \boldsymbol{F}_i^*\boldsymbol{p}^*$, $\boldsymbol{F}_i^*$ is a subderivative of $\boldsymbol{I}_i$ at $\boldsymbol{p}^*$. Since $\boldsymbol{I}_i$ is differentiable at $\boldsymbol{p}^*$, its subdifferential at $\boldsymbol{p}^*$ is a singleton set, i.e., $\partial \boldsymbol{I}_i(\boldsymbol{p}^*) = \{\nabla \boldsymbol{I}_i(\boldsymbol{p}^*)\}$. Hence we have $\nabla \boldsymbol{I}_i(\boldsymbol{p}^*) = \boldsymbol{F}_i^*$, i.e., $D_{\boldsymbol{p}}\boldsymbol{I}(\boldsymbol{p}^*) = \boldsymbol{F}^*$. By [11, Theorem 10.1.4], the root-convergence factor of the iteration $\boldsymbol{p}(k+1) = \boldsymbol{I}(\boldsymbol{p}(k))$ to the fixed point $\boldsymbol{p}^*$ is given by

$$\lim_{k\to\infty} \|\boldsymbol{p}(k) - \boldsymbol{p}^*\|_2^{\frac{1}{k}} = \rho(D_{\boldsymbol{p}}\boldsymbol{I}(\boldsymbol{p}^*)) = \rho(\boldsymbol{F}^*) = \sup_{\boldsymbol{F}\in\mathcal{F}} \rho(\boldsymbol{F}). \tag{78}$$

B Proof of Theorem 2 and Theorem 3

Proof To streamline the proof, denote $\tilde{\boldsymbol{p}}(s) = \boldsymbol{p}(s,0) = \boldsymbol{p}(sM)$. We first show that $\tilde{\boldsymbol{p}}(s)$ converges to $\boldsymbol{p}^*$. According to the (23), $\tilde{\boldsymbol{p}}$ satisfies the following iterative relation:

$$\tilde{\boldsymbol{p}}(s+1) = \boldsymbol{G}_M[\boldsymbol{v} + \boldsymbol{g}(\tilde{\boldsymbol{p}}(s))] + \boldsymbol{H}_M\tilde{\boldsymbol{p}}(s) \triangleq \boldsymbol{I}_M(\tilde{\boldsymbol{p}}(s)), \tag{79}$$

where $\boldsymbol{G}_M \triangleq \sum_{l=0}^{M-1} \bar{\boldsymbol{F}}^l = (\boldsymbol{I}-\bar{\boldsymbol{F}})^{-1}(\boldsymbol{I}-\bar{\boldsymbol{F}}^M)$ and $\boldsymbol{H}_M \triangleq \bar{\boldsymbol{F}}^M$. $\boldsymbol{I}_M$ can be shown to be a standard interference function in entirely analogous fashion as in Section 3. Then by by [3, Theorem 2], (79) converges to its fixed point, which is unique in view of [3, Theorem 1]. By (75), it is straightforward to verify that $\boldsymbol{p}^*$ is the fixed point of (79). Similarly to the proof of (20) in Appendix A, the root-convergence factor of (79) is given by $\rho(D_{\boldsymbol{p}}\boldsymbol{I}_M(\boldsymbol{p}^*) = \rho(\boldsymbol{G}_M\boldsymbol{F}^* + \boldsymbol{H}_M)$.

Next we show that the convergence of $\tilde{\boldsymbol{p}}(s)$ governs the convergence of $\boldsymbol{p}(k)$ as well as its exponent. Indeed,

$$\begin{aligned} \boldsymbol{p}(s,l) &= \boldsymbol{G}_l[\boldsymbol{v} + \boldsymbol{g}(\tilde{\boldsymbol{p}}(s))] + \boldsymbol{H}_l\tilde{\boldsymbol{p}}(s) && (80)\\ &= \boldsymbol{G}_l\boldsymbol{G}_M^{-1}\tilde{\boldsymbol{p}}(s+1) + (\boldsymbol{H}_l - \boldsymbol{G}_l\boldsymbol{G}_M^{-1}\boldsymbol{H}_M)\tilde{\boldsymbol{p}}(s), && (81)\end{aligned}$$

hence

$$\lim_{k\to\infty} \|\boldsymbol{p}(k) - \boldsymbol{p}^*\|_2^{\frac{1}{k}} = \lim_{s\to\infty} \|\tilde{\boldsymbol{p}}(s) - \boldsymbol{p}^*\|_2^{\frac{1}{sM}} = [\rho(\boldsymbol{G}_M\boldsymbol{F}^* + \boldsymbol{H}_M)]^{\frac{1}{M}}. \tag{82}$$

Next we proceed to prove Theorem 3. First we show that $\boldsymbol{g}$ is a $\sigma_{\mathcal{F}}$-Lipschitz function, where $\boldsymbol{g}$ is the protection function defined in (14). Indeed,

$$\|\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{y})\| \le \max_{\boldsymbol{F}\in\mathcal{F}} \left\|(\boldsymbol{F}-\bar{\boldsymbol{F}})(\boldsymbol{x}-\boldsymbol{y})\right\| \le \sigma_{\mathcal{F}}\|\boldsymbol{x}-\boldsymbol{y}\|.$$

Therefore $\|D_{\boldsymbol{p}}\boldsymbol{g}(\boldsymbol{p}^*)\| \le \sigma_F$. Hence

$$\begin{aligned} E(M,\mathcal{F}) &\ge \frac{1}{M}\log\frac{1}{\|\boldsymbol{G}_M D_{\boldsymbol{p}}\boldsymbol{g}(\boldsymbol{p}^*) + \boldsymbol{H}_M\|}\\ &\ge \frac{1}{M}\log\frac{1}{\frac{\sigma_{\mathcal{F}}}{1-\|\bar{\boldsymbol{F}}\|}(1-\|\bar{\boldsymbol{F}}\|^M) + \|\bar{\boldsymbol{F}}\|^M}\end{aligned}$$

where we have used the facts: $\rho(\boldsymbol{F}) \le \|\boldsymbol{F}\|$, $\|\boldsymbol{F}+\boldsymbol{G}\| \le \|\boldsymbol{F}\| + \|\boldsymbol{G}\|$ and $\left\|\boldsymbol{F}^k\right\| \le \|\boldsymbol{F}\|^k$.

To conclude the proof, we show (29). It follows from (79) that $\boldsymbol{I}_M$ is a Lipschitz function with Lipschtiz constant $\|\boldsymbol{G}_M\|\sigma_{\mathcal{F}} + \|\boldsymbol{H}_M\| \le \exp[-M\underline{E}(M,\sigma_{\mathcal{F}})]$. Therefore

$$\|\boldsymbol{p}(s,0) - \boldsymbol{p}^*\| \le \exp[-sM\underline{E}(M,\sigma_{\mathcal{F}})]\,\|\boldsymbol{p}(0) - \boldsymbol{p}^*\|.$$

By (80), for $k = sM + l$ $(0 \le l \le M-1)$, we have

$$\begin{aligned} \|\boldsymbol{p}(k) - \boldsymbol{p}^*\| \le\ & \|\boldsymbol{G}_l\|\,\|\boldsymbol{g}(\boldsymbol{p}(s,0)) - \boldsymbol{g}(\boldsymbol{p}^*)\| + \|\boldsymbol{H}_l\|\,\|\boldsymbol{p}(s,0) - \boldsymbol{p}^*\|\\ \le\ & \|\boldsymbol{p}(0) - \boldsymbol{p}^*\|\left(\frac{1-\|\bar{\boldsymbol{F}}\|^l}{1-\|\bar{\boldsymbol{F}}\|}\sigma_{\mathcal{F}} + \|\bar{\boldsymbol{F}}\|^l\right)\exp[-sM\underline{E}(M,\sigma_{\mathcal{F}})]\\ =\ & \|\boldsymbol{p}(0) - \boldsymbol{p}^*\|\exp[-sM\underline{E}(M,\sigma_{\mathcal{F}})],\end{aligned}$$

which implies the desired result (29).

C Proof of Proposition 2

Proof Let $\alpha = \frac{\sigma_{\mathcal{F}}}{1-\|\bar{\boldsymbol{F}}\|}$, $\beta = \|\bar{\boldsymbol{F}}\|$, then by the assumption of (43), $\alpha, \beta \in (0,1)$. For $x > 0$ define $L(x) = -\frac{1}{x}\log[(1-\alpha)\beta^x + \alpha]$, then $\underline{E}(M, \sigma_{\mathcal{F}}) = L(M)$ according to (28).

1. Positivity: since $\alpha, \beta \in (0,1)$, $(1-\alpha)\beta^x + \alpha \in (0,1)$. Therefore $L(x) > 0$ for any $x > 0$.
2. Strict monotonicity in $\sigma_{\mathcal{F}}$: directly from (28), we see that $\underline{E}(M, \sigma_{\mathcal{F}})$ decreases strictly with $\sigma_{\mathcal{F}}$. As $\sigma_{\mathcal{F}}$ tends to $1 - \left\|\bar{\boldsymbol{F}}\right\|_2$, α tends to 1, hence $L(M)$ vanishes.
3. Strict monotonicity in M: note that

$$L'(x) = \frac{h(x)}{x^2[(1-\alpha)\beta^x + \alpha]}, \tag{83}$$

where $h(x) = -\beta^x(1-\alpha)\log\beta^x + [(1-\alpha)\beta^x + \alpha]\log[(1-\alpha)\beta^x + \alpha]$. Since the denominator in (83) is positive, to show $L'(x) < 0$ for $x > 0$ it is sufficient to show that $h(0) = 0$ and $h'(x) < 0$. Indeed, we have $h'(x) = (1-\alpha)\beta^x \log\beta \log\frac{(\alpha+(1-\alpha)\beta^x)}{\beta^x} < 0$.

D Proof of Theorem 4

Proof By Theorem 1, it is sufficient to check

$$\max_{\boldsymbol{F}\in\mathcal{F}} \rho(\boldsymbol{F}) \leq \left\|\bar{\boldsymbol{F}}\right\|_2 + \|\boldsymbol{\epsilon}\|_2 \tag{84}$$

and

$$\max_{\boldsymbol{F}\in\mathcal{F}} \rho(\boldsymbol{F}) \leq \frac{\rho(\bar{\boldsymbol{F}} + \bar{\boldsymbol{F}}^{\mathrm{T}})}{2} + \|\boldsymbol{\epsilon}\|_2 \,. \tag{85}$$

Recall the Frobenius norm of a matrix is defined as $\|\boldsymbol{F}\|_{\mathrm{F}} = \sqrt{\sum_{i,j}|F_{ij}|^2}$, and $\|\boldsymbol{F}\|_{\mathrm{F}} \geq \|\boldsymbol{F}\|_2$ [18]. For $\boldsymbol{F}\in\mathcal{F}$, let $\boldsymbol{F} = \bar{\boldsymbol{F}} + \Delta\boldsymbol{F}$. Then $\|\Delta\boldsymbol{F}_i\|_2 \leq \epsilon_i$, hence $\|\Delta\boldsymbol{F}\|_{\mathrm{F}} \leq \|\boldsymbol{\epsilon}\|_2$. Therefore

$$\begin{aligned}\rho(\bar{\boldsymbol{F}} + \Delta\boldsymbol{F}) &\leq \left\|\bar{\boldsymbol{F}} + \Delta\boldsymbol{F}\right\|_2 \leq \left\|\bar{\boldsymbol{F}}\right\|_2 + \|\Delta\boldsymbol{F}\|_2 \\ &\leq \left\|\bar{\boldsymbol{F}}\right\|_2 + \|\Delta\boldsymbol{F}\|_{\mathrm{F}} \leq \left\|\bar{\boldsymbol{F}}\right\|_2 + \|\boldsymbol{\epsilon}\|_2 \,,\end{aligned}$$

this proves (84). By [19, Theorem 1.1],

$$\max_{\|\boldsymbol{G}\|_{\mathrm{F}}\leq 1} \rho(\bar{\boldsymbol{F}} + \boldsymbol{G}) \leq \frac{\rho(\bar{\boldsymbol{F}} + \bar{\boldsymbol{F}}^{\mathrm{T}})}{2} + 1,$$

hence

$$\max_{\boldsymbol{F}\in\mathcal{F}} \rho(\bar{\boldsymbol{F}} + \Delta\boldsymbol{F}) \leq \max_{\|\Delta\boldsymbol{F}\|_{\mathrm{F}}\leq\|\boldsymbol{\epsilon}\|_2} \rho(\bar{\boldsymbol{F}} + \Delta\boldsymbol{F}) \leq \frac{\rho(\bar{\boldsymbol{F}} + \bar{\boldsymbol{F}}^{\mathrm{T}})}{2} + \|\boldsymbol{\epsilon}\|_2 \,,$$

and (85) follows.

When $\bar{\boldsymbol{F}} = \bar{\boldsymbol{F}}^{\mathrm{T}}$, note that $\frac{\rho(\bar{\boldsymbol{F}}+\bar{\boldsymbol{F}}^{\mathrm{T}})}{2} = \rho(\bar{\boldsymbol{F}}) \leq \left\|\bar{\boldsymbol{F}}\right\|_2$, hence (36) reduces to (37).

E Proof of Theorem 6

Proof Note that

$$\rho(\bar{\boldsymbol{F}} + \Delta\boldsymbol{F}) \leq \left\|\bar{\boldsymbol{F}} + \Delta\boldsymbol{F}\right\|_1 \leq \left\|\bar{\boldsymbol{F}}\right\|_1 + \|\Delta\boldsymbol{F}\|_1 \leq \left\|\bar{\boldsymbol{F}}\right\|_1 + \max_j \sum_i t_{ij},$$

where the last inequality follows from [18]

$$\|\Delta\boldsymbol{F}\|_1 = \max_j \sum_i |\Delta F_{ij}| \leq \max_j \sum_i t_{ij}.$$

Therefore $\sup_{\boldsymbol{F}\in\mathcal{F}} \rho(\boldsymbol{F}) \leq \left\|\bar{\boldsymbol{F}}\right\|_1 + \max_j \sum_i t_{ij} < 1$, and the theorem holds by Theorem 1.

F Proof of Theorem 7

Proof Define $\boldsymbol{\Delta x} = [\Delta F_{i1}x_1, \ldots, \Delta F_{iL}x_L]^{\mathrm{T}}$ and $\boldsymbol{\Delta y}$ similarly. By [16, Proposition 3], $\langle\cdot\rangle_k \leq \sqrt{k}\,\|\cdot\|_2$. Then by triangle inequality of $\langle\cdot\rangle_{\Gamma_i}$,

$$|g_i^D(\Gamma_i, \boldsymbol{x}) - g_i^D(\Gamma_i, \boldsymbol{y})| \leq \langle \boldsymbol{\Delta x} - \boldsymbol{\Delta y}\rangle_{\Gamma_i} \leq \sqrt{\Gamma_i}\,\|\boldsymbol{\Delta x} - \boldsymbol{\Delta y}\|_2 .$$

Hence $\left\|\boldsymbol{g}^D(\boldsymbol{\Gamma}, \boldsymbol{x}) - \boldsymbol{g}^D(\boldsymbol{\Gamma}, \boldsymbol{y})\right\|_2^2 \leq \sum_i \Gamma_i \sum_j \Delta F_{ij}^2 (x_i - y_i)^2 \leq \|\boldsymbol{\Gamma}\|_\infty \|\boldsymbol{F}\|_{\mathrm{F}}^2 \|\boldsymbol{x} - \boldsymbol{y}\|_2^2$. Hence $\boldsymbol{I}(\boldsymbol{p}) = \bar{\boldsymbol{F}}\boldsymbol{p} + \boldsymbol{g}^D(\boldsymbol{\Gamma}, \boldsymbol{p}) + \boldsymbol{v}$ is a contractive mapping and the iteration converges. Optimality of the fixed point can be shown similarly.

Acknowledgements The authors would like to thank Bo Jiang for useful inputs on an earlier draft of this paper.

References

1. K. Yang, J. Huang, Y. Wu, X. Wang, and M. Chiang, "Distributed Robust Optimization (DRO) Part I: Framework and Example," 2008, Draft.
2. M. Chiang, P. Hande, T. Lan, and C. W. Tan, "Power control in wireless cellular networks," *Foundation and Trends in Networking*, vol. 2, no. 4, pp. 381 – 533, July 2008.
3. R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 1341 – 1347, 1995.
4. G. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *Vehicular Technology, IEEE Transactions on*, vol. 42, no. 4, pp. 641–646, Nov. 1993.
5. 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.
6. C. W. Tan, D. P. Palomar, and M. Chiang, "Exploiting hidden convexity for flexible and robust resource allocation in cellular networks," in *Proc. IEEE INFOCOM*, 2007.
7. J. Huang, R. Berry, and M. L. Honig, "Distributed interference compensation in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 5, pp. 1074–1084, May 2006.
8. J. Lee, M. Chiang, and A. Calderbank, "Utility-optimal random-access control," *IEEE transactions on wireless communications*, vol. 6, no. 7, pp. 2741 – 2751, 2007.
9. F. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, no. 3, pp. 237–252, 1998.
10. R. Schneider, *Convex bodies: the Brunn-Minkowski theory.* Cambridge University Press, 1993.
11. J. Ortega and W. Rheinboldt, *Iterative solution of nonlinear equations in several variables.* Academic Press, 1970.
12. D. P. Bertsekas, *Nonlinear programming 2nd ed.* Belmont, Mass. : Athena Scientific, 1999.
13. B. W. Levinger, "An inequality for nonnegative matrices," *Notices of the American Mathematical Society*, vol. 17, p. 260, 1970.
14. "Mosek." [Online]. Available: http://www.mosek.com
15. R. Jantti and S. Kim, "Second-order power control with asymptotically fast convergence," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 447 – 457, 2000.
16. D. Bertsimasa, D. Pachamanovab, and M. Simc, "Robust linear optimization under general norms," *Operations Research Letters*, vol. 32, pp. 510 – 516, 2004.
17. R. S. Varga, *Matrix iterative analysis.* Englewood Cliffs, N.J: Prentice-Hall, 1962.
18. R. A. Horn and C. R. Johnson, *Matrix Analysis.* Cambridge University Press, 1990.
19. L. Han, M. Neumann, and M. Tsatsomeros, "Spectral radii of fixed Frobenius norm perturbations of nonnegative matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 1, pp. 79–92, 2000.