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

Yihong Wu[1], Kai Yang[2], Jianwei Huang[3], Xiaodong Wang[2], Mung Chiang[1]

[1]Department of Electrical Engineering, Princeton University, Princeton, NJ, USA

[2]Department of Electrical Engineering, Columbia University, New York, NY, USA

[3]Deparmtent of Information Engineering, The Chinese University of Hong Kong, HKSAR, China

{kyee, wangx}@ee.columbia.edu, {yihongwu, chiangm}@princeton.edu, jwhuang@ie.cuhk.edu.hk

Abstract

Optimization formulations and distributed solution algorithms have long been used for resource allocation problems in wireless networks including power control. However, constant parameters in these formulations are often time-varying, unknown, or based on inaccurate estimates. Can we keep the algorithms distributed with little communication overhead while making the solutions robust to inevitable errors in problem formulations? We answer this question through the general framework of Distributed Robust Optimization, as in Part I of this two-part paper. Robustness against channel fluctuations, SIR measurement errors, and user dynamics is described through the ellipsoid, polyhedron, and D-norm uncertain sets, respectively. We then 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. Related tradeoffs such as robustness vs. nominal performance, and robustness vs. energy expenditure are also characterized, while convergence of new, robust power control algorithms proved under various sufficient conditions.

J. Huang is supported by CUHK Direct Research Grant 2050398.

Index Terms

Wireless power control, robust optimization, distributed algorithm, distributiveness-robustness-performance tradeoff.

I. 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) is a general framework to study 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*, including worst-case and probabilistic setup formulations, etc., and a *uncertainty set*, which is application-specific and approximates practical uncertainties. We are interesting 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 established the framework 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 model is applicable to building robustness with respect to three kinds of dynamics: channel variation, SIR measurement errors, and user entry. Utilizing nonnegative matrix theory and the standard interference function framework [3], we show that the *necessary and sufficient* condition of feasibility as well as convergence and optimality of the algorithm is that the maximal spectral radius of normalized channel matrices is bounded away from one. This generalizes well-known results in [4] to a robust scenario. Based upon this criterion algorithms and sufficient conditions for specific uncertainty sets are derived. Furthermore, we introduce the notion of infrequent feedback where the globally coupled parameter is updated less frequently than user-end iterations. This technique balances the tradeoff between robustness, distributiveness and performance in terms of convergence rate. A quantitative analysis of this tradeoff provides insights for network designers to choose the best operating point of robust power control protocols.

The rest of the paper is organized as follows. Section II provides a brief review on the nominal wireless power control problem and the DRO framework. Section III 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. Sections IV – VI are devoted to modeling channel fluctuations, SIR measurement inaccuracy and user dynamics using ellipsoids, polyhedrons and D-norm uncertainty sets respectively. Section VII concludes the paper. Proofs are relegated to the appendix.

II. Background

A. The Nominal Problem and DPC 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 a transmitter node and a receiver node. If all users are distinct, this could model a wireless ad hoc network. If all users share the same transmitter node (or receiver node), then 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 gains from the j^{th} user's transmitter to the i^{th} user'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 SNR $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_L]$, such that the total transmission power is minimized:

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

Introducing the notation $\boldsymbol{v} = \left[\frac{\gamma_1 n_1}{G_{11}}, \ldots, \frac{\gamma_L n_L}{G_{LL}}\right]$ and the normalized channel matrix $\boldsymbol{F} = [F_{ij}]$ with

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

we have

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

and Problem (2) can be written equivalently as:

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

where $\mathbf{1}$ and $\boldsymbol{I}$ denotes the all-one $L \times 1$ vector and $L \times L$ identity matrix respectively.

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

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

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

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

the system will globally converge to the optimal solution in (6). We 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 system.

Robust power control has been studied in limited contexts before our development of the DRO framework. 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, against a range of uncertainties: channel fluctuation, SIR measurement errors, and users entering and leaving the systems. In [6] the tradeoff between the robustness and the extra power consumption is studied with penalty-defined formulation, while the key focus of this paper is to study the tradeoff between robustness and distributiveness as in Section IV-D, though we also study the tradeoff of extra power allocation versus robustness in Section IV-E. 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.

B. DRO

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 $\boldsymbol{A}$ be $\boldsymbol{a}_j^T$, which lies in a compact uncertainty set $\mathcal{A}_j$. Then the *robust* optimization problem that we focus on 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],

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* $\bar{\boldsymbol{a}}_j$. $\boldsymbol{g}_j(\boldsymbol{x})$ *is a convex function for each* j.

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

III. 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 IV to VI 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). Further denote the i^{th} row of $\bar{\boldsymbol{F}}$ as $\bar{\boldsymbol{F}}_i$ and the corresponding uncertainty 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 in later subsections. The uncertainty set of $\boldsymbol{F}$ is $\mathcal{F} = \{\boldsymbol{F} : \boldsymbol{F}_i \in \mathcal{F}_i, \ i = 1, \ldots, L\}$, 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

$$\begin{array}{rl} \text{minimize} & \mathbf{1}^{\mathrm{T}}\boldsymbol{p} \\ \text{subject to} & \boldsymbol{p} \succeq \bar{\boldsymbol{F}}\boldsymbol{p} + \boldsymbol{v} + \boldsymbol{g}(\boldsymbol{p}), \\ \text{variables} & \boldsymbol{p} \succeq 0. \end{array} \tag{12}$$

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

$$g_i(\boldsymbol{p}) = \sup_{\boldsymbol{F}_i \in \mathcal{F}_i} (\boldsymbol{F}_i - \bar{\boldsymbol{F}}_i)^{\mathrm{T}}\boldsymbol{p}. \tag{13}$$

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, 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, 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 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 suits the worst-case modeling methodology well, since 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 (12) is feasible, that is, there exists a $\hat{\boldsymbol{p}} \succeq 0$, such that for any $\boldsymbol{F} \in \mathcal{F}$,

$$\hat{\boldsymbol{p}} \succeq \boldsymbol{F}\hat{\boldsymbol{p}} + \boldsymbol{v}. \tag{14}$$

2)

$$\sup_{\boldsymbol{F} \in \mathcal{F}} \rho(\boldsymbol{F}) < 1. \tag{15}$$

3) Starting from any initial point, the following iteration

$$\boldsymbol{p}(k+1) = \bar{\boldsymbol{F}}\boldsymbol{p}(k) + \boldsymbol{v} + \boldsymbol{g}(\boldsymbol{p}(k)) \tag{16}$$

converges asynchronously to the optimal solution of (12), given by the unique solution of

$$\boldsymbol{p}^* = \bar{\boldsymbol{F}}\boldsymbol{p}^* + \boldsymbol{g}(\boldsymbol{p}^*) + \boldsymbol{v}. \tag{17}$$

Moreover, the optimal power allocation admits the following characterization:

$$\boldsymbol{p}^* = \sup_{\boldsymbol{F} \in \mathcal{F}} (\boldsymbol{I} - \boldsymbol{F})^{-1} \boldsymbol{v}. \tag{18}$$

where the supremum is taken componentwise.

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

3) $\Rightarrow$ 1) is obvious.

1) $\Rightarrow$ 3): We make use of Yates' standard interference function framework here. Let the interference function be $\boldsymbol{I}(\boldsymbol{p}) = \bar{\boldsymbol{F}}\boldsymbol{p} + \boldsymbol{v} + \boldsymbol{g}(\boldsymbol{p})$, with $\boldsymbol{g}(\boldsymbol{p})$ as in (13). Then (14) implies that $\boldsymbol{p}_0 \succeq \boldsymbol{I}(\boldsymbol{p}_0)$, equivalent to the feasibility of Problem (12).

Now we check that $\boldsymbol{I}(\boldsymbol{p})$ is a *standard* interference function as defined in [3]:

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 nonnegative entries, for any $\boldsymbol{p}' \geq \boldsymbol{p}$, we have $\boldsymbol{F}\boldsymbol{p}' \geq \boldsymbol{F}\boldsymbol{p}$ and $\boldsymbol{I}(\boldsymbol{p}') \succeq \boldsymbol{I}(\boldsymbol{p})$.
3) Scalability: $\forall \alpha > 1$, $\forall j$,

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

Hence (16) converges to the optimal solution of (12) 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 (18). ■

Remark 1. *The equivalence of feasibility and convergence is due to the fact that our formulation falls naturally into the interference function framework introduced by Yates in [3]. The relationship between the supremum of spectral radius and the feasibility hinges on the nonnegativity of each $\boldsymbol{F}$ and the product set structure of $\mathcal{F}$, which follows from $\mathcal{F}$ being a* row-wise *uncertainty set.*

Theorem 1 implies that if the robust power control problem is feasible, the fixed-point iteration algorithm (16) will drive each user's power to the globally optimal allocation. Under certain regular uncertainty sets such as ellipsoid, polyhedron or D-norm, (16) could be implemented in a distributive fashion, as manifested later in (23), (47) and (52) respectively. Moreover, the technique of infrequent feedback could be used to further reduce global message passing and balance the *tradeoff* between distributiveness and robustness. In later subsections 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 between robustness, distributiveness, performance and total power consumption. Due to space limitations, we take ellipsoid uncertainty sets as the primary example.

IV. 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 $\boldsymbol{\epsilon} = [\epsilon_1, \dots, \epsilon_L]^{\mathrm{T}} \succeq 0$. The uncertainty set $\mathcal{F}_i$ for $\boldsymbol{F}_i$ under ellipsoid approximation can be represented by

$$\mathcal{F}_i = \left\{ \bar{\boldsymbol{F}}_i + \Delta \boldsymbol{F}_i : \sum_{j \neq i} |\Delta F_{ij}|^2 \leq \epsilon_i^2. \right\} \tag{19}$$

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 (19) is found as

$$g_i^{\mathrm{ell}}(\boldsymbol{p}) = \epsilon_i \sqrt{\sum_{j \neq i} p_j^2} = \epsilon_i \sqrt{\|\boldsymbol{p}\|_2^2 - p_i^2}, \tag{20}$$

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

$$\begin{array}{ll} \text{minimize} & \boldsymbol{1}^{\mathrm{T}} \boldsymbol{p} \\ \text{subject to} & (\boldsymbol{I} - \boldsymbol{F})\boldsymbol{p} \succeq \boldsymbol{v} + \boldsymbol{g}^{\mathrm{ell}}(\boldsymbol{p}), \\ \text{variables} & \boldsymbol{p} \succeq 0. \end{array} \tag{21}$$

This is the central problem we will solve in this subsection. We recognize it as an SOCP (Second-Order Cone Programming) problem [10], which could be rewritten in the standard form as

$$\begin{array}{ll} \text{minimize} & \boldsymbol{1}^{\mathrm{T}} \boldsymbol{p} \\ \text{subject to} & \sqrt{\boldsymbol{p}^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{e_i}\boldsymbol{e_i}^{\mathrm{T}})\boldsymbol{p}} \leq (\boldsymbol{e_i} - \boldsymbol{F_i})^{\mathrm{T}} \boldsymbol{p} - v_i, \\ & i = 1, \dots, L. \end{array} \tag{22}$$

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

A. Optimality Conditions and Distributed Algorithms

Plugging (20) into (16), we obtain the following distributive algorithm:

Algorithm 1. *For $i = 1, \ldots, L$, at each time slot k, the i^{th} user updates its transmission power according to*

$$p_i(k+1) = \frac{\gamma_i}{\mathsf{SIR}_i} p_i(k) + \epsilon_i \sqrt{Q^2(k) - p_i^2(k)} \tag{23}$$

where

$$Q(k) = \|\boldsymbol{p}(k)\|_2. \tag{24}$$

Note that (23) is a generalization of the classical DPC algorithm in (7) with a protection function $g_i^{\text{ell}}(\boldsymbol{p}(k)) = \epsilon_i \sqrt{Q^2(k) - p_i^2(k)}$ 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 reduces 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 2. *Problem (21) is feasible, or equivalently, Algorithm 1 converges to the optimal solution of Problem (21), if*

$$\min\left\{\left\|\bar{\boldsymbol{F}}\right\|_2, \frac{\rho(\bar{\boldsymbol{F}} + \bar{\boldsymbol{F}}^{\mathrm{T}})}{2}\right\} + \|\boldsymbol{\epsilon}\|_2 < 1, \tag{25}$$

where $\left\|\bar{\boldsymbol{F}}\right\|_2$ is the induced ℓ_2-norm of $\bar{\boldsymbol{F}}$, i.e., $\left\|\bar{\boldsymbol{F}}\right\|_2 = \sup_{\|\boldsymbol{p}\|_2 = 1} \left\|\bar{\boldsymbol{F}}\boldsymbol{p}\right\|_2$. If $\bar{\boldsymbol{F}}$ is symmetric, (25) reduces to

$$\|\boldsymbol{\epsilon}\|_2 + \rho(\bar{\boldsymbol{F}}) < 1. \tag{26}$$

Proof: See Appendix B. ■

Remark 2. *From Levinger's inequality [11],*

$$\frac{\rho(\bar{\boldsymbol{F}} + \bar{\boldsymbol{F}}^{\mathrm{T}})}{2} \geq \rho(\bar{\boldsymbol{F}}), \tag{27}$$

we see that $\rho(\bar{\boldsymbol{F}})$ is a lower bound to the sufficient condition in Theorem 2. 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 at the same time, the system becomes less resilient to noise or uncertainty in the channel matrix.

B. Infrequent Feedback

If Q in Algorithm 1 is allowed to be updated at a slower pace without affecting the convergence or optimality, the distributiveness of the algorithm can be further increased. In particular, we can choose a parameter $M \geq 1$ such that Q is updated every M time slots. As M increases, the amount of message passed among users decreases. To facilitate the discussions, we can represent any time index k as (s, l), such that $k = sM + l$, $s = 0, 1, \ldots$, and $0 \leq l \leq M - 1$. Then we can design the following distributive algorithms:

Algorithm 2. *For $i = 1, \ldots, L$, at each time slot $k = (s, l)$, the i^{th} user updates its transmission power according to*

$$p_i(k+1) = \frac{\gamma_i}{\mathsf{SIR}_i} p_i(k) + \epsilon_i \sqrt{Q^2(s,0) - p_i^2(k)}. \tag{28}$$

In Algorithm 2, each user i broadcasts its power $p_i(k)$ and computes $Q(k)$ using (24) every M iterations. Between these updates, users update their power based on the most recently computed value of Q. For the downlink transmission in a single cell network, Q could be simply broadcast by the base station.

Algorithm 2 globally converges to the optimal solution of Problem (21) under proper stability conditions of matrix $\bar{\boldsymbol{F}}$ and uncertainty parameter $\boldsymbol{\epsilon}$. Furthermore, the transient behavior of Algorithm 2 is controllable since the difference between the current power vector and the optimal power vector during each step of the iteration can be bounded. The algorithm converges exponentially and the speed of convergence *decreases* as M increases.

Theorem 3. *Assume*

$$\|\boldsymbol{\epsilon}\|_2 + \|\bar{\boldsymbol{F}}\|_2 < 1. \tag{29}$$

Algorithm 2 globally converges to the optimal solution of Problem (21), denoted as $\boldsymbol{p}^*$. *Moreover,*

$$\|\boldsymbol{p}(k) - \boldsymbol{p}^*\|_2 \leq \frac{(C_M)^{\lfloor k/M \rfloor}}{1 - \|\bar{\boldsymbol{F}}\|_2 - \|\boldsymbol{\epsilon}\|_2} \|\boldsymbol{p}(1) - \boldsymbol{p}(0)\|_2, \tag{30}$$

where

$$C_M = \|\bar{\boldsymbol{F}}\|_2^M + \frac{\|\boldsymbol{\epsilon}\|_2}{1 - \|\bar{\boldsymbol{F}}\|_2}(1 - \|\bar{\boldsymbol{F}}\|_2^M), \tag{31}$$

$M \geq 1$ *is the number of time slots between two adjacent updates of parameter* Q, *and* $\lfloor \cdot \rfloor$ *is the floor function.*

Proof: See Appendix C. ■

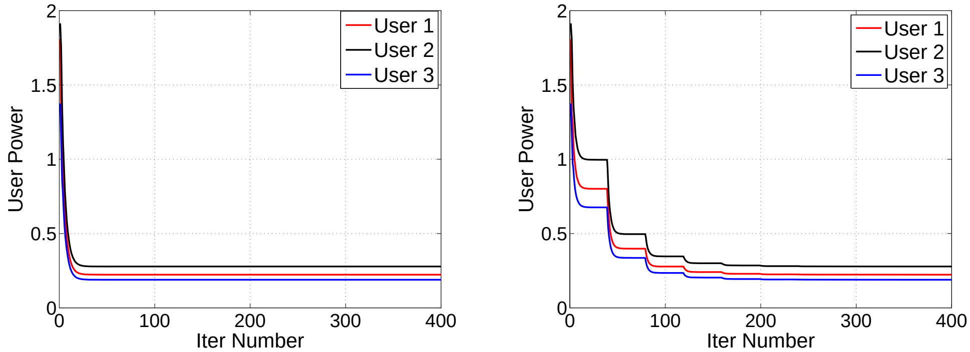
C. Numerical Results and Performance Comparison

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 feedback delay ($M = 1$) and with feedback delay ($M = 40$). In both cases, the algorithm converges to the optimal solution (verified by the centralized MOSEK toolbox [12]) 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 (5 in this case). As shown in Fig. 2, Algorithm 2 avoids channel outage since it considers the worst case of the uncertainty set, and the original DPC algorithm leads to frequent channel outages.

D. Robustness-Distributiveness Tradeoff

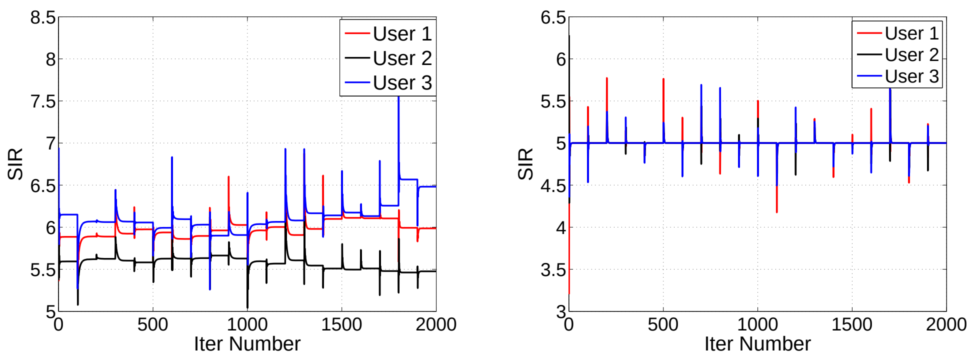
If we fix the total number of iterations as N and the desired optimality gap $\|\boldsymbol{p}(N) - \boldsymbol{p}^*\| = \delta$, then there exists an interesting tradeoff between robustness and distributiveness. In particular,



(a) Without feedback delay ($M = 1$)

(b) With feedback delay ($M = 40$).

Fig. 1. Convergence of Algorithm 2.



(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).

if more robustness is desired (i.e., a lager ϵ), we will have more message passing and less distributiveness (i.e., a smaller number of update interval M).

To characterize this tradeoff analytically, let us first see how the speed of convergence depends on with M and $\boldsymbol{\epsilon}$. By (30), the convergence rate of Algorithm 2 is at least exponentially. Let

$$E(M, \|\boldsymbol{\epsilon}\|_2) = \frac{1}{M} \log \frac{1}{C_M}, \tag{32}$$

where C_M is in (31). Then

$$\|\boldsymbol{p}(k) - \boldsymbol{p}^*\|_2 \leq \frac{\|\boldsymbol{p}(1) - \boldsymbol{p}(0)\|_2}{1 - \|\bar{\boldsymbol{F}}\|_2 - \|\boldsymbol{\epsilon}\|_2} \exp[-NE(M, \|\boldsymbol{\epsilon}\|_2)]. \tag{33}$$

We see that given $\boldsymbol{\epsilon}$, $E(M, \|\boldsymbol{\epsilon}\|_2)$ is a lower bound of the error exponent of convergence of

Algorithm 2 when Q is updated per M time slots.

Observe the following properties of $E(M, \|\boldsymbol{\epsilon}\|_2)$:

Proposition 2. *Under the condition (29) in Theorem 3,*

1) $E(M, \|\boldsymbol{\epsilon}\|_2) > 0$, *for* $M > 0$ *and* $0 \leq \|\boldsymbol{\epsilon}\|_2 < 1 - \left\|\bar{\boldsymbol{F}}\right\|_2$.
2) $E(M, \|\boldsymbol{\epsilon}\|_2)$ *is strictly decreasing in* M. $E(1, \|\boldsymbol{\epsilon}\|_2) = \log \frac{1}{\|\boldsymbol{\epsilon}\|_2 + \left\|\bar{\boldsymbol{F}}\right\|_2}$ *and*

$$\lim_{M\to\infty} E(M, \|\boldsymbol{\epsilon}\|_2) = 0.$$

3) $E(M, \|\boldsymbol{\epsilon}\|_2)$ *is strictly decreasing in* $\|\boldsymbol{\epsilon}\|_2$. $E(M, 0) = \log \frac{1}{\|\bar{\boldsymbol{F}}\|_2}$ *and*

$$\lim_{\|\boldsymbol{\epsilon}\|_2 \to 1 - \|\bar{\boldsymbol{F}}\|_2} E(M, \|\boldsymbol{\epsilon}\|_2) = 0.$$

Proof: See Appendix D. ■

$E(M, \|\boldsymbol{\epsilon}\|_2)$ characterizes the tradeoff between distributiveness $\frac{N}{M}$ and robustness $\|\boldsymbol{\epsilon}\|_2$, as plotted in Fig. 3. For a given robustness requirement $\boldsymbol{\epsilon}$, Algorithm 2 converges slower as less messages are passed. In the limit when M tends to infinity, the rate of convergence becomes arbitrarily small. For a fixed number of message passing $\frac{N}{M}$, i.e., fixed M, the speed of convergence of Algorithm 2 decreases as $\|\boldsymbol{\epsilon}\|_2$ increases, that is, more robustness are taken into account in the power allocation. Note that as a lower bound of convergence rate $E(M, \|\boldsymbol{\epsilon}\|_2)$ is not tight. For instance, in the nominal problem when $\boldsymbol{\epsilon} = 0$, $E(M, 0) = \log \frac{1}{\left\|\bar{\boldsymbol{F}}\right\|_2}$. But in this case the exponent of convergence should be $\log \frac{1}{\rho(\bar{\boldsymbol{F}})}$.

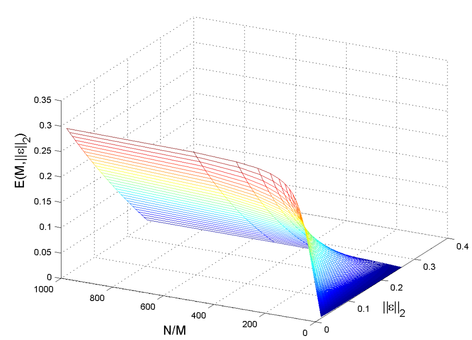


Fig. 3. Plot of exponent of convergence of Algorithm 2 against robustness $\|\boldsymbol{\epsilon}\|_2$ and number of message passing N/M.

By (33), a sufficient condition to achieve the desired optimality gap δ is

$$E(M, \|\boldsymbol{\epsilon}\|_2) = \frac{1}{N} \log \left[\frac{\|\boldsymbol{p}(1) - \boldsymbol{p}(0)\|_2}{\delta(1 - \left\|\bar{\boldsymbol{F}}\right\|_2 - \|\boldsymbol{\epsilon}\|_2)} \right]. \tag{34}$$

Since $E(M, \|\boldsymbol{\epsilon}\|_2)$ is strictly monotonic in M, it has an inverse function, denoted by E^{-1}. We can solve (34) and obtain a lower bound on the largest allowed value of update interval M:

$$M \geq E^{-1} \left[\frac{1}{N} \log \left(\frac{\delta(1 - \left\|\bar{\boldsymbol{F}}\right\|_2 - \|\boldsymbol{\epsilon}\|_2)}{\|\boldsymbol{p}(1) - \boldsymbol{p}(0)\|_2} \right) \right]. \tag{35}$$

Then an 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 is

$$\frac{N}{E^{-1}\left\{ \frac{1}{N} \log \left[\frac{\delta(1 - \left\|\bar{\boldsymbol{F}}\right\|_2 - \|\boldsymbol{\epsilon}\|_2)}{\|\boldsymbol{p}(1) - \boldsymbol{p}(0)\|} \right] \right\}}. \tag{36}$$

Here one message passing corresponds to each user announcing his power level once, or the base station evaluating the current Q in (24) and broadcasting it to the users. This upper bound is also plotted in Fig. 4, together with the simulated result. For notational simplicity assume that $\boldsymbol{\epsilon} = \epsilon \mathbf{1}$. The bound is quite tight when ϵ is small. 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 6 global message passing is needed for $\epsilon = 5\%$, while 25 messages must be passed in order to achieve robustness of $\epsilon = 15\%$. The simulated three-dimensional tradeoff among robustness ϵ, optimality gap δ, and the number of message passing is given in Fig. 5.

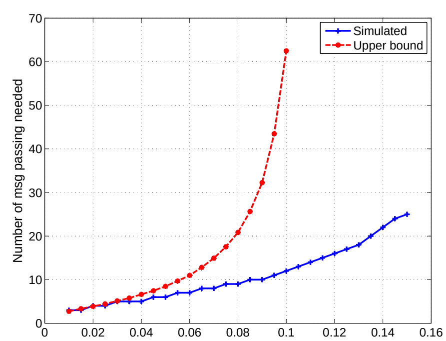


Fig. 4. Robustness-distributiveness tradeoff curve under ellipsoid uncertainty set, where $\boldsymbol{\epsilon} = \epsilon \mathbf{1}$.

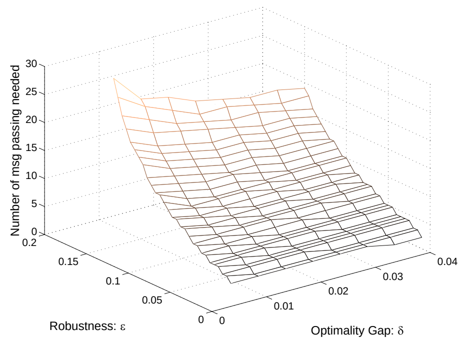


Fig. 5. Three-dimensional tradeoff surface between robustness ϵ, optimality gap δ and the number of message passing.

E. Power-Robustness Tradeoff

The tradeoff between increased power consumption and the robustness are shown in Fig. 6, with comparison with the total power allocation of DPC algorithm. Note here the allocated power are the solution of Problem (21) as well as the steady state of Algorithm 2. The power-robustness tradeoff could be analyzed as follows. Let $\bar{\boldsymbol{p}}$ and $\boldsymbol{p}^*$ denote the solution to the nominal problem (2) and the robust problem (23) respectively. Then

$$\bar{\boldsymbol{p}} = \bar{\boldsymbol{F}}\bar{\boldsymbol{p}} + \boldsymbol{v}, \tag{37}$$

$$\boldsymbol{p}^* = \bar{\boldsymbol{F}}\boldsymbol{p}^* + \boldsymbol{g}^{\text{ell}}(\boldsymbol{p}^*) + \boldsymbol{v}. \tag{38}$$

where $\boldsymbol{g}^{\text{ell}}$ is in (20). Let $\boldsymbol{p}^* = \bar{\boldsymbol{p}} + \Delta\boldsymbol{p}$, and $\mathbf{1}^{\text{T}}\Delta\boldsymbol{p}$ is the extra power consumption. From (37) and (38) we have $\bar{\boldsymbol{p}} + \Delta\boldsymbol{p} = \bar{\boldsymbol{F}}\bar{\boldsymbol{p}} + \bar{\boldsymbol{F}}\Delta\boldsymbol{p} + \boldsymbol{g}^{\text{ell}}(\bar{\boldsymbol{p}} + \Delta\boldsymbol{p}) + \boldsymbol{v}$, therefore $(\boldsymbol{I} - \bar{\boldsymbol{F}})\Delta\boldsymbol{p} = \boldsymbol{g}^{\text{ell}}(\bar{\boldsymbol{p}} + \Delta\boldsymbol{p}) = \boldsymbol{g}^{\text{ell}}(\bar{\boldsymbol{p}}) + D_{\boldsymbol{p}}\boldsymbol{g}^{\text{ell}}(\bar{\boldsymbol{p}}) + \mathcal{O}\left(|\Delta\boldsymbol{p}|^2\right)$, where $D_{\boldsymbol{p}}\boldsymbol{g}^{\text{ell}}(\bar{\boldsymbol{p}})$ is calculated in (69) in Appendix C. Hence a good estimate of the extra total power allocation is

$$\mathbf{1}^{\text{T}}\Delta\boldsymbol{p} \approx \mathbf{1}^{\text{T}}(\boldsymbol{I} - \bar{\boldsymbol{F}} - D_{\boldsymbol{p}}\boldsymbol{g}^{\text{ell}}(\bar{\boldsymbol{p}}))^{-1}\boldsymbol{g}^{\text{ell}}(\bar{\boldsymbol{p}}). \tag{39}$$

Since $\rho(\bar{\boldsymbol{F}} + D_{\boldsymbol{p}}\boldsymbol{g}^{\text{ell}}(\bar{\boldsymbol{p}})) \leq \left\|\bar{\boldsymbol{F}} + D_{\boldsymbol{p}}\boldsymbol{g}^{\text{ell}}(\bar{\boldsymbol{p}})\right\|_2 < 1$ as shown in (70), we have

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

By (20), observe that when $\|\boldsymbol{\epsilon}\|_2$ is small the first term dominates, and the extra power consumption scales with robustness as $\mathcal{O}(\|\boldsymbol{\epsilon}\|_2)$. Estimate in (39) is also plotted in Fig. 6 along with the power-robustness tradeoff curve, and the transient behavior is shown in Fig. 7. We see that the close-form approximation in (39) fits very well with the numerical calculation.

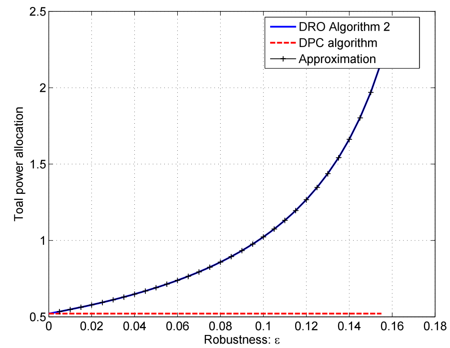
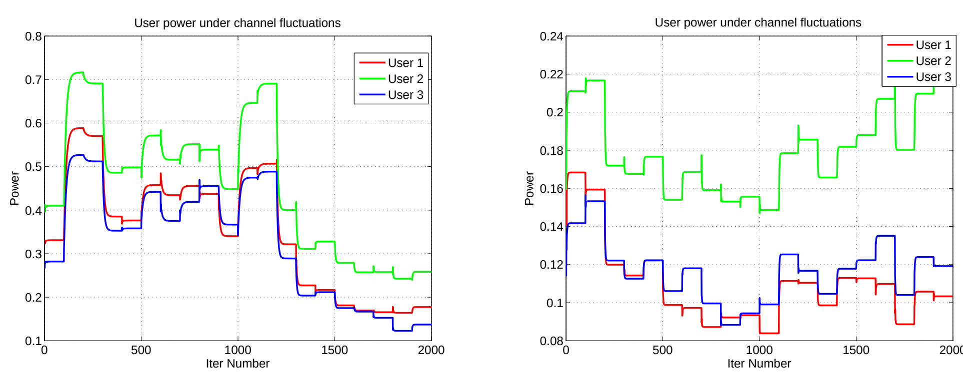


Fig. 6. Robustness-power tradeoff curve under ellipsoid uncertainty set, with approximation in (39).

Remark 3. *We mention the connections between the proposed distributed robust power control scheme and the traditional Gauss-Seidel and Jacobi methods. Indeed, 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 [13]. 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 (22), (23) actually solves a one-dimensional optimization problem with respect to the power of*



(a) Robust algorithm. (b) DPC algorithm.

Fig. 7. User power when the channel undergoes random fluctuation in the uncertainty set for 20 times ($\epsilon = 10\%$).

the i^{th} user, 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.

V. Modeling SIR Measurement Inaccuracy using Polyhedrons

To model SIR measurement errors, consider a row-wise uncertainty polyhedron

$$\mathcal{F}_i = \left\{ \bar{\boldsymbol{F}}_i + \Delta \boldsymbol{F}_i : \left| \sum_{j \neq i} \frac{\Delta F_{ij}}{t_{ij}} \right| \leq 1 \right\}, \tag{40}$$

where $t_{ij} > 0$ parametrizes the maximal deviation of F_{ij} from its nominal value. This uncertainty setup will be very useful in modeling inaccurate SIR measurements, in that the relative error could be equivalently characterized by a uncertainty polyhedron in terms of the normalized channel matrix $\boldsymbol{F}$. Assume that the i^{th} user measures the SIR at its receiver with relative error Δ_i, that is,

$$\frac{\mathsf{SIR}_i}{\overline{\mathsf{SIR}}_i} \in [1 - \Delta_i, 1 + \Delta_i]. \tag{41}$$

where $\overline{\mathsf{SIR}}_i$ is the nominal value corresponding to $\bar{\boldsymbol{F}}$. By (4), (41) is equivalent to

$$\frac{\sum_{j \neq i} \bar{F}_{ij} p_j + v_i}{\sum_{j \neq i} F_{ij} p_j + v_i} \in [1 - \Delta_i, 1 + \Delta_i], \tag{42}$$

that is,

$$\Big| \sum_{j \neq i} \Delta F_{ij} p_j \Big| \leq \Delta_i \Big(\sum_{j \neq i} F_{ij} p_j + v_i \Big) = \frac{\Delta_i p_i \gamma_i}{\mathsf{SIR}_i}. \tag{43}$$

Hence we see that (41) is a polyhedron in (40) with coefficients

$$t_{ij} = \frac{\Delta_i p_i \gamma_i}{\mathsf{SIR}_i p_j}. \tag{44}$$

Next we analyze the corresponding protection function of (40) and distributed solutions. Note that (40) is similar to (19) 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 (40) is

$$g_i^{\text{poly}}(\boldsymbol{p}) = \max_{j \neq i} t_{ij} p_j, \quad i = 1, \ldots, L, \tag{45}$$

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

$$\begin{aligned} \text{minimize} \quad & \mathbf{1}^{\mathrm{T}} \boldsymbol{p} \\ \text{subject to} \quad & \boldsymbol{p} \succeq \bar{\boldsymbol{F}} \boldsymbol{p} + \boldsymbol{g}^{\text{poly}}(\boldsymbol{p}) + \boldsymbol{v} \\ \text{variables} \quad & \boldsymbol{p} \succeq 0. \end{aligned} \tag{46}$$

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

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

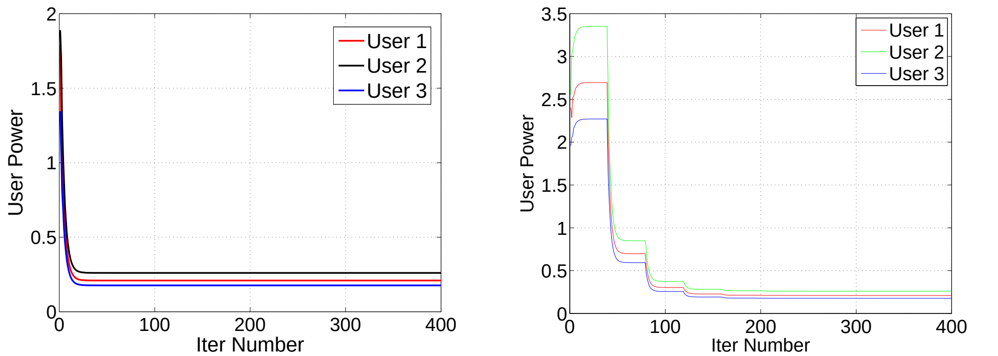
$$p_i(k+1) = \frac{\gamma_i}{\mathsf{SIR}_i} p_i(k) + \max_{j \neq i} t_{ij} p_j(k). \tag{47}$$

In this algorithm, 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 could be implemented in a full distributed fashion. Plugging (44) into (47), we have

$$\begin{aligned} p_i(k+1) =& \frac{\gamma_i}{\mathsf{SIR}_i} p_i(k) + \max_{j \neq i} \frac{\Delta_i p_i(k) \gamma_i}{\mathsf{SIR}_i} \\ =& \frac{\gamma_i}{\mathsf{SIR}_i} (1 + \Delta_i) p_i(k) \approx \frac{\gamma_i}{\mathsf{SIR}_i (1 - \Delta_i)} p_i(k), \end{aligned} \tag{48}$$

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 2, we derive a sufficient condition for feasibility under polyhedron un-



(a) Convergence of Algorithm 3 without feedback delay.

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

Fig. 8. Convergence of Algorithm 3 with and without feedback delay.

certainty set and convergence of Algorithm 3 to the optimality as follows:

Theorem 4. *Problem (46) is feasible, or equivalently, Algorithm 3 converges to the optimal solution of Problem (46) if*

$$\left\|\bar{\boldsymbol{F}}\right\|_1 + \max_j \sum_i t_{ij} < 1, \tag{49}$$

where $\left\|\bar{\boldsymbol{F}}\right\|_1$ *is the induced* ℓ_1*-norm of* $\bar{\boldsymbol{F}}$*, i.e.,* $\left\|\bar{\boldsymbol{F}}\right\|_1 = \sup_{\|\boldsymbol{p}\|_1=1} \left\|\bar{\boldsymbol{F}}\boldsymbol{p}\right\|_1$.

Proof: See Appendix E. ■

Simulation results of Algorithm 3 are shown in Fig. 8, with similar setup as in Fig. 1. The robustness-distributiveness tradeoff curve is plotted in Fig. 9. 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. The tradeoff curve is less smooth compared to Fig. 4, due to linear programming nature of Problem (46), whereas Problem (21) is an SOCP. Also observe that the algorithm for polyhedrons converges faster and needs less message passing than that for ellipsoids.

VI. Modeling user dynamics using D-norm

It is also possible to use D-norm [14] 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 $\langle \boldsymbol{y} \rangle_k = \max\{\sum_{i\in S} |y_i| : S \subset \{1,\ldots,L\}, |S| \le k\}$, where $k \in \{1,\ldots,L\}$, i.e., the sum of the k largest components in $|\boldsymbol{y}|$. It is an equivalent generalization of the ℓ_∞ norm, which corresponds to

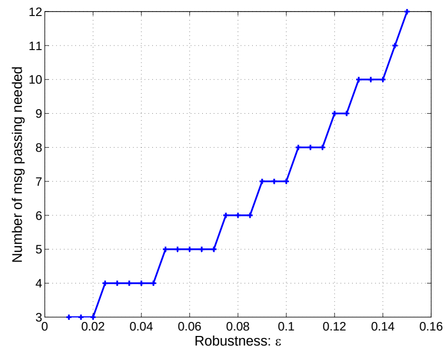


Fig. 9. Robustness-distributiveness tradeoff under polyhedron uncertainty set, where $\epsilon = \max t_{ij}$.

$k = 1$.

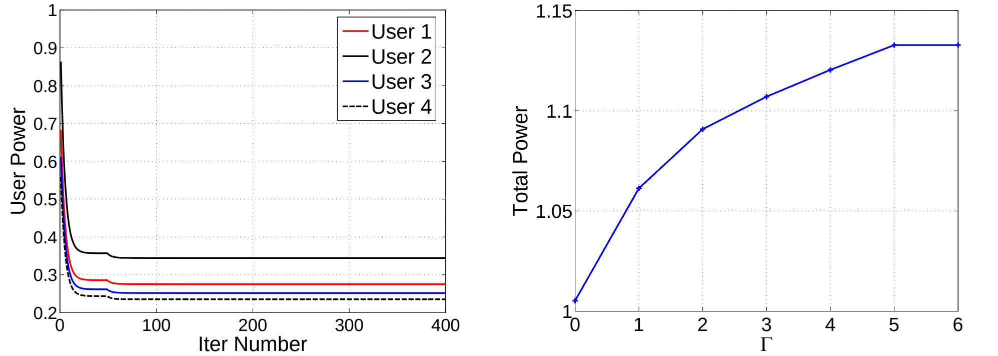
Let L and V be the total number of active and possible *virtual* users (i.e., users who are not active but might turn active) in the system respectively. Let $h_{iv} \in [0, \hat{h}_{iv}]$ denote the relative channel gain (normalized by G_{ii}) of virtual user v's transmitter to active user i's receiver. $\Delta\boldsymbol{F}$ 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:

$$g_i^D(\Gamma_i, \boldsymbol{p}) = \left\langle [\Delta F_{i1}p_1, \ldots, \Delta F_{iL}p_L, \hat{h}_{i1}\bar{p}_1^{\max}, \ldots, \hat{h}_{iV}\bar{p}_V^{\max}]^{\mathrm{T}} \right\rangle_{\Gamma_i}, \tag{50}$$

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 is

$$\begin{array}{rl} \text{minimize} & \mathbf{1}^{\mathrm{T}}\boldsymbol{p} \\ \text{subject to} & \boldsymbol{p} \succeq \bar{\boldsymbol{F}}\boldsymbol{p} + \boldsymbol{g}^D(\boldsymbol{\Gamma}, \boldsymbol{p}) + \boldsymbol{v} \\ \text{variables} & \boldsymbol{p} \succeq 0. \end{array} \tag{51}$$

which can be solved by the following distributed algorithm.



(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. 10. Numerical results for Algorithm 4

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

$$p_i(k+1) = \frac{\gamma_i}{\mathsf{SIR}_i} p_i(k) + g_i(\Gamma_i, \boldsymbol{p}(k)). \tag{52}$$

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

Theorem 5. *Algorithm 4 converges to the optimal solution of Problem (51) if*

$$\left\|\bar{\boldsymbol{F}}\right\|_2 + \sqrt{\|\boldsymbol{\Gamma}\|_\infty}\|\Delta \boldsymbol{F}\|_{\mathrm{F}} < 1, \tag{53}$$

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. 10, 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.

VII. Concluding remarks

DRO is a new framework that combines robustness and distributiveness into optimization models. In this paper we studied the robust power control problem under the DRO framework with respect to channel variation, SIR measurement errors, and user entry. 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. In the paper upper bounds for the maximal spectral radius are developed, hence sufficient for optimality. It is an interesting matrix theoretical problem to evaluate maximal spectral radius exactly in 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.

APPENDIX

A. Proof of Theorem 1

Proof: We complete the proof of Theorem 1 by showing $1) \Leftrightarrow 2)$ and (18).

$1) \Rightarrow 2)$: Suppose $\exists \hat{\boldsymbol{p}} \succeq 0$, such that $\forall \boldsymbol{F} \in \mathcal{F}$,

$$\hat{\boldsymbol{p}} \succeq \boldsymbol{F}\hat{\boldsymbol{p}} + \boldsymbol{v}. \tag{54}$$

Then $\max\limits_j \frac{(\boldsymbol{F}\hat{\boldsymbol{p}})_j}{\hat{p}_j} \leq 1 - \min\limits_j \frac{v_j}{\hat{p}_j}$. Let $c = \min\limits_j \frac{v_j}{\hat{p}_j}$. Then $0 < c < 1$ since $0 \prec \boldsymbol{v} \prec \hat{\boldsymbol{p}}$.

On the other hand, for nonnegative matrices ([15, 5.5.2] and [16, Exercise 2.3.2]) we have the following minimax characterization of spectral radius:

$$\rho(\boldsymbol{F}) = \min_{\boldsymbol{p} \succeq 0} \max_j \frac{(\boldsymbol{F}\boldsymbol{p})_j}{p_j}. \tag{55}$$

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

$2) \Rightarrow 1)$: Suppose that for some $\delta \in (0, 1)$,

$$\sup_{\boldsymbol{F} \in \mathcal{F}} \rho(\boldsymbol{F}) = 1 - \delta. \tag{56}$$

Step 1: Let us first show for $\mathcal{F}$ is *compact* (in the topology on all $L \times L$ matrices induced by the Frobenius norm $\|\cdot\|_{\mathrm{F}}$, or any other equivalent norm). Since $\rho(\boldsymbol{F})$ is continuous in $\boldsymbol{F}$, there exists a $\boldsymbol{F}^*$ such that $\rho(\boldsymbol{F}^*) = 1-\delta$. By Perron-Frobenius Theorem [16, Theorem 2.7]), $1-\delta$ is an eigenvalue of $\boldsymbol{F}^*$ and there exists a corresponding right eigenvector $\boldsymbol{p}^* \succeq 0$. We show that

$$(1-\delta)\boldsymbol{p}^* \succeq \boldsymbol{F}\boldsymbol{p}^*, \quad \forall \boldsymbol{F} \in \mathcal{F}. \tag{57}$$

Assume the opposite, that is, $\exists \boldsymbol{F}' \in \mathcal{F}$ and $l \in \{1,\dots,L\}$ such that

$$(\boldsymbol{F}'\boldsymbol{p}^*)_l > (1-\delta)\boldsymbol{p}^*_l. \tag{58}$$

Construct a new matrix $\boldsymbol{K}$ by assigning its rows as $\boldsymbol{K}_l = \boldsymbol{F}'_l$ and $\boldsymbol{K}_j = \boldsymbol{F}^*_j$ for $j \neq l$. Since $\boldsymbol{F}'_l \in \mathcal{F}_l$, $\boldsymbol{F}^*_j \in \mathcal{F}_j$ for $j \neq l$, and $\mathcal{F}$ is a *row-wise* uncertainty set, we have $\boldsymbol{K} \in \mathcal{F}$.

Observe that by $\boldsymbol{F}^*\boldsymbol{p}^* = (1-\delta)\boldsymbol{p}^*$ and (58), we have

$$\begin{aligned}(\boldsymbol{K}\boldsymbol{p}^*)_l >&(1-\delta)\boldsymbol{p}^*_l,\\ (\boldsymbol{K}\boldsymbol{p}^*)_j \leq&(1-\delta)\boldsymbol{p}^*_j, \quad j \neq l.\end{aligned} \tag{59}$$

Now claim that $\rho(\boldsymbol{K}) > 1-\delta$, which contradicts with (56). To show the claim, denote by $\boldsymbol{q} \succeq 0$ the left-eigenvector of $\boldsymbol{K}$ corresponding to $\rho(\boldsymbol{K})$. Premultiplying $\boldsymbol{K}\boldsymbol{p}^* \succeq (1-\delta)\boldsymbol{p}^*$ by $\boldsymbol{q}^{\mathrm{T}}$ and by (59), we have $\rho(\boldsymbol{K})\boldsymbol{q}^{\mathrm{T}}\boldsymbol{p}^* > (1-\delta)\boldsymbol{q}^{\mathrm{T}}\boldsymbol{p}^*$, i.e., $\rho(\boldsymbol{K}) > 1-\delta$.

Now with (57), we have $(1-\delta)\boldsymbol{p}^* \succeq \max_{\boldsymbol{F}\in\mathcal{F}} \boldsymbol{F}\boldsymbol{p}^*$, where $\max$ is taken componentwise. Since $\boldsymbol{v}$ is fixed, let $\hat{\boldsymbol{p}} = t\boldsymbol{p}^*$ for sufficiently large t, then

$$\hat{\boldsymbol{p}} = t(1-\delta)\boldsymbol{p}^* + \delta t\boldsymbol{p}^* \succ \max_{\boldsymbol{F}\in\mathcal{F}} \boldsymbol{F}\hat{\boldsymbol{p}} + \boldsymbol{v},$$

and the proof for compact $\mathcal{F}$ is completed.

Step 2: For general $\mathcal{F}$, denote $\overline{\mathcal{F}}$ the closure of $\mathcal{F}$. By continuity of $\rho(\boldsymbol{F})$, we have

$$\sup_{\boldsymbol{F}\in\mathcal{F}} \rho(\boldsymbol{F}) = \max_{\boldsymbol{F}\in\overline{\mathcal{F}}} \rho(\boldsymbol{F}). \tag{60}$$

Next we approximate $\overline{\mathcal{F}}$ from interior by *compact* uncertainty sets: let $\overline{\mathcal{F}}_j = \overline{\mathcal{F}} \cap \{\boldsymbol{F} \succeq 0 : \|\boldsymbol{F}\|_{\mathrm{F}} \leq j\}$. Each $\overline{\mathcal{F}}_j$ is compact since it is bounded and closed in $\mathbb{R}^{L\times L}$. Moreover $\overline{\mathcal{F}}_j$ is increasing and $\bigcup_{j\geq 1} \overline{\mathcal{F}}_j = \overline{\mathcal{F}}$. Let $\rho_j = \max_{\boldsymbol{F}\in\overline{\mathcal{F}}_j} \rho(\boldsymbol{F})$ with $\boldsymbol{F}^*_j \in \overline{\mathcal{F}}_j$ satisfying $\rho(\boldsymbol{F}^*_j) = \rho_j$. Then ρ_j is a non-decreasing sequence converging to $\lim_j \rho_j = \sup_{\boldsymbol{F}\in\mathcal{F}} \rho(\boldsymbol{F}) = 1-\delta$. For each j denote by $\boldsymbol{p}^*_j \succeq 0$ the right eigenvector of $\rho(\boldsymbol{F}^*_j)$ corresponding to ρ_j, and $\left\|\boldsymbol{p}^*_j\right\|_2 = 1$. Since

$\mathcal{P} = \{\boldsymbol{p} \succeq 0 : \|\boldsymbol{p}\|_2 = 1\}$ is a bounded subset in $\mathbb{R}^L$, there exists a convergent subsequence $\boldsymbol{p}^*_{j_k}$ converging to some $\boldsymbol{p}^* \in \mathcal{P}$. Now claim that this $\boldsymbol{p}^*$ satisfies

$$(1-\delta)\boldsymbol{p}^* \succeq \boldsymbol{F}\boldsymbol{p}^*. \tag{61}$$

For any $\boldsymbol{F} \in \mathcal{F}$, there exists a K such that $\forall k > K$, $\boldsymbol{F} \in \overline{\mathcal{F}}_{j_k}$. Since $\overline{\mathcal{F}}_{j_k}$ is compact, by (57) in Step 1 we know that $\rho_{j_k}\boldsymbol{p}^*_{j_k} \succeq \boldsymbol{F}^*_{j_k}\boldsymbol{p}^*_{j_k}$ for all $k > K$. Sending $k \to \infty$ on both sides yields (61). Then by the same scaling argument, there exists a $\hat{\boldsymbol{p}}$, such that (54) holds.

Finally, we show (18), that is, the optimal power allocation, i.e., the solution of

$$\boldsymbol{p} = \bar{\boldsymbol{F}}\boldsymbol{p} + \boldsymbol{g}(\boldsymbol{p}) + \boldsymbol{v} = \sup_{\boldsymbol{F}\in\mathcal{F}} \boldsymbol{F}\boldsymbol{p} + \boldsymbol{v} \tag{62}$$

is given by

$$\boldsymbol{p}^* = \sup_{\boldsymbol{F}\in\mathcal{F}} (\boldsymbol{I} - \boldsymbol{F})^{-1}\boldsymbol{v}. \tag{63}$$

where the supremum above is taken componentwise.

Assume for now that $\mathcal{F}$ is compact. Since $\boldsymbol{F}\boldsymbol{p}$ and $(\boldsymbol{I} - \boldsymbol{F})^{-1}\boldsymbol{v}$ are continuous function of $\boldsymbol{F}$, supremum in (62) and (63) could be replaced by maximum. Let $\tilde{\boldsymbol{F}}_i \in \arg\max_{\boldsymbol{F}\in\mathcal{F}}[(\boldsymbol{I} - \boldsymbol{F})^{-1}\boldsymbol{v}]_i$, and $\tilde{\boldsymbol{p}}_i = (\boldsymbol{I} - \tilde{\boldsymbol{F}}_i)^{-1}\boldsymbol{v}$. By definition, $\boldsymbol{p}^*$ is the componentwise maximum of $\{\tilde{\boldsymbol{p}}_1, \dots, \tilde{\boldsymbol{p}}_L\}$. Since

$$\tilde{\boldsymbol{p}}_i = \tilde{\boldsymbol{F}}_i\tilde{\boldsymbol{p}}_i + \boldsymbol{v} \succeq \max_{\boldsymbol{F}\in\mathcal{F}} \boldsymbol{F}\tilde{\boldsymbol{p}}_i + \boldsymbol{v} \succeq \max_{\boldsymbol{F}\in\mathcal{F}} \boldsymbol{F}\boldsymbol{p}^* + \boldsymbol{v},$$

we have $\boldsymbol{p}^* \succeq \max_{\boldsymbol{F}\in\mathcal{F}} \boldsymbol{F}\boldsymbol{p}^* + \boldsymbol{v}$.

On the other hand, let $\boldsymbol{F}^*$ be a componentwise maximizer of $\boldsymbol{F}\boldsymbol{p}^*$, that is, its i^{th} row satisfies $\boldsymbol{F}^*_i \in \arg\max_{\boldsymbol{F}_i\in\mathcal{F}_i} \boldsymbol{F}_i^{\mathrm{T}}\boldsymbol{p}^*$.By the row-wise structure of $\mathcal{F}$, $\boldsymbol{F}^* \in \mathcal{F}$. Then $\boldsymbol{p}^* \succeq \boldsymbol{F}^*\boldsymbol{p}^* + \boldsymbol{v}$ and $\rho(\boldsymbol{F}) < 1$. Hence $(\boldsymbol{I} - \boldsymbol{F}^*)^{-1} = \sum_{k\geq 0}(\boldsymbol{F}^*)^k$ has positive entries. Therefore $\boldsymbol{p}^* \succeq (\boldsymbol{I} - \boldsymbol{F}^*)^{-1}\boldsymbol{v}$. By (63), equality holds in the above equation, and we conclude that $\boldsymbol{F}^*$ simultaneously maximizes $\boldsymbol{F}\boldsymbol{p}^*$ and $(\boldsymbol{I} - \boldsymbol{F})^{-1}\boldsymbol{v}$. Hence $\boldsymbol{p}^* = (\boldsymbol{I} - \boldsymbol{F}^*)^{-1}\boldsymbol{v}$ is a solution to (62). By the uniqueness shown in Theorem 1 before, we complete the proof for compact $\mathcal{F}$.

By approximating $\overline{\mathcal{F}}$ by compact sets from interior similarly as in showing $2) \Rightarrow 1)$, the same conclusion holds for general $\mathcal{F}$. ■

B. Proof of Theorem 2

Proof: By Theorem 1, it is sufficient to check

$$\max_{\boldsymbol{F}\in\mathcal{F}} \rho(\boldsymbol{F}) \leq \left\|\bar{\boldsymbol{F}}\right\|_2 + \|\boldsymbol{\epsilon}\|_2 \tag{64}$$

and

$$\max_{\boldsymbol{F}\in\mathcal{F}} \rho(\boldsymbol{F}) \leq \frac{\rho(\bar{\boldsymbol{F}}+\bar{\boldsymbol{F}}^{\mathrm{T}})}{2} + \|\boldsymbol{\epsilon}\|_2. \tag{65}$$

Recall the Frobenius norm of a matrix is defined as $\|\boldsymbol{F}\|_{\mathrm{F}} = \sqrt{\sum_{i,j}|F_{ij}|^2}$, and $\|\boldsymbol{F}\|_{\mathrm{F}} \geq \|\boldsymbol{F}\|_2$ [17]. 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 (64). By [18, 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

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

Hence (65) holds.

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 (25) reduces to (26). ■

C. Proof of Theorem 3

Proof: Case 1: $M = 1$. We first prove that updates in (28) converge, then show that the limit is the optimal solution of Problem (21).

Step 1: Substituting (4) into (28), we have

$$\boldsymbol{p}(k+1) = \bar{\boldsymbol{F}}\boldsymbol{p}(k) + \boldsymbol{g}^{\mathrm{ell}}(\boldsymbol{p}(k)) + \boldsymbol{v} \triangleq \boldsymbol{\varphi}(\boldsymbol{p}(k)), \tag{66}$$

where $\boldsymbol{g}^{\mathrm{ell}}(\boldsymbol{p})$ is the protection function for ellipsoid uncertainty sets defined in (20). To prove convergence, it suffices to prove that $\boldsymbol{\varphi}(\boldsymbol{p})$ is a contractive mapping.

Since $\|\boldsymbol{\varphi}(\boldsymbol{x})-\boldsymbol{\varphi}(\boldsymbol{y})\|_2 \leq \left\|\bar{\boldsymbol{F}}\right\|_2 \|\boldsymbol{x}-\boldsymbol{y}\|_2 + \left\|\boldsymbol{g}^{\text{ell}}(\boldsymbol{x})-\boldsymbol{g}^{\text{ell}}(\boldsymbol{y})\right\|_2$, and

$$\left\|\boldsymbol{g}^{\text{ell}}(\boldsymbol{x})-\boldsymbol{g}^{\text{ell}}(\boldsymbol{y})\right\|_2^2 = \sum_{i=1}^{L}\left(\epsilon_i\sqrt{\sum_{j\neq i} x_j^2} - \epsilon_i\sqrt{\sum_{j\neq i} y_j^2}\right)^2 \leq \|\boldsymbol{\epsilon}\|_2^2 \|\boldsymbol{x}-\boldsymbol{y}\|_2^2, \tag{67}$$

we have

$$\|\boldsymbol{\varphi}(\boldsymbol{x})-\boldsymbol{\varphi}(\boldsymbol{y})\|_2 \leq \left(\left\|\bar{\boldsymbol{F}}\right\|_2 + \|\boldsymbol{\epsilon}\|_2\right)\|\boldsymbol{x}-\boldsymbol{y}\|_2. \tag{68}$$

Therefore (29) implies that $\boldsymbol{\varphi}(\boldsymbol{p})$ is a contractive mapping. By Contraction Principle [19], (66) converges to a unique fixed point $\boldsymbol{p}^*$.

Step 2: Next we prove that $\boldsymbol{p}^*$ is the optimal solution of (21). The constraints of (21) can be written as $\boldsymbol{f}(\boldsymbol{p}) \triangleq (\boldsymbol{I}-\bar{\boldsymbol{F}})\boldsymbol{p} - \boldsymbol{g}^{\text{ell}}(p) - \boldsymbol{v} \geq 0$. Since Problem (21) is an SOCP hence convex, by the Karush-Kuhn-Tucker theorem [10], it suffices to prove the existence of a Lagrangian multiplier $\boldsymbol{\mu} \succeq 0$ such that $\mathbf{1}^T = \boldsymbol{\mu}^T D_{\boldsymbol{p}}\boldsymbol{f}(\boldsymbol{p}^*)$.

At the fixed point $\boldsymbol{p}^*$, the partial derivative of $\boldsymbol{f}$ with respect to $\boldsymbol{p}$ is $D_{\boldsymbol{p}}\boldsymbol{f}(\boldsymbol{p}^*) = \boldsymbol{I} - \bar{\boldsymbol{F}} - \epsilon D_p\boldsymbol{g}^{\text{ell}}(\boldsymbol{p}^*)$, whose ij^{th} entry is

$$\left[D_{\boldsymbol{p}}\boldsymbol{g}^{\text{ell}}(\boldsymbol{p}^*)\right]_{ij} = \frac{\partial g_i^{\text{ell}}}{\partial p_j} = \begin{cases} 0 & i=j, \\ \frac{\epsilon_i p_j^*}{\sqrt{\|\boldsymbol{p}\|_2^2 - p_i^2}} & i\neq j. \end{cases} \tag{69}$$

Then $\bar{\boldsymbol{F}} + D_{\boldsymbol{p}}\boldsymbol{g}^{\text{ell}}(\boldsymbol{p}^*)$ has nonnegative entries. Also,

$$\left\|\bar{\boldsymbol{F}} + D_{\boldsymbol{p}}\boldsymbol{g}^{\text{ell}}(p^*)\right\|_2 \leq \left\|\bar{\boldsymbol{F}}\right\|_2 + \left\|D_{\boldsymbol{p}}\boldsymbol{g}^{\text{ell}}(p^*)\right\| = \left\|\bar{\boldsymbol{F}}\right\|_2 + \|\boldsymbol{\epsilon}\|_2 < 1, \tag{70}$$

where $\left\|D_{\boldsymbol{p}}\boldsymbol{g}^{\text{ell}}(p^*)\right\| \leq \|\boldsymbol{\epsilon}\|_2$, since by Cauchy-Schwartz inequality, for any $\boldsymbol{x}$ we have

$$\left\|D_{\boldsymbol{p}}\boldsymbol{g}^{\text{ell}}(p^*)\boldsymbol{x}\right\|^2 = \sum_i\left(\sum_{j\neq i}\epsilon_i\frac{p_j^*}{\sqrt{\|\boldsymbol{p}\|_2^2 - p_i^2}}x_j\right)^2 \leq \sum_i \epsilon_i^2 \sum_{j\neq i} x_j^2 \leq \|\boldsymbol{\epsilon}\|_2^2\|\boldsymbol{x}\|_2^2.$$

Therefore, $D_{\boldsymbol{p}}f(\boldsymbol{p}^*) = \boldsymbol{I} - \bar{\boldsymbol{F}} - D_{\boldsymbol{p}}\boldsymbol{g}^{\text{ell}}(\boldsymbol{p}^*)$ is invertible. Note that since $\rho(\bar{\boldsymbol{F}} + D_{\boldsymbol{p}}\boldsymbol{g}^{\text{ell}}(\boldsymbol{p}^*)) \leq \left\|\bar{\boldsymbol{F}} + D_{\boldsymbol{p}}\boldsymbol{g}^{\text{ell}}(\boldsymbol{p}^*)\right\|_2 < 1$,

$$[D_{\boldsymbol{p}}f(\boldsymbol{p}^*)]^{-1} = \left[\boldsymbol{I} - \bar{\boldsymbol{F}} - D_{\boldsymbol{p}}\boldsymbol{g}^{\text{ell}}(\boldsymbol{p}^*)\right]^{-1} = \sum_{k=0}^{\infty}\left(\bar{\boldsymbol{F}} + D_{\boldsymbol{p}}\boldsymbol{g}^{\text{ell}}(\boldsymbol{p}^*)\right)^k,$$

thus $[D_{\boldsymbol{p}}f(\boldsymbol{p}^*)]^{-1}$ has positive entries. Then $\boldsymbol{\mu}^T = [D_{\boldsymbol{p}}f(\boldsymbol{p}^*)]^{-1}\mathbf{1}^T \succeq 0$, which concludes the proof for $M = 1$.

Case 2: $M > 1$. For $1 \leq l \leq M-1$, by definition of $\left\|\bar{\boldsymbol{F}}\right\|_2$, we have

$$\|\boldsymbol{p}(s,l+1)-\boldsymbol{p}(s,l)\|_2 \leq \left\|\bar{\boldsymbol{F}}\right\|_2 \|\boldsymbol{p}(s,l)-\boldsymbol{p}(s,l-1)\|_2\,,$$

where $\boldsymbol{p}(s,M)=\boldsymbol{p}(s{+}1,0)$ and $\boldsymbol{p}(s,-1)=\boldsymbol{p}(s{-}1,M{-}1)$. It follows that, for $0 \leq l \leq m \leq M$,

$$\|\boldsymbol{p}(s,m)-\boldsymbol{p}(s,l)\|_2 \leq \frac{\left\|\bar{\boldsymbol{F}}\right\|_2^l - \left\|\bar{\boldsymbol{F}}\right\|_2^m}{1-\|\boldsymbol{F}\|_2} \|\boldsymbol{p}(s,1)-\boldsymbol{p}(s,0)\|_2\,, \tag{71}$$

and

$$\|\boldsymbol{p}(s+1,0)-\boldsymbol{p}(s,0)\|_2 \leq \frac{1-\left\|\bar{\boldsymbol{F}}\right\|_2^M}{1-\|\boldsymbol{F}\|_2} \|\boldsymbol{p}(s,1)-\boldsymbol{p}(s,0)\|_2\,, \tag{72}$$

Now claim that

$$\|\boldsymbol{p}(s,1)-\boldsymbol{p}(s,0)\|_2 \leq (C_M)^s \|\boldsymbol{p}(0,1)-\boldsymbol{p}(0,0)\|_2\,, \tag{73}$$

where C_M is defined in (31). To see this, by (67), (71) and (72),

$$\begin{aligned}
&\|\boldsymbol{p}(s,1)-\boldsymbol{p}(s,0)\| \\
\leq&\left\|\bar{\boldsymbol{F}}\right\|_2 \|\boldsymbol{p}(s-1,M)-\boldsymbol{p}(s-1,M-1)\|_2 + \|\boldsymbol{\epsilon}\|_2 \|\boldsymbol{p}(s,0)-\boldsymbol{p}(s-1,0)\|_2 \\
\leq& \left\|\bar{\boldsymbol{F}}\right\|_2^M \|\boldsymbol{p}(s-1,1)-\boldsymbol{p}(s-1,0)\|_2 + \frac{\|\boldsymbol{\epsilon}\|_2\left(1-\left\|\bar{\boldsymbol{F}}\right\|_2^M\right)}{1-\left\|\bar{\boldsymbol{F}}\right\|_2} \|\boldsymbol{p}(s-1,1)-\boldsymbol{p}(s-1,0)\|_2 \\
=&C_M \|\boldsymbol{p}(s-1,1)-\boldsymbol{p}(s-1,0)\|_2\,.
\end{aligned}$$

Therefore (73) follows from induction.

The proof for case 2 is as follows. (29) implies $C_M < 1$. For $sM+m \geq tM+l$,

$$\begin{aligned}
&\|\boldsymbol{p}(s,m)-\boldsymbol{p}(t,l)\|_2 \\
\leq &\|\boldsymbol{p}(s,m)-\boldsymbol{p}(s,0)\|_2 + \|\boldsymbol{p}(s,0)-\boldsymbol{p}(t+1,0)\|_2 + \|\boldsymbol{p}(t,M)-\boldsymbol{p}(t,l)\|_2 \\
\leq &\frac{1-\left\|\bar{\boldsymbol{F}}\right\|_2^m}{1-\left\|\bar{\boldsymbol{F}}\right\|_2} \|\boldsymbol{p}(s,1)-\boldsymbol{p}(s,0)\|_2 + \sum_{r=t+1}^{s-1} \frac{1-\left\|\bar{\boldsymbol{F}}\right\|_2^M}{1-\left\|\bar{\boldsymbol{F}}\right\|_2} \|\boldsymbol{p}(r,1)-\boldsymbol{p}(r,0)\|_2 \\
+ &\frac{\left\|\bar{\boldsymbol{F}}\right\|_2^l - \left\|\bar{\boldsymbol{F}}\right\|_2^M}{1-\left\|\bar{\boldsymbol{F}}\right\|_2} \|\boldsymbol{p}(t,1)-\boldsymbol{p}(t,0)\|_2 \\
\leq &\frac{1-\left\|\bar{\boldsymbol{F}}\right\|_2^M}{1-\left\|\bar{\boldsymbol{F}}\right\|_2} \sum_{r=t}^{s} \|\boldsymbol{p}(r,1)-\boldsymbol{p}(r,0)\|_2 \leq \frac{1-\left\|\bar{\boldsymbol{F}}\right\|_2^M}{1-\left\|\bar{\boldsymbol{F}}\right\|_2} \sum_{r=t}^{s} C_M^r \|\boldsymbol{p}(1)-\boldsymbol{p}(0)\|_2 \\
\leq &\frac{1-\left\|\bar{\boldsymbol{F}}\right\|_2^M}{1-\left\|\bar{\boldsymbol{F}}\right\|_2} \frac{(C_M)^t}{1-C_M} \|\boldsymbol{p}(1)-\boldsymbol{p}(0)\|_2 \leq \frac{(C_M)^t}{1-\left\|\bar{\boldsymbol{F}}\right\|_2 - \|\boldsymbol{\epsilon}\|_2} \|\boldsymbol{p}(1)-\boldsymbol{p}(0)\|_2 .
\end{aligned}$$

This shows that $\boldsymbol{p}(k)$ is a Cauchy sequence and hence converges. Similarly we can verify that the limit point is the fixed point of the constraints, and the optimality follows as in case 1. The error bound is obtained by letting $s \to \infty$ above. ∎

D. Proof of Proposition 2

Proof: By (32) and (31),

$$E(M, \|\boldsymbol{\epsilon}\|_2) = \frac{1}{M} \log \frac{1}{\left\|\bar{\boldsymbol{F}}\right\|_2^M + \frac{\|\boldsymbol{\epsilon}\|_2}{1-\left\|\bar{\boldsymbol{F}}\right\|_2}(1-\left\|\bar{\boldsymbol{F}}\right\|_2^M)}. \tag{74}$$

Let $\alpha = \frac{\|\boldsymbol{\epsilon}\|_2}{1-\left\|\bar{\boldsymbol{F}}\right\|_2}$, $\beta = \left\|\bar{\boldsymbol{F}}\right\|_2$, then by the assumption of (29), $\alpha, \beta \in (0,1)$. For $x > 0$ define $L(x) = -\frac{1}{x}\log[(1-\alpha)\beta^x + \alpha]$, then $E(M, \|\boldsymbol{\epsilon}\|_2) = L(M)$.

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 $\|\boldsymbol{\epsilon}\|_2$: directly from (74), we see that $E(M, \|\boldsymbol{\epsilon}\|_2)$ decreases strictly with $\|\boldsymbol{\epsilon}\|_2$. As $\|\boldsymbol{\epsilon}\|_2$ tends to $1-\left\|\bar{\boldsymbol{F}}\right\|_2$, α tends to 1, hence $L(M)$ vanishes.
3) Strict monotonicity in M: note that

$$L'(x) = \frac{h(x)}{x^2[(1-\alpha)\beta^x + \alpha]}, \tag{75}$$

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 (75) is positive, to show $L'(x) < 0$ for $x > 0$ it is sufficient to show that $h(0) = 0$ and

$h'(x) < 0$, which is true because $h'(x) = (1-\alpha)\beta^x \log\beta \log\frac{(\alpha+(1-\alpha)\beta^x)}{\beta^x} < 0$.

■

E. Proof of Theorem 4

Proof: Note that

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

where the last inequality follows from [17]

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

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

F. Proof of Theorem 5

Proof: We follow the same steps of Appendix C. Define $\Delta\boldsymbol{x} = [\Delta F_{i1}x_1, \ldots, \Delta F_{iL}x_L]^{\mathrm{T}}$ and $\Delta\boldsymbol{y}$ similarly. By [14, Proposition 3], $\langle\cdot\rangle_k \le \sqrt{k}\|\cdot\|_2$. Then by triangle inequality of $\langle\cdot\rangle_{\Gamma_i}$,

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

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

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 wirelesscommunication 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. W. Lee, M. Chiang, and R. A. Calderbank, “Utility-optimal random-access control,” to appear in IEEE Trans. Wireless Comm., 2006.

[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] D. P. Bertsekas, *Nonlinear programming 2nd ed.* Belmont, Mass. : Athena Scientific, 1999.

[11] B. W. Levinger, “An inequality for nonnegative matrices,” *Notices of the American Mathematical Society*, vol. 17, p. 260, 1970.

[12] “Mosek.” [Online]. Available: http://www.mosek.com

[13] 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.

[14] D. Bertsimasa, D. Pachamanovab, and M. Simc, “Robust linear optimization under general norms,” *Operations Research Letters*, vol. 32, pp. 510 – 516, 2004.

[15] M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities*. Boston: Allyn and Bacon, 1964.

[16] R. S. Varga, *Matrix iterative analysis*. Englewood Cliffs, N.J: Prentice-Hall, 1962.

[17] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

[18] L. Han, M. Neumann, and M. Tsatsomeros, “Spectral radii of fixed frobenius norm perturbations of nonnegative matrices,” *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 1, pp. 79–92, 1999.

[19] D. P. Bertsekas and J. Tsitsikklis, *Parallel and distributed computation: numerical methods*. Englewood Cliffs, NJ: Prentice Hall,1989, 1999.