Distributed Robust Optimization for Communication Networks

Kai Yang[1], Yihong Wu[2], Jianwei Huang[3], Xiaodong Wang[1], Sergio Verdú[2]

[1]Department of Electrical Engineering, Columbia University, New York, NY, USA
[2]Department of Electrical Engineering, Princeton University, Princeton, NJ, USA
[3]Deparmtent of Information Engineering, The Chinese University of Hong Kong, HKSAR, China
{kyee, wangx}@ee.columbia.edu, {yihongwu, verdu}@princeton.edu, jwhuang@ie.cuhk.edu.hk

Abstract—Robustness of optimization models for networking problems has been an under-explored area. Yet most existing algorithms for solving robust optimization problems are centralized, thus not suitable for many communication networking problems that demand distributed solutions. This paper represents the first step towards building a framework for designing distributed robust optimization algorithms. We first discuss several models for describing parameter uncertainty sets that can lead to decomposable problem structures. These models include general polyhedron, D-norm, and ellipsoid. We then apply these models to solve robust power control in wireless networks and robust rate control in wireline networks. In both applications, we propose distributed algorithms that converge to the optimal robust solution. Various tradeoffs among performance, robustness, and distributiveness are illustrated both analytically and through simulations.

I. Introduction

Despite the importance and success of using optimization theory to study communication and networking problems, most work in this area makes the important assumption that the data defining the constraints and objective function of the optimization problem can be obtained precisely. We call the corresponding problems "nominal". However, in many practical problems, these data are typically inaccurate, time-varying, or uncertain. Solving the nominal optimization problems may lead to poor or even infeasible solutions of the real problems.

Over the last ten years, robust optimization has emerged in operations research community as a field that tackle optimization problems under data uncertainty (e.g., [1]–[5]). The basic idea of the robust optimization is to seek a solution which remains feasible and near-optimal under the perturbation of parameters in the nominal optimization problem. Each robust optimization is defined by three-tuple: a nominal formulation, a definition of robustness, and an uncertainty set. The process of making an optimization formulation robust can be viewed as a mapping that maps from one optimization problem to another. A central question here is when properties such as convexity and decomposability are preserved under such mapping.

So far, most of the work on robust optimization focuses on how to find a proper set to characterize the data uncertainty,

which leads to a tractable robust counter part of the nominal problem. For example, it has been shown that under the assumption of ellipsoid set of data uncertainty, a robust linear optimization problem can be converted into a second-order cone problem; and a robust second-order cone problem can be reformulated as a semi-definite optimization problem [6]. In general, the previous focus in this area is to formulate the robust optimization problem such that it preserves the *convexity* of the original nominal problem, such that we can use effective centralized algorithms (e.g., interior point method) to solve it. Here we will focus instead on the *distributiveness*-preserving formulation of the robust optimization, which is desirable for many practical problems in communications and networking.

In this paper, we first show how to properly define a uncertainty set, which not only captures the data uncertainty in the model but also leads to a distributively solvable optimization problem. Second, in the case where full distributed algorithm is not obtainable, we focus on the investigation of the *tradeoff* between robustness and distributiveness. While distributed computation has long been studied [7], unlike convexity of a problem, distributiveness of an algorithm does not have a widely-agreed definition. It is often quantified by the amount and frequency of communication overhead required, whose tradeoff with the degree of robustness is interesting to study.

In Section II, we review some background of robust optimization, with focus on the characterization of uncertainty sets that are useful for designing distributed algorithms. Applications on robust power control and robust rate control are given in Sections III and IV, where we discuss various tradeoffs between robustness, distributiveness, and performance through both analysis and numerical studies. Conclusions are given in Section V. All proofs can be found in the online technical report [8].

II. Robust Optimization With Linear Constraints

To make our discussions concrete, 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,

$$\begin{aligned} \text{maximize} \quad & f_0\left(\boldsymbol{x}\right) & (1) \\ \text{subject to} \quad & \boldsymbol{Ax} \preceq \boldsymbol{b} \\ \text{variables} \quad & \boldsymbol{x}, \end{aligned}$$

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., [9]–[12]).

The uncertainty of Problem (1) may exist in the objective function f_0, matrix parameter $\boldsymbol{A}$, and vector parameter $\boldsymbol{b}$. In many cases, the uncertainty in objective function f_0 can be converted into uncertainty of the parameters defining constraints [13]. And later in Section IV we show that it is also possible to convert the uncertainty in $\boldsymbol{b}$ into uncertainty in $\boldsymbol{A}$ (although this could be difficult in general). Therefore, in the rest of the paper, we will focus on studying the uncertainty in $\boldsymbol{A}$.

In the robust counterpart of Problem (1), we require the constraints $\boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b}$ to be valid for any $\boldsymbol{A} \in \mathcal{A}$, where $\mathcal{A}$ denotes the uncertainty set of $\boldsymbol{A}$, and the definition of robustness is in the *worst-case* sense [14]. If we allow arbitrary uncertainty set $\boldsymbol{A}$, then the robust optimization problem is difficult to solve even in a centralized manner [15]. In this paper, we will focus on the study of *constraint-wise* (i.e. *row-wise*) uncertainty set, where the uncertainties between different rows in matrix $\boldsymbol{A}$ are decoupled. This restricted class of uncertainty set characterizes the data uncertainty in many practical problems, and it also allows us to convert the robust optimization problem into a formulation that are distributively solvable.

Denote the jth 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:

$$\text{maximize } f_0(\boldsymbol{x}), \tag{2}$$
$$\text{subject to } \boldsymbol{a}_j^T \boldsymbol{x} \leq b_j, \ \forall \boldsymbol{a}_j \in \mathcal{A}_j, \ \forall 1 \leq j \leq M.$$
$$\text{variables } \boldsymbol{x}.$$

We can show that the robust optimization problem (2) can be equivalently written in a form represented by *protection functions* instead of uncertainty sets.

Denote the nominal counterpart of problem (2) with a coefficient matrix $\bar{\boldsymbol{A}}$ (i.e., the values when there is no uncertainty), with the jth row's coefficient $\bar{\boldsymbol{a}}_j \in \mathcal{A}_j$. Then

Proposition 1. *Assume the uncertainty sets $\mathcal{A}_j$ is compact for all j. Problem (2) is equivalent to*

$$\text{maximize} \quad f_0(\boldsymbol{x}), \tag{3}$$
$$\text{subject to} \quad \bar{\boldsymbol{a}}_j^T \boldsymbol{x} + \boldsymbol{g}_j(\boldsymbol{x}) \leq b_j, \ \forall 1 \leq j \leq M.$$
$$\text{variables} \quad \boldsymbol{x},$$

where

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

is the protection function for the jth constraint, which depends on the uncertainty set $\mathcal{A}_j$ and the nominal row $\bar{\boldsymbol{a}}_j$.

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 $\boldsymbol{g}_j(\boldsymbol{x})$.

A. Robust Formulation Defined By General Polyhedron

In this case, the uncertainty set $\mathcal{A}_j$ is a polyhedron characterized by a set of linear inequalities, i.e., $\mathcal{A}_j \triangleq \{\boldsymbol{a}_j : \boldsymbol{D}_j \boldsymbol{a}_j \preceq \boldsymbol{c}_j\}$. The protection function is

$$\boldsymbol{g}_j(\boldsymbol{x}) = \max_{\boldsymbol{a}_j : \boldsymbol{D}_j \boldsymbol{a}_j \preceq \boldsymbol{c}_j} (\boldsymbol{a}_j - \bar{\boldsymbol{a}}_j)^T \boldsymbol{x}, \tag{4}$$

which involves solving a linear programming (LP). We next show that the uncertainty set (or protection function) can be translated into a set of linear constraints. In the jth constraint in (2), if we fix $\boldsymbol{x} = \hat{\boldsymbol{x}}$, we can characterize the set $\forall \boldsymbol{a}_j \in \mathcal{A}_j$ by solving the following LP:

$$v_j^* = \max_{\boldsymbol{a}_j : \boldsymbol{D}_j \boldsymbol{a}_j \preceq \boldsymbol{c}_j} \boldsymbol{a}_j^T \hat{\boldsymbol{x}}. \tag{5}$$

If $v_j^* \leq b_j$, then $\hat{\boldsymbol{x}}$ is feasible for (2). However, this approach is not very useful since it requires solving one LP in (5) for each possible $\hat{\boldsymbol{x}}$. Alternatively, we can take the dual of (5),

$$v_j^* = \min_{\boldsymbol{p}_j : \boldsymbol{D}_j^T \boldsymbol{p}_j \succeq \hat{\boldsymbol{x}}, \boldsymbol{p}_j \succeq \boldsymbol{0}} \boldsymbol{c}_j^T \boldsymbol{p}_j. \tag{6}$$

If we can find a feasible solution $\hat{\boldsymbol{p}}_j$ to (6), and $\boldsymbol{c}_j^T \hat{\boldsymbol{p}}_j \leq b_j$, then we must have $v_j^* \leq \boldsymbol{c}_j^T \hat{\boldsymbol{p}}_j \leq b_j$. We can thus replace constraint in (2) by the following constraints:

$$\boldsymbol{c}_j^T \boldsymbol{p}_j \leq b_j, \ \boldsymbol{D}_j^T \boldsymbol{p}_j \succeq \boldsymbol{x}, \ \boldsymbol{p}_j \succeq \boldsymbol{0}, \ \forall 1 \leq j \leq M, \tag{7}$$

and we now have an equivalent and *deterministic* formulation for Problem (2), where all the constraints are linear.

B. Robust Formulation Defined by D-norm

D-norm approach [13] is another method to model the uncertainty set, and has advantages such as guarantee of feasibility independent of uncertainty distributions and flexibility in terms of tradeoff between robustness and performance.

Consider the jth constraint $\boldsymbol{a}_j^T \boldsymbol{x} \leq b_j$. Denote the set of all uncertain coefficients in $\boldsymbol{a}_j$ as $\mathcal{E}_j$. The size of $\mathcal{E}_j$ is $|\mathcal{E}_j|$, which might be smaller than the total number of coefficients N (i.e., a_{ij} for some i might not have uncertainty). For each $a_{ij} \in \mathcal{E}_j$, assume the actual value falls into the range of $[\bar{a}_{ij} - \hat{a}_{ij}, \bar{a}_{ij} + \hat{a}_{ij}]$, in which $\hat{a}_{ij}$ is a given error bound. Also choose a nonnegative integer $\Gamma_i \leq |\mathcal{E}_j|$. The definition of robustness associated with the D-norm formulation is to maintain feasibility if at most Γ_i out of all possible $|\mathcal{E}_j|$ parameters are perturbed. Let's denote $\mathcal{S}_i$ as the set of Γ_i uncertain coefficients. The above robustness definition can be characterized by the following protection function,

$$\boldsymbol{g}_j(\Gamma_j, \boldsymbol{x}) = \max_{\mathcal{S}_j : \mathcal{S}_j \subseteq \mathcal{E}_j, |\mathcal{S}_j| = \Gamma_j} \sum_{i \in \mathcal{S}_j} \hat{a}_{ij} |x_i|. \tag{8}$$

Notice that if $\Gamma_j = 0$, then $\boldsymbol{g}_j(\Gamma_j, \boldsymbol{x}) = 0$ and the jth constraint is reduced to the nominal constraint, i.e., no protection against uncertainty. If $\Gamma_j = |\mathcal{E}_j|$, then $g_j(\Gamma_j, \boldsymbol{x}) = \sum_{i \in \mathcal{E}_j} \hat{a}_{ij} |x_i|$ and the jth constraint becomes Soyster's worst-case formulation [13]. The tradeoff between robustness and performance can be obtained by adjusting Γ_j.

Note that the nonlinearity of $\gamma_j(\Gamma_j, \boldsymbol{x})$ is difficult to deal with in the constraint. Alternatively, we can reformulate it into

the following LP problem,

$$\max_{\{0 \leq s_{ij} \leq 1\}_{\forall i \in \mathcal{E}_j}} \sum_{i \in \mathcal{E}_j} \hat{a}_{ij} |x_i| s_{ij}, \text{ s.t. } \sum_{i \in \mathcal{E}_j} s_{ij} \leq \Gamma_j. \quad (9)$$

Taking the dual of Problem (9), we have

$$\min_{\{p_{ij} \geq 0\}_{\forall i \in \mathcal{E}_j}, q_j \geq 0} q_j \Gamma_j + \sum_{i \in \mathcal{E}_j} p_{ij}, \text{ s.t. } q_j + p_{ij} \geq \hat{a}_{ij} |x_i|, \forall i \in \mathcal{E}_j.$$
$$(10)$$

Similar to Section II-A, we can substitute (10) into the robust Problem (2) to obtain an equivalent formulation:

$$\text{maximize} \quad f_0(\boldsymbol{x}) \quad (11)$$
$$\text{subject to} \quad \sum_i a_{ij} x_i + q_j \Gamma_j + \sum_{i \in \mathcal{E}_k} p_{ij} \leq b_j, \forall j,$$
$$q_j + p_{ij} \geq \hat{a}_{ij} y_i, \ \forall i \in \mathcal{E}_k, \forall j,$$
$$- y_i \leq x_i \leq y_i, \forall i,$$
$$\text{variables} \quad \boldsymbol{x}, \ \boldsymbol{y} \succeq \boldsymbol{0}, \ \boldsymbol{p} \succeq \boldsymbol{0}, \ \boldsymbol{q} \succeq \boldsymbol{0}.$$

The new problem only has linear constraints. We provide such an example in Section IV.

C. Robust Formulation Defined by Ellipsoid

Ellipsoid is commonly used to approximate complicated uncertainty sets based on statistical reasons [15] and to succinctly describe a set of discrete points in Euclidean geometry [14]. Here we consider the case where coefficient a_j falls in an ellipsoid centered at the nominal $\bar{a}_j$. Specifically,

$$\mathcal{A}_j = \{\bar{\boldsymbol{a}}_j + \Delta \boldsymbol{a}_j : \sum_i |\Delta a_{ij}|^2 \leq \epsilon_j^2\}. \quad (12)$$

The protection function is given by

$$g_j(\boldsymbol{x}) = \max_{\boldsymbol{a}_j^T \in \mathcal{A}_j} \left(\boldsymbol{a}_j^T - \bar{\boldsymbol{a}}_j^T\right) \boldsymbol{x} = \epsilon_j \sqrt{\sum_i x_i^2}, \quad (13)$$

which can be derived through Cauchy-Schwartz inequality. Although the resulting constraint in Problem (2) is not readily decomposable using standard decomposition techniques, we will see in Section III that this leads to tractable formulation in some important applications (e.g., power control) where users can obtain most network information through local measurement without global message passing.

III. APPLICATION: DISTRIBUTED ROBUST POWER CONTROL

A. The Nominal Problem

Consider the following system model as in the seminal work by Foschini and Miljanic [9]. There exists a set of $\mathcal{L} = \{1, ..., L\}$ 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 single cellular network. The signal to interference ratio (SIR) on the link of user i is

$$\text{SIR}_i = \frac{G_{ii} p_i}{\sum_{j \neq l} G_{ij} p_j + n_i} \quad (14)$$

where G_{ij} is the channel gains from user j's transmitter to user i' 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:

$$\begin{aligned} \text{minimize} \quad & \sum_{i \in \mathcal{L}} p_i \\ \text{subject to} \quad & \text{SIR}_i(\boldsymbol{p}) \geq \gamma_i, \forall i \in \mathcal{L} \\ \text{variables} \quad & p_i \geq 0, \forall i \in \mathcal{L} \end{aligned} \quad (15)$$

The constraints of Problem (15) can be equivalently represented as:

$$(\boldsymbol{I} - \boldsymbol{F})\boldsymbol{p} \succeq \boldsymbol{v} \quad (16)$$

where vector $\boldsymbol{v} = \left[\frac{\gamma_1 n_1}{G_{11}}, \ldots, \frac{\gamma_L n_L}{G_{LL}}\right]$, $\boldsymbol{I}$ is the $L \times L$ identity matrix, and $\boldsymbol{F} = [F_{ij}]$ with

$$F_{ij} = \begin{cases} 0 & i = j \\ \frac{G_{ij} \gamma_i}{G_{ii}} & i \neq j \end{cases}. \quad (17)$$

It has been proved in [9] that if the spectral radius $\|\boldsymbol{F}\| < 1$, the unique global optimal solution of Problem (15) is

$$\boldsymbol{p}^* = (\boldsymbol{I} - \boldsymbol{F})^{-1}\boldsymbol{v}. \quad (18)$$

Furthermore, if each user i locally measures its SIR value $\text{SIR}_i(k)$ at each time slot k, and updates its transmission power by

$$p_i(k+1) = \frac{\gamma_i}{\text{SIR}_i(k)}, \forall i \in \mathcal{L}, \quad (19)$$

the system will globally converge to the optimal solution in (18). We refer to this fully distributed power control algorithm in (19) as the *FM algorithm*. In the rest of the section, we will consider the robust optimization problems under either uncertainties in channel coefficients $\boldsymbol{F} = [F_{ij}]$ or randomness in terms of users entering and leaving the system.

B. Robust Formulation under Ellipsoid Uncertainty Set

In this section we consider the uncertainty in channel matrix $\boldsymbol{F}$ due to fluctuation of the channels. Let $\mathcal{F}_i$ denotes the uncertainty set of the ith row of matrix $\boldsymbol{F}$, 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 the sources of uncertainty. In this section we use ellipsoid to characterize set $\mathcal{F}_i$ for all i, as considered in Section II-C.

Assume the actual normalized channel gain between user j's transmitter and user i's receiver as $F_{ij} + \Delta F_{ij}$, where F_{ij} is the nominal value (e.g., long term average value). Further denote the ith row of $\boldsymbol{F}$ as $\boldsymbol{F}_i$ and the corresponding uncertainty as $\Delta \boldsymbol{F}_i$. Then the uncertainty set $\mathcal{F}_i$ under ellipsoid approximation can be represented by

$$\mathcal{F}_i = \{\boldsymbol{F}_i + \Delta \boldsymbol{F}_i : \sum_{j \neq i} |\Delta F_{ij}|^2 \leq \epsilon_i^2.\} \quad (20)$$

Notice that we always have $F_{ii} = 0$. The robust counterpart

of the nominal Problem (15) is

$$\text{minimize} \quad \sum_{i \in \mathcal{L}} p_i$$
$$\text{subject to} \quad p_i - \sum_{j \neq i}(F_{ij} + \Delta F_{ij})p_j \leq v_i \quad \forall i \in \mathcal{L}$$
$$\sum_{j \neq i}|\Delta F_{ij}|^2 \leq \epsilon_i^2, \forall i \in \mathcal{L} \quad (21)$$
$$\text{variables} \quad p_i \geq 0, i \in \mathcal{L}$$

By (13), we can transform the constraints in (21) into

$$p_i - \sum_{j \neq i} F_{ij}p_j - \epsilon_i \sqrt{\sum_{j \neq i} p_j^2} \geq v_i, \forall i \in \mathcal{L} \quad (22)$$

If we further define matrix $\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{F}$, then Problem (21) can be written equivalently as

$$\text{minimize} \quad \sum_{i \in \mathcal{L}} p_i$$
$$\text{subject to} \quad \epsilon_i \sqrt{\sum_{j \neq i} p_j^2} \leq \boldsymbol{A}_i^{\mathrm{T}}\boldsymbol{p} - v_i, \ i \in \mathcal{L} \quad (23)$$
$$\text{variables} \quad p_i \geq 0, \forall i \in \mathcal{L}$$

This is the problem that we will solve in this subsection.

1) Distributed Algorithm with Delayed Feedback: Let us first consider the following power update of each user i at time slot k,

$$p_i(k+1) = \frac{\gamma_i}{\mathsf{SIR}_i} p_i(k) + \epsilon_i \sqrt{Q^2(k) - p_i^2(k)} \quad (24)$$

where

$$Q(k) = \sqrt{\sum_{j \in \mathcal{L}} p_j^2(k)}. \quad (25)$$

Note that (24) is a modified version of the FM algorithm in (19) with a protection function $\epsilon_i \sqrt{Q^2(k) - p_i^2(k)}$ for each user i. Similar as (19), it involves no diminishing step size or dual variables. The only coupled parameter is $Q(k)$, which needs to be updated at every time slot k.

It we allow Q to be updated at a slower pace, we can increase the distributiveness of the updates. 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 also 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 1. *For each time slot $k = (s, l)$, each user i updates its transmission power as*

$$p_i(k+1) = \boldsymbol{F}_i^T p(k) + v_i + \epsilon_i \sqrt{Q^2(s, 0) - p_i^2(k)}. \quad (26)$$

In Algorithm 1, each user i broadcasts its power $p_i(k)$ and computes $Q(k)$ using (25) 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.
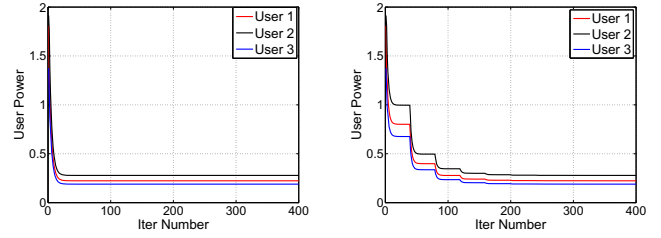
(a) Without feedback delay ($M = 1$)   (b) With feedback delay ($M = 40$).

Fig. 1. Convergence of Algorithm 1.

Algorithm 1 globally converges to the optimal solution of Problem (23) under proper technical conditions of matrix $\boldsymbol{F}$ and uncertainty parameter ϵ_i. Furthermore, we can bound the difference between the current power vector and the optimal power vector during each step of the iteration. The algorithm converges exponentially and the speed of convergence increases as M decreases.

Theorem 1. *Assume $\|\epsilon\| + \|\boldsymbol{F}\| < 1$, where $\|\epsilon\|^2 = \sum \epsilon_i^2$. Algorithm 1 globally converges to the optimal solution of Problem (23), denoted as $\boldsymbol{p}^*$. Moreover,*

$$\|\boldsymbol{p}(k) - \boldsymbol{p}^*\| \leq \frac{1 - \|\boldsymbol{F}\|^M}{1 - \|\boldsymbol{F}\|} \frac{(C_M)^{\lfloor k/M \rfloor}}{1 - C_M} \|\boldsymbol{p}(1) - \boldsymbol{p}(0)\| \quad (27)$$

where

$$C_M = \|\boldsymbol{F}\|^M + \frac{\|\epsilon\|}{1 - \|\boldsymbol{F}\|}(1 - \|\boldsymbol{F}\|^M), \quad (28)$$

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

Notice that the condition of $\|\boldsymbol{F}\| < 1$ is needed for the convergence of the original FM algorithm in (19), and the maximal robustness the algorithm can guarantee is characterized by $1 - \|\boldsymbol{F}\|$.

2) Numerical Results and Performance Comparison: We simulate the performance of Algorithm 1 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 [16]) exponentially fast.

We also compare the performance of Algorithm 1 and the original FM 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 1 avoids channel outage since it considers the worst case of the uncertainty set, and the original FM algorithm leads to frequent channel outages.

3) 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, if more
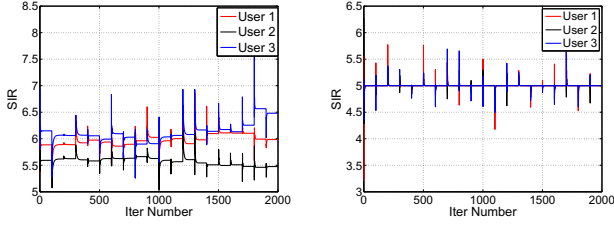
(a) Algorithm 1 avoids channel outage. (b) FM algorithm leads to frequent channel outage.

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

Fig. 3.   robustness-distributiveness tradeoff curve under ellipsoid uncertainty set.

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 understand this tradeoff theoretically, we use (27), which gives an upper bound on the convergence rate of Algorithm 1. A sufficient condition to achieve the desired optimality gap δ is

$$\frac{1 - \|\boldsymbol{F}\|^M}{1 - \|\boldsymbol{F}\|} \frac{C_M^{N/M}}{1 - C_M} \|\boldsymbol{p}(1) - \boldsymbol{p}(0)\| \leq \delta. \tag{29}$$

Let $\alpha = \frac{\|\epsilon\|}{1 - \|\boldsymbol{F}\|}$, $L(M) = \frac{1}{M} \log C_M$, and the equality hold in (29), i.e.,

$$L(M) = \frac{1}{N} \log \left[\frac{\delta \left(1 - \|\boldsymbol{F}\| \right) \left(1 - \alpha \right)}{\|\boldsymbol{p}(1) - \boldsymbol{p}(0)\|} \right]. \tag{30}$$

Since $\alpha < 1$, $\|\boldsymbol{F}\| < 1$, and $\log C_M$ decreases in M, thus $L(M)$ is a monotonically decreasing function and has an inverse function. We can solve (30) and obtain a lower bound on the largest allowed value of update interval M, i.e.,

$$M(\epsilon) \geq L^{-1} \left[\frac{1}{L} \log \left(\frac{\delta \left(1 - \|\boldsymbol{F}\| \right) \left(1 - \alpha \right)}{\|\boldsymbol{p}(1) - \boldsymbol{p}(0)\|} \right) \right]. \tag{31}$$

Then an upper bound on the total number of message passing for reaching an optimality gap of δ with an uncertainty ellipsoid of radius ϵ and a total of N iterations is

$$\frac{N}{L^{-1} \left\{ \frac{1}{N} \log \left[\frac{\delta(1 - \|F\|)}{\|p(1) - p(0)\|} \left(1 - \frac{\|\epsilon\|}{1 - \|F\|} \right) \right] \right\}}. \tag{32}$$

Here one message passing corresponds to each user announcing his power level once, or the base station evaluating the current Q in (25) and broadcasting it to the users. This upper bound is also plotted in Fig. 3, together with the simulated result. 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 3-dimensional tradeoff among robustness ϵ, optimality gap δ, and the number of message passing is given in Fig. 4.

C. Robust Formulation Under Polyhedron Uncertainty Set

Instead of modeling channel uncertainty using ellipsoid as in the last subsection, we can also model it using polyhedron
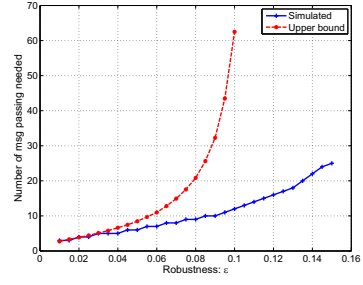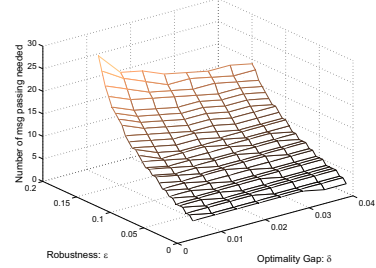

Fig. 4.   Three-dimensional tradeoff surface between robustness ϵ, optimality gap δ and the number of message passing.

as discussed in Section III-B. In particular, we consider a row-wise uncertainty set

$$\mathcal{F}_i = \{\boldsymbol{F}_i + \Delta \boldsymbol{F}_i : \sum_{j \neq i} |\Delta F_{ij}|/t_{ij} \leq 1\}, \tag{33}$$

where $t_{ij} > 0$ are weight coefficients, corresponding to the maximal deviation of F_{ij} from its nominal value. This uncertainty setup will be very useful in modeling the uncertainty in the SIR measurement [8]. Note that (33) is similar to (20) with quadratic terms replaced by linear terms. Since

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

we can derive the robust formulation similar to Problem (23):

$$\begin{aligned}
&\text{minimize} && \sum_{i \in \mathcal{L}} p_i \\
&\text{subject to} && p_i \geq \sum_{j \neq i} F_{ij} p_j + \max_{j \neq i} t_{ij} p_j + v_i, \forall i \in \mathcal{L} \quad (34)\\
&\text{variables} && p_i \geq 0, \forall i \in \mathcal{L}
\end{aligned}$$

A distributed algorithm with limited message passing is derived similar to Algorithm 1.

Algorithm 2. *User i updates its power at time k accordingly to*

$$p_i(k + 1) = \frac{\gamma_i}{\mathsf{SIR}_i} p_i(k) + \max_{j \neq i} t_{ij} p_j(k). \tag{35}$$

In this algorithm, the largest two of user's weighted power need to be communicated globally, through broadcasting of the users. Convergence and performance bounds can be proved similar to Theorem 1.

D. Modeling users' entering and leaving the system with D-norm

It is also possible to use D-norm to model the uncertainty due to both channel fluctuation and users randomly entering the system. Let N and U 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. Also let $\boldsymbol{p} \triangleq [p_1, p_2, ..., p_N]^T$, $\bar{\boldsymbol{p}} \triangleq [\bar{p}_1, \bar{p}_2, ..., \bar{p}_V]^T$, $\bar{p}_v^{\min}$ and $\bar{p}_v^{\max}$ as the lower and upper bound of the transmit power from v^{th} virtual user.

Consider the following protection function for the ith constraint with uncertainty parameter Γ_i :

$$g_i(\Gamma_i, \boldsymbol{p}^*) = \max_{(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathcal{H}} \sum_{j \neq i} |\Delta F_{ij}| \alpha_{ij} p_j^* + \sum_v \hat{h}_{iv} \beta_{iv} \bar{p}_v^{\max},$$

with $\mathcal{H} = \{(\boldsymbol{\alpha}, \boldsymbol{\beta}) | \sum_{j \neq i} \alpha_{ij} + \sum_v \beta_{iv} \leq \Gamma_i, \alpha_{ij}, \beta_{iv} \in \{0, 1\}\}$. 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 $\hat{h}_{iv} \beta_{iv} \bar{p}_v^{\max}$ for all $j \neq i$ and v in the descending order, and choose the Γ_i largest elements as the solution. Notice this can be done at the base station and send to each mobile station. The robust power control problem under both channel fluctuation and user entering uncertainty is

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^N p_j \\ \text{subject to} \quad & p_i - \sum_{j \neq i} F_{ij} p_j - g_i(\Gamma_i, \boldsymbol{p}) \geq v_i, \quad \forall i, \\ \text{variable} \quad & \boldsymbol{p} \succeq \boldsymbol{0}. \end{aligned} \tag{36}$$

which can be solved by the following distributed algorithm.

Algorithm 3. *Each user i updates its transmission power in time slot k as*

$$p_i(k+1) = \frac{\gamma_i}{\mathsf{SIR}_i} p_i(k) + g_i(\Gamma_i, \boldsymbol{p}(k)). \tag{37}$$

where $g_i(\Gamma_i, \boldsymbol{p})$ is computed by the base station and broadcast to each user in the downlink channel.

Optimality, convergence and performance bound of Algorithm 3 can be similarly proved as in the previous two subsections.

E. Related work on robust power control

Reference [18] 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 the 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 [17] the tradeoff between the robustness and the extra power consumed is studied with penalty-defined formulation, while the key focus

of this paper is to study the tradeoff between robustness and distributiveness. Moreover, the algorithm in [17] is primal-dual based and involves centralized computation by the base station, while Algorithm 1 we proposed has less complexity and only requires global message passing of a single parameter.

IV. APPLICATION: DISTRIBUTED ROBUST RATE CONTROL

A. Nominal and Robust Formulations

Consider a wireline network where some links might fail due to reasons such as human mistakes, software bugs, hardware defects, or natural hazard. Network operators typically reserve some bandwidth for backup paths. When the primary paths fail, some or all of the traffic will be re-routed to the corresponding disjoint backup paths. Thus fast system recovery schemes are required to guarantee service availability in the presence of link failure. There are three key components for fast system recovery [19]: identifying a backup path disjoint from the primary path, computing network resource (such as bandwidth) in reservation prior to link failure, detecting the link failure in real-time and re-route the traffic. The first component has been investigated extensively in graph theory. The third component has been extensively studied in system research community. Here we consider the robust rate control and bandwidth reservation in the face of possible failure of primary path, which is related to the second component.

First consider the nominal problem with no link failures. Following similar notations as in Kelly's seminal work [12], we consider a network with S users, L links and T paths, indexed by s, l and t, respectively. Each user is a unique flow from one source node to one destination node. There could be multiple users between the same source-destination node pair. The network is characterized by the $L \times T$ path-availability $0 - 1$ matrix

$$[\boldsymbol{D}]_{lt} = \begin{cases} d_{lt} = 1, & \text{if link } l \text{ is on path } t, \\ 0, & \text{otherwise.} \end{cases}$$

and $T \times S$ primary-path-choice nonnegative matrix

$$[\boldsymbol{W}]_{ts} = \begin{cases} w_{ts}, & \text{if user } s \text{ uses path } t \text{ as the primary path,} \\ 0, & \text{otherwise.} \end{cases}$$

where w_{ts} indicates the percentage that user s allocates its rate to primary path t, and satisfies $w_{ts} > 0$ and $\sum_t w_{ts} = 1$. Let $\boldsymbol{x}$, $\boldsymbol{c}$, and $\boldsymbol{y}$ denote source rates, link capacities, and aggregated path rates, respectively. The nominal multi-path rate control problem is

$$\begin{aligned} \text{maximize} \quad & \sum_s f_s(x_s) \\ \text{subject to} \quad & \boldsymbol{Dy} \preceq \boldsymbol{c}, \quad \boldsymbol{Wx} \preceq \boldsymbol{y}, \\ \text{variables} \quad & \boldsymbol{x} \succeq \boldsymbol{0}, \boldsymbol{y} \succeq \boldsymbol{0}, \end{aligned} \tag{38}$$

where $f_s(x_{s)}$ is the utility of user s, which is increasing and strictly concave in x_s.

In order to guarantee the data transmission is robust against the link failure, each user also determines a backup path when it joins the network. The nonnegative backup path choice

matrix is

$$[\boldsymbol{B}]_{ts} = \begin{cases} b_{ts}, & \text{if user } s \text{ uses path } t \text{ as the backup path,} \\ 0, & \text{otherwise.} \end{cases}$$

where b_{ts} indicates the maximum percentage that user s allocates its rate to path t and satisfies $b_{ts} > 0$. The actual rate allocation will be a random variable between 0 and b_{ts}, depending on whether the primary paths fail. We further assume that a path can only be used as either a primary path or a backup path for the same user. The corresponding robust multi-path routing rate allocation problem is given by

$$\text{maximize} \quad \sum_s f_s(x_s) \tag{39}$$
$$\text{subject to} \quad \boldsymbol{D}\boldsymbol{y} \preceq \boldsymbol{c},$$
$$\sum_s w_{ts}x_s + g_t(\boldsymbol{b}_t, \boldsymbol{x}) \le y_t, \quad \forall t.$$
$$\text{variables} \quad \boldsymbol{x} \succeq \boldsymbol{0}, \boldsymbol{y} \succeq \boldsymbol{0}.$$

Here $\sum_s w_{ts}x_s$ denotes the aggregate rate from users who utilize path t as their primary path, and $g_t(\boldsymbol{b}_t, \boldsymbol{x})$ corresponds to the protection function for the traffic from users who use path t as their backup path, and $\boldsymbol{b}_t$ is the tth row of matrix $\boldsymbol{B}$. There are many ways of characterizing the protection function. Here we consider the choice of D-norm.

Let $\mathcal{E}_t = \{s : b_{ts} > 0, \forall s\}$ denote the set of users who utilize path t as the backup path, and $\mathcal{F}_{t,\Gamma_t}$ denote a subset of $\mathcal{E}_t$ with size Γ_t, where $0 \le \Gamma_t \le |\mathcal{E}_t|$ and controls the tradeoff between robustness and performance. Then the protection function is

$$g_t(\boldsymbol{b}_t, \boldsymbol{x}) = \max_{\mathcal{F}_{t,\Gamma_t} \subseteq \mathcal{E}_t} \sum_{s \in \mathcal{F}_{t,\Gamma_t}} b_{ts}x_s, \forall t. \tag{40}$$

B. Distributed Algorithms

Following the approach in Section II-B, we can convert the robust optimization problem into an equivalent problem with only linear constraints and solve it distributively by dual-based decompositions.

This approach, however, leads to a large amount of extra message passing (due to the new auxiliary variables and constraints) and is computationally expensive to calculate local projections. In this section, we propose a fast distributed algorithm based on a combination of column generation method [20] and dual-based decomposition method.

We first show that the nonlinear constraints in Problem (39) can be replaced by a set of linear constraints:

Proposition 2. *For any path t, the constraint*

$$\sum_s w_{ts}x_s + g_t(\boldsymbol{b}_t, \boldsymbol{x}) \le y_t, \tag{41}$$

is equivalent to the following set of constraints

$$\sum_s w_{ts}x_s + \sum_{s \in \mathcal{F}_{t,\Gamma_t}} b_{ts}x_s \le y_t, \quad \forall \mathcal{F}_{t,\Gamma_t} \in \mathcal{E}_t. \tag{42}$$

Based on Proposition 2, we can convert robust optimization problem (39) into a problem with only linear constraints. However, the number of new linear constraints grows approximately in the order of $M\Gamma_t$, where M is the number of linear constraints in the nominal problem. More importantly, we found the resultant new optimization problem is difficult to solve by the dual decomposition method in a distributed fashion. This motivates us to design an alternative sequential optimization algorithm.

Let $\mathcal{H}_t = \{\mathcal{F}_{t,\Gamma_t} | \mathcal{F}_{t,\Gamma_t} \subseteq \mathcal{E}_t\}$ denote the set of all subsets of $\mathcal{E}_t$ with size Γ_t. The basic idea is to iteratively generate a set $\bar{\mathcal{H}}_t \subseteq \mathcal{H}_t$, and use the following set of constraints to approximate (42):

$$\sum_s w_{ts}x_s + \sum_{s \in \mathcal{F}_{t,\Gamma_t}} b_{ts}x_s \le y_t, \quad \forall \mathcal{F}_{t,\Gamma_t} \in \bar{\mathcal{H}}_t. \tag{43}$$

This leads to a relaxed approximation of Problem (39):

$$\text{maximize} \quad \sum_s f_s(x_s) \tag{44}$$
$$\text{subject to} \quad \boldsymbol{D}\boldsymbol{y} \preceq \boldsymbol{c},$$
$$\sum_s w_{ts}x_s + \sum_{s \in \mathcal{F}_{t,\Gamma_t}} b_{ts}x_s \le y_t, \quad \forall \mathcal{F}_{t,\Gamma_t} \in \bar{\mathcal{H}}_t, \quad \forall t.,$$
$$\text{variables} \quad \boldsymbol{x} \succeq \boldsymbol{0}, \boldsymbol{y} \succeq \boldsymbol{0}.$$

Let $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ denote an optimal solution of (44) and $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$ denote an optimal solution of (39). If $\bar{\mathcal{H}}_t = \mathcal{H}_t$, then we have $\sum_s f_s(x_s^*) = \sum_s f_s(\bar{x}_s)$. Even if $\bar{\mathcal{H}}_t \subset \mathcal{H}_t$, the two optimal objective values can still be the same as shown in the following theorem:

Theorem 2. $\sum_s f_s(x_s^*) = \sum_s f_s(\bar{x}_s)$ *if the following condition holds*

$$g_t(\boldsymbol{b}_t, \boldsymbol{x}^*) = \max_{\mathcal{F}_{t,\Gamma_t} \in \bar{\mathcal{H}}_t} \sum_{s \in \mathcal{F}_{t,\Gamma_t}} b_{ts}x_s^*, \forall t. \tag{45}$$

Next we develop a distributed algorithm (Algorithm 4) to solve Problem (44) for a fixed $\bar{\mathcal{H}}_t$ for each t, which is suboptimal for solving Problem (39). We then design an optimal distributed algorithm (Algorithm 5) that achieves the optimal solution of Problem (39) by iteratively using Algorithm 4.

We first give an equivalent representation of Problem (44) to facilitate the presentation of our distributed algorithms. For each path t, we let $\mathcal{F}_{t,\Gamma_t}(i)$ represent the ith element in set $\bar{\mathcal{H}}_t$, and define a group of auxiliary variables $\{y_{ti}, 1 \le i \le |\bar{\mathcal{H}}_t|\}$. It can be shown that

Proposition 3. *Consider the case where link l is on path t, i.e., $d_{lt} = 1$. Then, given l and t, the set of constraints,*

$$\sum_{j:j \ne t} d_{lj}y_j + d_{lt}y_t \le c_l,$$
$$\sum_s w_{ts}x_s + \sum_{s \in \mathcal{F}_{t,\Gamma_t}} b_{ts}x_s \le y_t, \quad \forall \mathcal{F}_{t,\Gamma_t} \in \bar{\mathcal{H}}_t, \tag{46}$$

are equivalent to the following set of constraints

$$\sum_{j:j \ne t} d_{lj}y_j + d_{lt}y_{ti} \le c_{lti}, c_{lti} = c_l, \ 1 \le i \le |\bar{\mathcal{H}}_t|,$$
$$\sum_s w_{ts}x_s + \sum_{s \in \mathcal{F}_{t,\Gamma_t}} b_{ts}x_s \le y_{ti}, \ 1 \le i \le |\bar{\mathcal{H}}_t|. \tag{47}$$

With the auxiliary variables $\{y_{ti}\}$ and $\{c_{lti}\}$, we can convert

(44) into the following form,

$$\text{maximize} \quad \sum_s f_s(x_s) \tag{48}$$

$$\text{subject to} \quad \bar{D}\bar{y} \preceq \bar{c},$$

$$\sum_s w_{ts}x_s + \sum_{s \in \mathcal{F}_{t,\Gamma_t}(i)} b_{ts}x_s \leq y_{ti}, \ 1 \leq i \leq |\bar{\mathcal{H}}_t|,,$$

$$\text{variables} \quad \boldsymbol{x} \succeq \boldsymbol{0}, \boldsymbol{y} \succeq \boldsymbol{0},$$

where $\bar{y} = \{\{y_{ti}\}_{i=1}^{|\bar{\mathcal{H}}_t|}\}_{t=1}^T$, and $\bar{c}$ and $\bar{D}$ are defined similarly from c and D, respectively.

By relaxing the constraints in Problem (48) using dual variables $\boldsymbol{\lambda} = \{\{\{\lambda_{lti}\}_{i=1}^{|\bar{\mathcal{H}}_t|}\}_{t=1}^T\}_{l=1}^L$, $\boldsymbol{\mu} = \{\{\mu_{ti}\}_{i=1}^{|\bar{\mathcal{H}}_t|}\}_{t=1}^T$, we obtain the following Lagrangian,

$$Z(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{x}, \bar{\boldsymbol{y}}) = \sum_s f_s(x_s) + \boldsymbol{\lambda}^T(\bar{c} - \bar{D}\bar{y}) +$$

$$\sum_t \mu_{ti} \sum_{i=1}^{|\bar{\mathcal{H}}_t|} \left(y_{ti} - \sum_s w_{ts}x_s - \sum_{s \in \mathcal{F}_{t,\Gamma_t}(i)} b_{ts}x_s \right)$$

and the dual function is

$$Z(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \max_{\boldsymbol{x} \succeq \boldsymbol{0}, \bar{\boldsymbol{y}} \succeq \boldsymbol{0}} Z(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{x}, \bar{\boldsymbol{y}}). \tag{49}$$

The optimization over $\boldsymbol{x}$ in (49) can be decomposed into one problem for each user s:

$$\max_{x_s \geq 0} \left(f_s(x_s) - \sum_t \left(\sum_{i=1}^{|\bar{\mathcal{H}}_t|} u_{ti}w_{ts} + \sum_i \sum_{s:s\in\mathcal{F}_t(i)} u_{ti}b_{ts} \right) x_s \right). \tag{50}$$

Due to the problem reformulation in Proposition 3, link l is now associated with a group of dual variables λ_{lti}. Likewise, path t is associated with a group of dual variables μ_{ti}. Each user s determines it transmission rate x_s by considering prices from both its primary path and backup path.

The optimization over $\bar{\boldsymbol{y}}$ in (49) leads to the following relationship between dual variables,

$$\boldsymbol{\mu} = \bar{D}^T\boldsymbol{\lambda},$$

otherwise the dual function is unbounded.

The master dual problem that we want to solve is

$$\max_{\boldsymbol{\lambda} \succeq \boldsymbol{0}, \boldsymbol{\mu} \succeq \boldsymbol{0}} Z(\boldsymbol{\lambda}, \boldsymbol{\mu}) \tag{51}$$

which can be solved by the subgradient method. For each dual variable λ_{lti}, its subgradient can be calculated as

$$\zeta_{lti}(\lambda_{lti}) = c_l - \sum_s w_{ts}x_s - \sum_{s \in \mathcal{F}_t(i)} b_{ts}x_s, \tag{52}$$

and $\boldsymbol{\mu} = \bar{D}^T\boldsymbol{\lambda}$. The value of λ_{lti} will be updated using the subgradient information correspondingly. The complete algorithm is given as in Algorithm 4.

Algorithm 4. *(Suboptimal Distributed Algorithm)*

1) *Set time $k = 0$, $\boldsymbol{\lambda}(0) = \boldsymbol{0}$, and $\boldsymbol{\mu}(0) = \boldsymbol{0}$.*
2) *Let $k = k + 1$.*
3) *Each user s determines $x_s(k)$ by solving Problem (50).*

4) *Each user passes its tentative rate $x_s(k)$ to each link associated with this user.*
5) *Each link l calculates the subgradients $\boldsymbol{\zeta}_l(\boldsymbol{\lambda}_l(k)) = \{\zeta_{lti}(\lambda_{lti}(k)), \forall t, i\}$ as in (52).*
6) *If $|\boldsymbol{\zeta}(\boldsymbol{\lambda})| \leq \epsilon$, stop. Otherwise, each link l updates*

$$\boldsymbol{\lambda}_l(k+1) = \max\{\boldsymbol{\lambda}_l(k) + \theta(k)\boldsymbol{\zeta}_l(\boldsymbol{\lambda}_l(k)), 0\}.$$

7) *Each user calculates the associated dual prices $\mu_{ti}(k+1) = \sum_l d_{lt}\lambda_{lti}(k+1)$ by passing messages over the tth path from the destination to source.*

Here $\theta(k)$ is the step-size at time k and ϵ is the stopping criterion. At each iteration of the above algorithm, a link only needs to send an aggregated price variable to each user associated with this link, i.e., generating the price value $\mu_{ti}(k+1) = \sum_l d_{lt}\lambda_{lti}(k+1)$. Then each user use the collected price variable to update its source rate. Then every link will collect tentative decision rate $\boldsymbol{x}^*$ from sources and perform the subgradient projection as in (52). The number of message passing required for the first step *independent* of the parameter Γ_t, and the number of message passing for the third step increases only linearly with $|\bar{\mathcal{H}}_t|$. So the total number of message passing during one iteration for one user is $O(1 + |\bar{\mathcal{H}}_t|)$.

Building on Algorithm 4, we propose the optimal distributed algorithm in Algorithm 5 to find the optimal solution of Problem (39).

Algorithm 5. *(Optimal Distributed Algorithm)*

1) *Each path randomly generates a set $\mathcal{F}_{t,\Gamma_t}$ and let $\bar{H}_t = [\mathcal{F}_{t,\Gamma_t}]$.*
2) *Path t passes $\bar{H}_t$ to every link associated with it.*
3) *Run Algorithm 4 to obtain a tentative result $\boldsymbol{x}^*$.*
4) *The sth user passes the tentative date rate x_s^* to every path associated with this user.*
5) *Rank $\{b_{ts}x_s^*\}_{s\in\mathcal{E}_t}$ in descending order for path t and take the Γ_t biggest item to obtain a new set $\mathcal{F}_{t,\Gamma_t}$.*
6) *For the tth path, if every new generated set $\mathcal{F}_{t,\Gamma_t}$ is already contained in the corresponding set $\bar{H}_t$, then the stopping criterion stated in Theorem 2 is satisfied, stop.*
7) *Otherwise, path t passes the new generated set $\mathcal{F}_{t,\Gamma_t}$ to every link associated with this path. Every link in the tth path adds $\mathcal{F}_{t,\Gamma_t}$ into $\bar{H}_t$, and go to step 3.*

Algorithm 5 iteratively generates a group of relaxed problems to approximate the original problem (39), and eventually converges to optimal solution. Note in the worst case we may need to generate all $\mathcal{F}_{t,\Gamma_t} \in H_t$. In practice, however, column generation method typically converges to the optimal solution very fast [20].

C. Numerical Results

Here we consider a simple network model with three nodes, 13 links and 13 paths, as shown in Fig. 5. Paths $1 - 12$ are single link paths, and use links $1 - 12$, respectively. Path 13 consists of links 12 and 13. The first 11 paths are used as primary paths by 11 users in the network. Path 12 is used as the backup path by users $1 - 8$, and path 13 is used as

the backup path by user $9-11$. Each user s has a logarithmic utility function $\log(x_s)$, where unit of x_s is kbps. The capacity of each link is fixed at 1Mbps.

Figure 6(a) shows the tradeoff between robustness and performance. The performance is measured by the total network utility $\sum_s \log(x_s)$, and the robustness level is measured by the number of failures that is guaranteed to be protected on path 12, i.e., Γ_{12}. The value of Γ_{13} is fixed at 3. As we see, the performance decreases as the robustness (Γ_{13}) increases. Also the centralized algorithm and distributed optimal Algorithm 5 achieve the same performance.

Figure 6(b) shows the convergence behavior of the proposed distributed suboptimal algorithm 4. Here $\Gamma_{12} = \Gamma_{13} = 3$. It is seen the distributed method can quickly converge to the optimal solution.
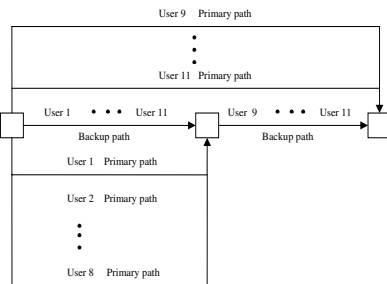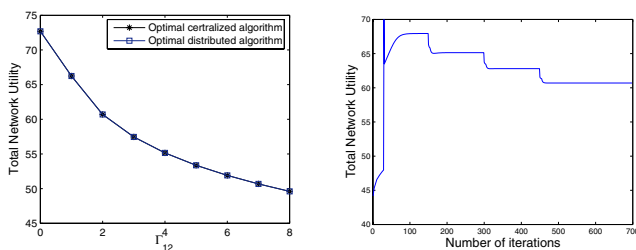

Fig. 5. Network Topology.

V. Conclusions

Making optimization models of communication network design robust is an important and under-explored area. This paper initiates the study of robust formulations that preserve a large degree of distributiveness of solution algorithms. We first describe several models for describing parameter uncertainty sets that can lead to distributed solutions for linearly constrained nominal problems. These models include general polyhedron, D-norm, and ellipsoid. We then apply these models in two representative applications. For robust power control with channel fluctuations and user entering uncertainty, several distributive algorithms are proposed under different uncertainty set modeling choices. The proposed algorithms globally and geometrically converge to the optimal solution, with provable error bounds during the transience. These algorithms can be interpreted as extended versions of the nominal Foschini-Miljanic algorithm [9]. We also characterize the tradeoff between robustness (i.e., the size of uncertainty set) and distributiveness (i.e., the amount of message passing needed) both analytically and numerically. For robust rate control under link failures, we design a fast sequential optimization algorithm based on distributed column generation method and dual decomposition. The algorithm can quickly converges to the optimal solution. The tradeoff between robustness (i.e., the maximum of link failures allowed) and performance is demonstrated through simulations.

The study of distributed robust optimization in general remains wide open, with many challenging issues and possible applications where robustness to uncertainty is as important as optimality in the nominal model.

References

[1] A. Ben-Tal and A. Nemirovski, "Robust solutions to uncertain programs," *Operations Research Letter*, vol. 25, pp. 1–13, Feb. 1999.

[2] ——, "Selected topics in robust convex optimization," *Mathematical Programming*, vol. 1, no. 1, pp. 125–158, July 2007.

[3] A. Nemirovski, "On tractable approximations of randomly perturbed convext constaints," *Proceedings of the 42nd IEEE Conference on Decision and Control*, pp. 2419–2422, 2003.

[4] L. El-Ghaoui and H. Lebret, "Robust solutions to least-square problems to uncertain data matrices," *SIAM J. Matrix Anal. Appl.*, vol. 18, pp. 1035–1064, 1997.

[5] L. El-Ghaoui, F. Oustry, and H. Lebret, "Robust solutions to uncertain semidefinite programs," *SIAM J. Optim.*, vol. 9, pp. 33–52, 1998.

[6] A. Ben-Tal and A. Nemirovski, "Robust solutions to linear programming problems contaminated with uncertain data," *Mathematical Programming*, vol. 88, pp. 411–424, 2000.

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

[8] K. Yang, Y. Wu, J. Huang, X. Wang, and M. Chiang, "Distributed Robust Optimization for Communication Neworks," (Technical Report) [Online]. Available: http://personal.ie.cuhk.edu.hk/~jwhuang/publication/DRO_TR.pdf

[9] 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.

[10] 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.

[11] J. W. Lee, M. Chiang, and R. A. Calderbank, "Utility-optimal random-access control," to appear in IEEE Trans. Wireless Comm., 2006.

[12] 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.

[13] B. Dimitris and M. Sim, "The price of robustness," *Operations Research*, vol. 52, no. 1, pp. 35–53, Feb. 2004.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[15] A. Ben-Tal and A. Nemirovski, "Robust solutions to uncertain linear problems," Research Report 6/95, Optimization Laboratory, Faculty of Industrial Engineering and Management, Technion, Tech. Rep., 1995.

[16] http://www.mosek.com, "Mosek." [Online]

[17] 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.

[18] S. Bambos, 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.

[19] D. Xu, Y. Li, M. Chiang, and A. R. Calderbank, "Optimal provision of elastic service availability," in *Proc. IEEE INFOCOM*, 2007.

[20] M. Johansson and L. Xiao, "Cross-layer optimization of wireless networks using nonlinear column generation," *Wireless Communications, IEEE Transactions on*, vol. 5, no. 2, pp. 435–445, 2006.

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

(a) Performance-robustness tradeoff under the D-norm uncertainty set.

(b) Convergence of the suboptimal algorithm

Fig. 6. Convergence of the suboptimal algorithm 4