POWER CONTROL BY GEOMETRIC PROGRAMMING

MUNG CHIANG *, CHEE WEI TAN †, DANIEL P. PALOMAR ‡, DANIEL O'NEILL, § DAVID JULIAN ¶

Abstract. In wireless cellular or ad hoc networks where qualities of service are interference-limited, a variety of power control problems can be formulated as nonlinear optimization with a system-wide objective, *e.g.*, maximizing total system throughput or achieving maxmin fairness, and many QoS constraints from individual users, *e.g.*, on data rate, delay, and outage probability. We show that in the high SIR regime, these nonlinear and apparently difficult, nonconvex optimization problems can be transformed into convex optimization problems through geometric programming, thus can be very efficiently solved for global optimality even in large networks. In the medium to low SIR regime, these constrained nonlinear optimization of power control cannot be turned into tractable convex formulations, but a heuristic can be used to compute the optimal solution by solving a series of geometric programs. While efficient and robust algorithms have been extensively studied for centralized solutions of geometric programs, distributed algorithms have not been fully explored. We present a systematic method of distributed algorithms for power control based on geometric programs. These techniques for power control, together with their implications to admission control and pricing in wireless networks, are illustrated through several numerical examples.

1. Introduction. Due to the broadcast nature of radio transmission, data rates and other qualities of service in a wireless network are affected by interference. This is particularly important in CDMA systems where users transmit at the same time over the same frequency bands and their spreading codes are not perfectly orthogonal. Transmit power control is often used to tackle this problem of signal interference. In this chapter, we study how to optimize over the transmit powers to create the optimal set of Signal-to-Interference Ratios (SIR) on wireless links. Optimality here may be referring to maximizing a system-wide efficiency metric (*e.g.*, the total system throughput), or maximizing a Quality of Service (QoS) metric for a user in the highest QoS class, or maximizing a QoS metric for the user with the minimum QoS metric value (*i.e.*, a maxmin optimization).

While the objective represents a system-wide goal to be optimized, individual users' QoS requirements also need to be satisfied. Any power allocation must therefore be constrained by a feasible set formed by these minimum requirements from the users. Such a constrained optimization captures the tradeoff between user-centric constraints and some network-centric objective. Because a higher power level from one transmitter increases the interference levels at other receivers, there may not be any feasible power allocation to satisfy the requirements from all the users. Sometimes an existing set of requirements can be satisfied, but when a new user is admitted into the system, there exists no more feasible power control solutions, or the maximized objective is reduced due to the tightening of the constraint set, leading to the need for admission control and admission pricing, respectively.

Because many QoS metrics are nonlinear functions of SIR, which is in turn a nonlinear (and neither convex nor concave) function of transmit powers, the above power control optimization or feasibility problems are difficult nonlinear optimization problems that may appear to be not efficiently solvable. This chapter shows that, when SIR is much larger than 0dB, a class of nonlinear optimization called Geometric

*ELECTRICAL ENGINEERING DEPARTMENT, PRINCETON UNIVERSITY
†ELECTRICAL ENGINEERING DEPARTMENT, PRINCETON UNIVERSITY
‡ELECTRICAL ENGINEERING DEPARTMENT, PRINCETON UNIVERSITY
§ELECTRICAL ENGINEERING DEPARTMENT, STANFORD UNIVERSITY
¶QUALCOMM

Programming (GP) can be used to efficiently compute the globally optimal power control in many of these problems, and efficiently determine the feasibility of user requirements by returning either a feasible (and indeed optimal) set of powers or a certificate of infeasibility. This leads to an effective admission control and admission pricing method. The key observation is that despite the *apparent* nonconvexity, the GP technique turns these constrained optimization of power control into nonlinear yet still convex optimization, which is intrinsically tractable despite its nonlinearity in objective and constraints. When SIR is comparable to or below 0dB, the power control problems are *truly* nonconvex with no efficient and global solution methods. In this case, we present a heuristic that empirically almost always compute the globally optimal power allocation by solving a sequence of GPs.

The GP approach reveals the hidden convexity structure, thus efficient solution methods, in power control problems with nonlinear objective functions, and clearly differentiates the tractable formulations in high-SIR regime from the intractable ones in low-SIR regime. Power control by GP is applicable to formulations in both cellular networks with single-hop transmission between mobile users and base stations and ad hoc networks with mulithop transmission among the nodes, as illustrated through several numerical examples in this chapter. Traditionally, GP is solved by centralized computation through the highly efficient interior point methods. We also outline a new result on how GP can be solved distributively with message passing.

The rest of this chapter is organized as follows. In section 2, we provide a concise introduction to GP. Section 3 is the core section of this chapter, with several subsections each discussing GP based power control with different representative formulations in cellular and multihop networks, with centralized computation and for high-SIR regime. This section's results are based on an Infocom 2002 paper [14]. Two previously unpublished extensions are then presented: solution method for low-SIR regime in section 4 and distributed algorithm in section 5.

2. Geometric Programming. GP is a class of nonlinear, nonconvex optimization with many useful theoretical and computational properties. Since a GP can be turned into a convex optimization problem [1], a local optimum is also a global optimum, duality gap is zero under mild conditions, and a global optimum can be computed very efficiently. Numerical efficiency holds both in theory and in practice: interior point methods applied to GP have provably polynomial time complexity [18], and are very fast in practice (see, *e.g.*, the algorithms and discussions on numerical efficiency in [16]) with high-quality software downloadable from the Internet (*e.g.*, the MOSEK package). Convexity and duality properties of GP are well understood, and large-scale, robust numerical solvers for GP are available. Furthermore, special structures in GP and its Lagrange dual problem lead to distributed algorithms, physical interpretations, and computational acceleration beyond the generic results for convex optimization.

GP was invented in 1960s [12] and applied to primarily mechanical and chemical engineering problems in 1960s and 1970s and then to several other science and engineering disciplines [1, 3, 4, 5]. Since mid-1990s, GP has been used to solve a variety of analysis and design problems in communication systems, including recently to wireless network power control [14, 15], resource allocation [9], joint congestion control and power control [6], queuing systems [10], information theory [8], as well as

[1] Minimizing a convex objective function subject to upper bound inequality constraints on convex constraint functions and linear equality constraints is a convex optimization problem.

to coding and signal processing. A detailed tutorial of GP and comprehensive survey of its recent applications to communication systems can be found in [7]. This section contains a brief introduction of GP terminology for applications to be shown in the next three sections.

2.1. Basic formulations. There are two equivalent forms of GP: standard form and convex form. The first is a constrained optimization of a type of function called posynomial, and the second form is obtained from the first through logarithmic change of variable. [2]

We first define a monomial as a function $f : \mathbf{R}^n_{++} \to \mathbf{R}$:

$$f(\mathbf{x}) = d x_1^{a^{(1)}} x_2^{a^{(2)}} \dots x_n^{a^{(n)}}$$

where the multiplicative constant $d \geq 0$ and the exponential constants $a^{(j)} \in \mathbf{R}, j = 1, 2, \dots, n$. A sum of monomials, indexed by k below, is called a posynomial:

$$f(\mathbf{x}) = \sum_{k=1}^{K} d_k x_1^{a_k^{(1)}} x_2^{a_k^{(2)}} \dots x_n^{a_k^{(n)}}.$$

where $d_k \geq 0,\ k = 1, 2, \dots, K$, and $a_k^{(j)} \in \mathbf{R},\ j = 1, 2, \dots, n, k = 1, 2, \dots, K$. The key features about posynomial are its positivity and convexity (in log domain).

For example, $2x_1^{-\pi} x_2^{0.5} + 3x_1 x_3^{100}$ is a posynomial in $\mathbf{x}$, $x_1 - x_2$ is not a posynomial, and x_1/x_2 is a monomial, thus also a posynomial.

Minimizing a posynomial subject to posynomial upper bound inequality constraints and monomial equality constraints is called GP in *standard form*:

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 1, \quad i = 1, 2, \dots, m, \\ & h_l(\mathbf{x}) = 1, \quad l = 1, 2, \dots, M \end{array} \tag{2.1}$$

where $f_i,\ \ i = 0, 1, \dots, m$, are posynomials: $f_i(\mathbf{x}) = \sum_{k=1}^{K_i} d_{ik} x_1^{a_{ik}^{(1)}} x_2^{a_{ik}^{(2)}} \dots x_n^{a_{ik}^{(n)}}$, and $h_l,\ \ l = 1, 2, \dots, M$ are monomials: $h_l(\mathbf{x}) = d_l x_1^{a_l^{(1)}} x_2^{a_l^{(2)}} \dots x_n^{a_l^{(n)}}$. [3]

GP in standard form is not a convex optimization problem, because posynomials are not convex functions. However, with a logarithmic change of the variables and multiplicative constants: $y_i = \log x_i, b_{ik} = \log d_{ik}, b_l = \log d_l$, and a logarithmic change of the functions' values, we can turn it into the following equivalent problem [4] in $\mathbf{y}$:

$$\begin{array}{ll} \text{minimize} & p_0(\mathbf{y}) = \log \sum_{k=1}^{K_0} \exp(\mathbf{a}_{0k}^T \mathbf{y} + b_{0k}) \\ \text{subject to} & p_i(\mathbf{y}) = \log \sum_{k=1}^{K_i} \exp(\mathbf{a}_{ik}^T \mathbf{y} + b_{ik}) \leq 0, \quad i = 1, 2, \dots, m, \\ & q_l(\mathbf{y}) = \mathbf{a}_l^T \mathbf{y} + b_l = 0, \quad l = 1, 2, \dots, M. \end{array} \tag{2.2}$$

[2]Standard form GP is often used in network resource allocation problems, and convex form GP in problems based on stochastic models such as information theoretic problems.

[3]Note that a monomial equality constraint can also be expressed as two monomial inequality constraints: $h_l(\mathbf{x}) \geq 1$ and $h_l(\mathbf{x}) \leq 1$. Thus a standard form GP can be defined as the minimization of a posynomial under upper bound inequality constraints on posynomials.

[4]Equivalence relationship between two optimization problems is used in a loose way here. If the optimized value of problem A is a simple (*e.g.*, monotonic and invertible) function of the optimized value of problem B, and an optimizer of problem B can be easily computed from an optimizer of problem A (*e.g.*, through a simple mapping), then problems A and B are said to be equivalent.

This is referred to as GP in *convex form*, which is a convex optimization problem since it can be verified that the log-sum-exp function is convex [5].

In summary, GP is a nonlinear, nonconvex optimization problem that can be transformed into a nonlinear, convex problem. Therefore, a local optimum for GP is also a global optimum, and the duality gap is zero under mild technical conditions.

2.2. Duality, feasibility, and sensitivity analysis. The Lagrange dual problem of GP has interesting structures. In particular, dual GP is linearly constrained and its objective function is a generalized entropy function [12]. Following the standard procedure of deriving the Lagrange dual problem [5], it is readily verified that for the following GP over $\mathbf{y}$ with m posynomial constraints,

$$\begin{array}{ll} \text{minimize} & \log \sum_{k=1}^{K_0} \exp(\mathbf{a}_{0k}^T \mathbf{y} + b_{0k}) \\ \text{subject to} & \log \sum_{k=1}^{K_i} \exp(\mathbf{a}_{ik}^T \mathbf{y} + b_{ik}) \leq 0, \ \ i = 1, \dots, m, \end{array}$$

the Lagrange dual problem is

$$\begin{array}{ll} \text{maximize} & \mathbf{b}_0^T \boldsymbol{\nu}_0 - \sum_{j=1}^{K_0} \nu_{0j} \log \nu_{0j} + \sum_{i=1}^{m} \left(\mathbf{b}_i^T \boldsymbol{\nu}_i - \sum_{j=1}^{K_i} \nu_{ij} \log \frac{\nu_{ij}}{\mathbf{1}^T \boldsymbol{\nu}_i} \right) \\ \text{subject to} & \boldsymbol{\nu}_i \succeq 0, \ \ i = 0, \dots, m, \\ & \mathbf{1}^T \boldsymbol{\nu}_0 = 1, \\ & \sum_{i=0}^{m} \mathbf{A}_i^T \boldsymbol{\nu}_i = 0 \end{array} \tag{2.3}$$

where the optimization variables are $(m+1)$ vectors: $\boldsymbol{\nu}_i, \ \ i = 0, 1, \dots, m$. The length of $\boldsymbol{\nu}_i$ is K_i, *i.e.*, the number of monomial terms in the ith posynomial, $i = 0, 1, \dots, m$. Here, $\mathbf{A}_0$ is the matrix of the exponential constants in the objective function, where each row corresponds to each monomial term (*i.e.*, $\mathbf{a}_{0k}^T$ is the kth row in matrix $\mathbf{A}_0$), and $\mathbf{A}_i, \ \ i = 1, 2, \dots, m$, are the matrices of the exponential constants in the constraint functions, again with each row corresponding to each monomial term. The multiplicative constants in the objective function are denoted as $\mathbf{b}_0$ and those in the ith constraint as $\mathbf{b}_i, \ i = 1, 2, \dots, m$. *Linearity* of dual problem constraints is utilized in some very efficient GP solvers (*e.g.*, [16]) that solve both the primal and dual problems of a GP simultaneously.

Testing whether there is any variable $\mathbf{x}$ that satisfies a set of posynomial inequality and monomial equality constraints:

$$f_i(\mathbf{x}) \leq 1, \ \ i = 1, \dots, m, \ \ h_l(\mathbf{x}) = 1, \ \ l = 1, \dots, M \tag{2.4}$$

is called a GP *feasibility problem*. Solving feasibility problem is useful when we would like to determine whether the constraints are too tight to allow any feasible solution, or when it is necessary to generate a feasible solution as the initial point of a interior-point algorithm.

Feasibility of the monomial equality constraints can be verified by checking feasibility of the linear system of equations that the monomial constraints get logarithmically transformed into. Feasibility of the posynomial inequality constraints can then be verified by solving the following GP, introducing an auxiliary variable $s \in \mathbf{R}$ in addition to variables $\mathbf{x} \in \mathbf{R}^n$ [12, 4]:

$$\begin{array}{ll} \text{minimize} & s \\ \text{subject to} & f_i(\mathbf{x}) \leq s, \ \ i = 1, \dots, m \\ & g_l(\mathbf{x}) = 1, \ \ l = 1, \dots, M, \\ & s \geq 1. \end{array} \tag{2.5}$$

This GP always has a feasible solution: $s = \max\{1, \max_i\{f_i(\mathbf{x})\}\}$ for any $\mathbf{x}$ that satisfies the monomial equality constraints. Now solve problem (2.5) and obtain the optimal $(s^*, \mathbf{x}^*)$. If $s^* = 1$, then the set of posynomial constraints $f_i(\mathbf{x}) \leq 1$ are feasible, and the associated $\mathbf{x}^*$ is a feasible solution to the original feasibility problem (2.4). Otherwise, the set of posynomial constraints is infeasible.

The constant parameters in a GP may be based on inaccurate estimates or vary over time. As constant parameters change a little, we may not want to solve the slightly perturbed GP from scratch. It is useful to directly determine the impact of small perturbations of constant parameters on the optimal solution. Suppose we loosen the ith inequality constraint (with $u_i > 0$) or tighten it (with $u_i < 0$), and shift the jth equality constraint (with $v_j \in \mathbf{R}$) in a standard form GP:

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq e^{u_i}, \;\; i = 1, \ldots, m, \\ & g_j(\mathbf{x}) = e^{v_j}, \;\; j = 1, \ldots, M. \end{array} \tag{2.6}$$

Consider the optimal value of a GP p^* as a function of the perturbations $\mathbf{u}, \mathbf{v}$. The sensitivities of a GP with respect to the ith inequality constraint and lth equality constraint are defined as:

$$S_i = \frac{\partial \log p^*(0,0)}{\partial u_i} = \frac{\partial p^*(0,0)/\partial u_i}{p^*(0,0)},$$
$$T_j = \frac{\partial \log p^*(0,0)}{\partial v_j} = \frac{\partial p^*(0,0)/\partial v_j}{p^*(0,0)}.$$

A large sensitivity S_i with respect to an inequality constraint means that if the constraint is tightened (or loosened), the optimal value of GP increases (or decreases) considerably. Sensitivity can be obtained from the corresponding Lagrange dual variables of (2.6): $S_i = -\lambda_i$ and $T_j = -\nu_j$ where $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ are the Lagrange multipliers of the inequality and equality constraints in the convex form of (2.6), respectively.

3. Power Control by Geometric Programming: High SIR Case. GP in standard form can be used to formulate network resource allocation problems with nonlinear objectives under nonlinear QoS constraints. The basic idea is that resources are often allocated proportional to some parameters, and when resource allocations are optimized over these parameters, we are maximizing an inverted posynomial subject to lower bounds on other inverted posynomials, which are equivalent to GP in standard form. This section presents how GP can be used to efficiently solve QoS constrained power control problems with nonlinear objectives, based on results in [14, 7].

Various schemes for power control, centralized or distributed, based on different transmission models and application needs, have been extensively studied since 1990s, *e.g.*, in [2, 13, 17, 20, 21, 22] and many other publications. This chapter summarizes the approach of formulating power control problems through GP (and an extension of GP called Signomial Programming). The key advantage is that globally optimal power allocations can be efficiently computed for a variety of *nonlinear* system-wide objectives and user QoS constraints, even when these nonlinear problems *appear* to be nonconvex optimization.

3.1. Basic model. Consider a wireless (cellular or multihop) network with n logical transmitter/receiver pairs. Transmit powers are denoted as $P_1, \ldots, P_n$. In the cellular uplink case, all logical receivers may reside in the same physical receiver, *i.e.*, the base station. In the multihop case, since the transmission environment can be different on the links comprising an end-to-end path, power control schemes must consider each link along a flow's path.

Under Rayleigh fading, the power received from transmitter j at receiver i is given by $G_{ij}F_{ij}P_j$ where $G_{ij} \geq 0$ represents the path gain and is often modeled as proportional to $d_{ij}^{-\gamma}$ where d_{ij} is distance and γ is the power fall-off factor. We also let G_{ij} encompass antenna gain and coding gain. The numbers F_{ij} model Rayleigh fading and are independent and exponentially distributed with unit mean. The distribution of the received power from transmitter j at receiver i is then exponential with mean value $\mathbf{E}\left[G_{ij}F_{ij}P_j\right] = G_{ij}P_j$. The distribution of the received power from transmitter j at receiver i is exponential with mean value $\mathbf{E}\left[G_{ij}F_{ij}P_j\right] = G_{ij}P_j$. The SIR for the receiver on logical link i is:

$$\text{SIR}_i = \frac{P_i G_{ii} F_{ii}}{\sum_{j \neq i}^{N} P_j G_{ij} F_{ij} + n_i} \tag{3.1}$$

where n_i is the noise for receiver i.

The constellation size M used by a link can be closely approximated for MQAM modulations as follows: $M = 1 + \frac{-\phi_1}{\ln(\phi_2 \text{BER})}\text{SIR}$ where BER is the bit error rate and ϕ_1, ϕ_2 are constants that depend on the modulation type. Defining $K = \frac{-\phi_1}{\ln(\phi_2 \text{BER})}$ leads to an expression of the data rate R_i on the ith link as a function of SIR: $R_i = \frac{1}{T}\log_2(1 + K\text{SIR}_i)$, which will be approximated as

$$R_i = \frac{1}{T}\log_2(K\text{SIR}_i) \tag{3.2}$$

when KSIR is much larger than 1. This approximation is reasonable either when the signal level is much higher than the interference level or, in CDMA systems, when the spreading gain is large. For notational simplicity in the rest of this chapter, we redefine G_{ii} as K times the original G_{ii}, thus absorbing constant K into the definition of SIR.

The aggregate data rate for the system can then be written as the sum

$$R_{system} = \sum_i R_i = \frac{1}{T}\log_2\left[\prod_i \text{SIR}_i\right].$$

So in the high SIR regime, aggregate data rate maximization is equivalent to maximizing a product of SIR. The system throughput is the aggregate data rate supportable by the system given a set of users with specified QoS requirements.

Outage probability is another important QoS parameter for reliable communication in wireless networks. A channel outage is declared and packets lost when the received SIR falls below a given threshold SIR_{th}, often computed from the BER requirement. Most systems are interference dominated and the thermal noise is relatively small, thus the ith link outage probability is

$$\begin{aligned} P_{o,i} &= \textbf{Prob}\{\text{SIR}_i \leq \text{SIR}_{th}\} \\ &= \textbf{Prob}\{G_{ii}F_{ii}P_i \leq \text{SIR}_{th}\sum_{j\neq i} G_{ij}F_{ij}P_j\}. \end{aligned}$$

The outage probability can be expressed as $P_{o,i} = 1 - \prod_{j\neq i} \frac{1}{1+\frac{\text{SIR}_{th}G_{ij}P_j}{G_{ii}P_i}}$ [15], which means that an upper bound on $P_{o,i} \leq P_{o,i,max}$ can be written as an upper bound on a posynomial in $\mathbf{P}$:

$$\prod_{j \neq i} \left(1 + \frac{\mathrm{SIR}_{th} G_{ij} P_j}{G_{ii} P_i} \right) \leq \frac{1}{1 - P_{o,i,max}}. \tag{3.3}$$

3.2. Cellular wireless networks. We first present how GP-based power control applies to cellular wireless networks with one-hop transmission from N users to a base station, extending the scope of power control by the classical solution in CDMA systems that equalizes SIRs, and those by the iterative algorithms (*e.g.*, in [2, 13, 17]) that minimize total power (a linear objective function) subject to SIR constraints.

We start the discussion on the suite of power control problem formulations with a simple objective function and simple constraints.

PROPOSITION 1. *The following constrained problem of maximizing the SIR of a particular user i^* is a GP:*

$$\begin{array}{ll} \textit{maximize} & SIR_{i^*}(\mathbf{P}) \\ \textit{subject to} & SIR_i(\mathbf{P}) \geq SIR_{i,min}, \ \ \forall i, \\ & P_{i1} G_{i1} = P_{i2} G_{i2}, \\ & 0 \leq P_i \leq P_{i,max}, \ \ \forall i. \end{array}$$

The first constraint, equivalent to $R_i \geq R_{i,min}$, sets a floor on the SIR of other users and protects these users from user i^* increasing her transmit power excessively. The second constraint reflects the classical power control criterion in solving the near-far problem in CDMA systems: the expected received power from one transmitter $i1$ must equal that from another $i2$. The third constraint is regulatory or system limitations on transmit powers. All constraints are verified to be inequality upper bounds on posynomials.

Alternatively, we can use GP to maximize the minimum SIR among all users. The maxmin fairness objective:

$$\mathrm{maximize}_{\mathbf{P}} \ \ \min_k \{\mathrm{SIR}_k\}$$

can be accommodated in GP-based power control because it can be turned into equivalently maximizing an auxiliary variable t such that $\mathrm{SIR}_k \geq t$, $\forall k$, which has posynomial objective and constraints in $(\mathbf{P}, t)$.

Example 1. A simple system comprised of five users is used for a numerical example. The five users are spaced at distances d of $1, 5, 10, 15$, and 20 units from the base station. The power fall-off factor $\gamma = 4$. Each user has a maximum power constraint of $P_{max} = 0.5mW$. The noise power is 0.5μW for all users. The SIR of all users, other than the user we are optimizing for, must be greater than a common threshold SIR level β. In different experiments, β is varied to observe the effect on the optimized user's SIR. This is done independently for the near user at $d = 1$, a medium distance user at $d = 15$, and the far user at $d = 20$. The results are plotted in Figure 3.1.

Several interesting effects are illustrated. First, when the required threshold SIR in the constraints is sufficiently high, there are no feasible power control solutions. At moderate threshold SIR, as β is decreased, the optimized SIR initially increases rapidly. This is because it is allowed to increase its own power by the sum of the power reductions in the four other users, and the noise is relatively insignificant. At low threshold SIR, the noise becomes more significant and the power trade-off from

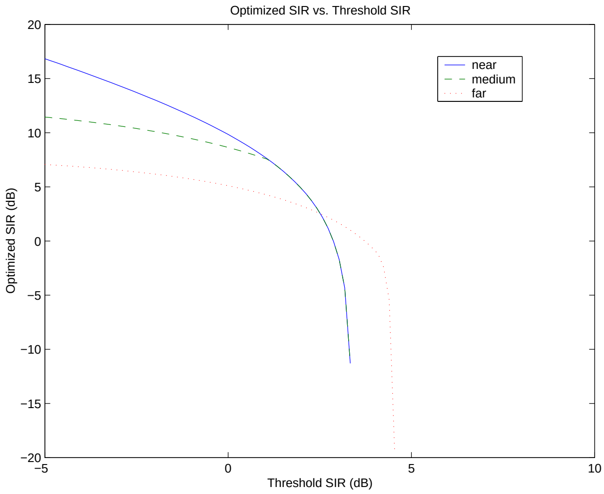

FIG. 3.1. *Constrained optimization of power control in a cellular network (Example 1).*

the other users less significant, so the curve starts to bend over. Eventually, the optimized user reaches its upper bound on power and cannot utilize the excess power allowed by the lower threshold SIR for other users. Therefore, during this stage, the only gain in the optimized SIR is the lower power transmitted by the other users. This is exhibited by the transition from a sharp bend in the curve to a much shallower sloped curve. We also note that the most distant user in the constraint set dictates feasibility.

We now proceed to show that GP can also be applied to the problem formulations with an overall system objective of total system throughput, under both user data rate constraints and outage probability constraints.

PROPOSITION 2. *The following constrained problem of maximizing system throughput is a GP:*

$$\begin{array}{ll} \textit{maximize} & R_{system}(\mathbf{P}) \\ \textit{subject to} & R_i(\mathbf{P}) \geq R_{i,min}, \quad \forall i, \\ & P_{o,i}(\mathbf{P}) \leq P_{o,i,max}, \quad \forall i, \\ & 0 \leq P_i \leq P_{i,max}, \quad \forall i \end{array} \tag{3.4}$$

where the optimization variables are the transmit powers $\mathbf{P}$.

The objective is equivalent to minimizing the posynomial $\prod_i \mathrm{ISR}_i$, where ISR is $\frac{1}{\mathrm{SIR}}$. Each ISR is a posynomial in $\mathbf{P}$ and the product of posynomials is again a posynomial. The first constraint is from the data rate demand $R_{i,min}$ by each user. The second constraint represents the outage probability upper bounds $P_{o,i,max}$. These inequality constraints put upper bounds on posynomials of $\mathbf{P}$, as can be readily verified through (3.2,3.3). Thus (3.4) is indeed a GP, and efficiently solvable for global optimality.

There are several obvious variations of problem (3.4) that can be solved by GP, *e.g.*, we can lower bound R_{system} as a constraint and maximize R_{i^*} for a particular user i^*, or have a total power $\sum_i P_i$ constraint or objective function.

The objective function to be maximized can also be generalized to a weighted sum of data rates: $\sum_i w_i R_i$ where $\mathbf{w} \succeq 0$ is a given weight vector. This is still a GP because maximizing $\sum_i w_i \log \mathrm{SIR}_i$ is equivalent to maximizing $\log \prod_i \mathrm{SIR}_i^{w_i}$, which is in turn equivalent to minimizing $\prod_i \mathrm{ISR}_i^{w_i}$. Now use auxiliary variables t_i, and minimize $\prod_i t_i^{w_i}$ over the original constraints in (3.4) plus the additional constraints $\mathrm{ISR}_i \le t_i$ for all i. This is readily verified to be a GP, and is equivalent to the original problem.

In addition to efficient computation of the globally optimal power allocation with nonlinear objectives and constraints, GP can also be used for admission control based on feasibility study described in subsection 2.2, and for determining which QoS constraint is a performance bottleneck, *i.e.*, meet tightly at the optimal power allocation, based on sensitivity analysis in subsection 2.2.

3.3. Extension: Multihop wireless networks. In wireless multihop networks, system throughput may be measured either by end-to-end transport layer utilities or by link layer aggregate throughput. GP application to the first approach has appeared in [6], and we focus on the second approach in this section by extending problem formulations, such as (3.4), to multihop case as in the following example.

Example 2. Consider a simple four node multihop network shown in Figure 3.2. There are two connections $A \to B \to D$ and $A \to C \to D$. Nodes A and D, as well as B and C, are separated by a distance of 20m. Path gain between a transmitter and a receiver is the distance to the power -4. Each link has a maximum transmit power of 1mW. All nodes use MQAM modulation. The minimum data rate for each connection is 100bps, and the target BER is 10^{-3}. Assuming Rayleigh fading, we require outage probability be smaller than 0.1 on all links for an SIR threshold of 10dB. Spreading gain is 200. Using GP formulation (3.4), we find the maximized system throughput $R^* = 216.8$kbps, $R_i^* = 54.2$kbps for each link, $P_1^* = P_3^* = 0.709$mW and $P_2^* = P_4^* = 1$mW. The resulting optimized SIR is 21.7dB on each link.

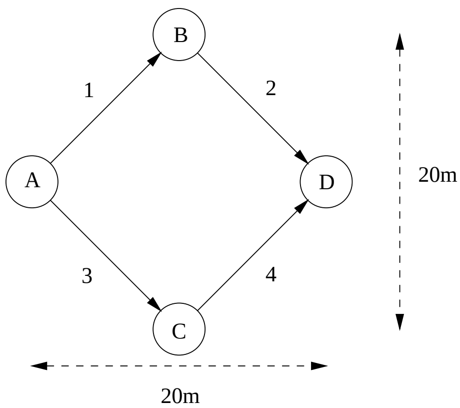

FIG. 3.2. *A small wireless multihop network (Example 2).*

For this topology, we also consider an illustrative example of admission control and pricing. Three new users U_1, U_2, and U_3 are going to arrive to the network in order. U_1 and U_2 require 30kbps sent along the upper path $A \to B \to D$, while U_3 requires 10kbps sent from $A \to B$. All three users require the outage probability to be less than 0.1. When U_1 arrives at the system, her price is the baseline price. Next, U_2 arrives, and her QoS demands decrease the maximum system throughput from 216.82kbps to 116.63kbps, so her price is the baseline price plus an amount proportional to the reduction in system throughput. Finally, U_3 arrives, and her QoS demands produce no feasible power allocation solution, so she is not admitted to the system.

3.4. Extension: Queuing models. We now turn to delay and buffer overflow properties to be included in constraints or objective function of power control optimization. The average delay a packet experiences traversing a network is an important design consideration in some applications. Queuing delay is often the primary source of delay, particularly for bursty data traffic in multihop networks. A node i first buffers the received packets in a queue and then transmits these packets at a rate R set by the SIR on the egress link, which is in turn determined by the transmit powers $\mathbf{P}$. A FIFO queuing discipline is used here for simplicity. Routing is assumed to be fixed, and is feed-forward with all packets visiting a node at most once.

Packet traffic entering the multihop network at the transmitter of link i is assumed to be Poisson with parameter λ_i and to have an exponentially distributed length with parameter Γ. Using the model of an $M/M/1$ queue, the probability of transmitter i having a backlog of $N_i = k$ packets to transmit is well known to be $\mathbf{Prob}\{N_i = k\} = (1-\rho)\rho^k$ where $\rho = \frac{\lambda_i}{\Gamma R_i(\mathbf{P})}$, and the expected delay is $\frac{1}{\Gamma R_i(\mathbf{P})-\lambda_i}$. Under the feed-forward routing and Poisson input assumptions, Burke's theorem can be applied. Thus the total packet arrival rate at node i is $\Lambda_i = \sum_{j\in I(i)} \lambda_j$ where $I(i)$ is the set of connections traversing node i. The expected delay $\bar{D}_i$ can be written as

$$\bar{D}_i = \frac{1}{\Gamma R_i(\mathbf{P}) - \Lambda_i}. \tag{3.5}$$

A bound $\bar{D}_{i,max}$ on $\bar{D}_i$ can thus be written as $\frac{1}{\frac{\Gamma}{T}\log_2(\mathrm{SIR}_i)-\Lambda_i} \leq \bar{D}_{i,max}$, or equivalently, $\mathrm{ISR}_i(\mathbf{P}) \leq 2^{-\frac{T}{\Gamma}(\bar{D}_{max}^{-1}+\Lambda_i)}$, which is an upper bound on a posynomial ISR of $\mathbf{P}$.

The probability P_{BO} of dropping a packet due to buffer overflow at a node is also important in several applications. It is again a function of $\mathbf{P}$ and can be written as $P_{BO,i} = \mathbf{Prob}\{N_i > B\} = \rho^{B+1}$ where B is the buffer size and $\rho = \frac{\Lambda_i}{\Gamma R_i(\mathbf{P})}$. Setting an upper bound $P_{BO,i,max}$ on the buffer overflow probability also gives a posynomial lower bound constraint in $\mathbf{P}$: $\mathrm{ISR}_i(\mathbf{P}) \leq 2^{-\Psi}$ where $\Psi = \frac{T\Lambda_i}{\Gamma(P_{BO,i,max})^{\frac{1}{B+1}}}$.

PROPOSITION 3. *The following nonlinear problem of optimizing powers to maximize system throughput, subject to constraints on outage probability, expected delay, and the probability of buffer overflow, is a GP:*

$$\begin{array}{ll} \textit{maximize} & R_{system}(\mathbf{P}) \\ \textit{subject to} & \bar{D}_i(\mathbf{P}) \leq \bar{D}_{i,max}, \;\; \forall i, \\ & P_{BO,i}(\mathbf{P}) \leq P_{BO,i,max}, \;\; \forall i, \\ & \textit{Same constraints as in problem (3.4)} \end{array} \tag{3.6}$$

where the optimization variables are the transmit powers $\mathbf{P}$.

Example 3. Consider a numerical example of the optimal tradeoff between maximizing the system throughput and bounding the expected delay for the network shown in Figure 3.3. There are six nodes, eight links, and five multihop connections. All sources are Poisson with intensity $\lambda_i = 200$ packets per second, and exponentially distributed packet lengths with an expectation of of 100 bits. The nodes use CDMA transmission scheme with a symbol rate of 10k symbols per second and the spreading gain is 200. Transmit powers are limited to 1mW and the target BER is 10^{-3}. The path loss matrix is calculated based on a power falloff of d^{-4} with the distance d, and a separation of 10m between any adjacent nodes along the perimeter of the network.

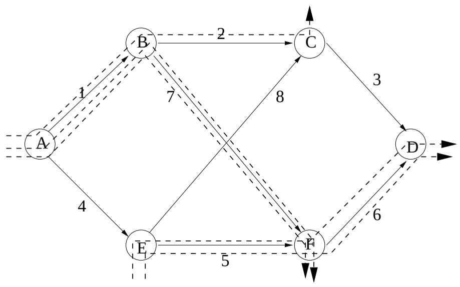

FIG. 3.3. *Topology and flows in a multihop wireless network (Example 3).*

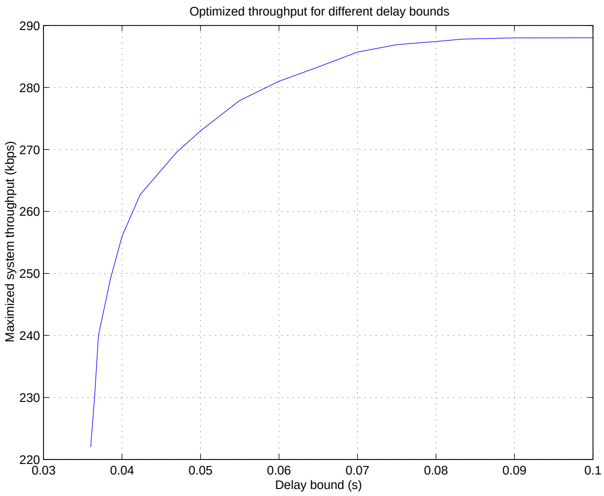

FIG. 3.4. *Optimal tradeoff between maximized system throughput and average delay constraint (Example 3).*

Figure 3.4 shows the maximized system throughput for different upper bound numerical values in the expected delay constraints, obtained by solving a sequence of GPs, one for each point on the curve. There is no feasible power allocation to achieve delay smaller than 0.036s. As the delay bound is relaxed, the maximized system throughput increases sharply first, then more slowly until the delay constraints are no longer active. GP efficiently returns the globally optimal tradeoff between system throughput and queuing delay.

4. Power Control by Geometric Programming: Medium to Low SIR Case. There are two main limitations in the GP-based power control methods discussed so far: high-SIR assumption and centralized computation. Both can be overcome as discussed in this and the next section.

The first limitation is the assumption that SIR is much larger than 0dB, which can be removed by the condensation method for Signomial Programming. When SIR is not much larger than 0dB, the approximation of $\log(1+\mathrm{SIR})$ as $\log \mathrm{SIR}$ does not hold. Unlike SIR, which is an inverted posynomial, $1 + \mathrm{SIR}$ is not an inverted posynomial.

Instead, $\frac{1}{1+\text{SIR}}$ is a ratio between two posynomials:

$$\frac{\sum_{j \neq i} G_{ij} P_j + n_i}{\sum_j G_{ij} P_j + n_i}.$$

Minimizing or upper bounding a ratio between two posynomials is a truly nonconvex problem that is intrinsically intractable.

4.1. Signomial programming. To overcome this difficulty, we need a generalization of GP called Signomial Programming (SP): minimizing a signomial subject to upper bound inequality constraints on signomials, where a signomial $s(\mathbf{x})$ is a sum of monomials, possibly with *negative* multiplicative coefficients:

$$s(\mathbf{x}) = \sum_{i=1}^{N} c_i g_i(\mathbf{x})$$

where $\mathbf{c} \in \mathbf{R}^N$ and $g_i(\mathbf{x})$ are monomials.

We first convert a SP into a Complementary GP, which allows upper bound constraints on the ratio between two posynomials, and then apply a monomial approximation iteratively. This is known in operations research community as the condensation method [3, 11], which is an instance of the cutting-plane method for nonlinear programming.

The conversion between a SP and a Complementary GP is trivial. An inequality in SP of the following form

$$f_{i1}(\mathbf{x}) - f_{i2}(\mathbf{x}) \leq 1,$$

where f_{i1}, f_{i2} are posynomials, is clearly equivalent to

$$\frac{f_{i1}(\mathbf{x})}{1 + f_{i2}(\mathbf{x})} \leq 1. \tag{4.1}$$

Now we have two choices to make the monomial approximation. One is to approximate the denominator $1+f_{i2}(\mathbf{x})$ with a monomial but leave the numerator $f_{i1}(\mathbf{x})$ as a posynomial. This is called the (single) condensation method, and results in a GP approximation of a SP. An iterative procedure can then be carried out: given a feasible $\mathbf{x}^k$, from which a monomial approximation using $\mathbf{x}^k$ can be made and a GP formed, from which an optimizer can be computed and used as $\mathbf{x}^{k+1}$: the starting point for the next iteration. This sequence of computation of $\mathbf{x}$ may converge to $\mathbf{x}^*$, a global optimizer of the original SP.

There are many ways to make a monomial approximation of a posynomial. One possibility is based on the following simple inequality: arithmetic mean is greater than or equal to geometric mean, *i.e.*,

$$\sum_i \alpha_i v_i \geq \prod_i v_i^{\alpha_i},$$

where $\mathbf{v} \succ 0$ and $\boldsymbol{\alpha} \succeq 0$, $\mathbf{1}^T\boldsymbol{\alpha} = 1$. Letting $u_i = \alpha_i v_i$, we can write this basic inequality as

$$\sum_i u_i \geq \prod_i \left(\frac{u_i}{\alpha_i}\right)^{\alpha_i}.$$

Let $\{u_i(\mathbf{x})\}$ be the monomial terms in a posynomial $f(\mathbf{x}) = \sum_i u_i(\mathbf{x})$. For a given $\boldsymbol{\alpha} \succeq 0$ such that $\mathbf{1}^T\boldsymbol{\alpha} = 1$, a lower bound inequality on posynomial $f(\mathbf{x})$ can now be approximated by an upper bound inequality on the following monomial:

$$\prod_i \left(\frac{u_i(\mathbf{x})}{\alpha_i} \right)^{-\alpha_i}. \tag{4.2}$$

This approximation is in the conservative direction because the original constraint is now tightened. There are many choices of $\boldsymbol{\alpha}$. One possibility is to let

$$\alpha_i(\mathbf{x}) = u_i(\mathbf{x})/f(\mathbf{x}), \;\; \forall i,$$

for any fixed positive $\mathbf{x}$, which obviously satisfies the condition that $\boldsymbol{\alpha} \succeq 0$ and $\mathbf{1}^T\boldsymbol{\alpha} = 1$. Given an $\boldsymbol{\alpha}$ for each of the upper bound inequalities on ratio between two posynomials (4.1), a standard form GP can be obtained based on the above geometric mean approximation.

Another choice is to make the monomial approximation for both the denominator $1 + f_{i2}(\mathbf{x})$ and numerator posynomials $f_{i2}(\mathbf{x})$ in (4.1). That turns all the constraints into monomials, and after a log transformation, approximate SP as a linear program (LP). This is called the double condensation method, and a similar iterative procedure can be carried out as in the single condensation method. The main difference from the single condensation method is that this LP approximation always generate solutions that are infeasible in the original SP. At the kth step of the iteration, the most violated constraint is condensed at $\mathbf{x}^k$, *i.e.*, the monomial approximation is applied to this constraint inequality using $\boldsymbol{\alpha}(\mathbf{x}^k)$. The resulting new constraint is added to the LP approximation for the $(k+1)$th step of the iteration. Iterations in double condensation stop at the solution $\mathbf{x}^*$ where all constraints in the original SP are satisfied.

4.2. Applications to power control. Figure 4.2 shows the approach of GP-based power control for general SIR regime. In the high SIR regime, we solve only *one* GP. In the medium to low SIR regimes, we solve truly nonconvex power control problems that cannot be turned into convex formulation through a *series* of GPs.

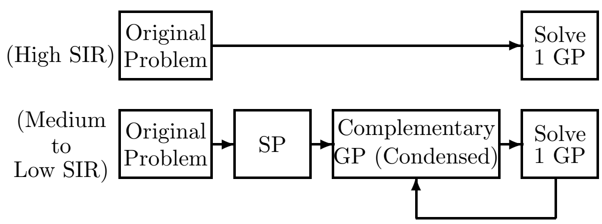

FIG. 4.1. *GP-based power control for general SIR regime.*

GP-based power control problems in the medium to low SIR regimes become Signomial Programs, which can be solved by the single or double condensation method. We focus on the single condensation method here. Consider a representative problem formulation of maximizing total system throughput in a cellular wireless network subject to user rate and outage probability constraints from proposition 2 which is explicitly written out as:

$$
\begin{array}{lll}
 & \text{minimize} & \prod_{i=1}^{N} \frac{1}{1+\mathrm{SIR}_i} \\
(4.3) & \text{subject to} & (2^{TR_{i,min}} - 1)\frac{1}{\mathrm{SIR}_i} \leq 1, \;\; i = 1, \ldots, N, \\
 & & (\mathrm{SIR}_{th})^{N-1}(1 - P_{o,i,max}) \prod_{j\neq i}^{N} \frac{G_{ij}P_j}{G_{ii}P_i} \leq 1, \;\; i = 1, \ldots, N, \\
 & & P_i(P_{i,max})^{-1} \leq 1, \;\; i = 1, \ldots, N.
\end{array}
$$

All the constraints are posynomials. However, the objective is not a posynomial, but a ratio between two posynomials. This power control problem is a SP (equivalently, a Complementary GP), and can be solved by the condensation method by solving a series of GPs. Specifically, we have the following algorithm:

STEP 0: Choose an initial power vector: a feasible $\mathbf{P}$.

STEP 1: Evaluate the denominator posynomial of the (4.3) objective function with the given $\mathbf{P}$.

STEP 2: Compute for each term i in this posynomial,

$$
\alpha_i = \frac{\text{value of } i\text{th term in posynomial}}{\text{value of posynomial}}.
$$

STEP 3: Condense the denominator posynomial of the (4.3) objective function into a monomial using (4.2) with weights α_i.

STEP 4: Solve the resulting GP using an interior point method.

STEP 5: Go to STEP 1 using $\mathbf{P}$ of STEP 4.

STEP 6: Terminate the kth loop if $\parallel \mathbf{P}^{(k)} - \mathbf{P}^{(k-1)} \parallel \leq \epsilon$ where ϵ is the error tolerance for exit condition.

As condensing the objective in the above problem gives us an underestimate of the objective value, each GP in the condensation iteration loop tries to improve the accuracy of the approximation to a particular minimum in the original feasible region.

Example 4. We consider a wireless cellular network with 3 users. Let $T = 10^{-6}$s, $G_{ii} = 1.5$, and generate $G_{ij}, i \neq j$, as independent random variables uniformly distributed between 0 and 0.3. Threshold SIR is $\mathrm{SIR}_{th} = -10$dB, and minimal data rate requirements are 100 kbps, 600 kbps and 1000 kbps for logical links 1, 2 and 3 respectively. Maximal outage probabilities are 0.01 for all links, and maximal transmit powers are 3mW, 4mW and 5mW for link 1, 2 and 3 respectively.

For each instance of SP power control (4.3), we pick a random initial feasible power vector $\mathbf{P}$ uniformly between 0 and $\mathbf{P}_{max}$. Figure 4.2 compares the maximized total network throughput achieved over five hundred sets of experiments with different initial vectors. With the (single) condensation method, SP converges to different optima over the entire set of experiments, achieving (or coming very close to) the global optimum at 5290 bps (96% of the time) and a local optimum at 5060 bps (4% of the time), thus is very likely to converge to or very close to the global optimum. The average number of GP iterations required by the condensation method over the same set of experiments is 15 if an extremely tight exit condition is picked for SP condensation iteration: $\epsilon = 1 \times 10^{-10}$. This average can be substantially reduced by using a larger ϵ, *e.g.*, increasing ϵ to 1×10^{-2} requires on average only 4 GPs.

Other constant parameter values are tried, and a particular instance of these further trials is shown in Table 4.1, where the maximized total system throughput and the corresponding power vector are shown for exhaustive search (which always generates the globally optimal results) and the condensation method (which solves the SP formulation through a series of GPs).

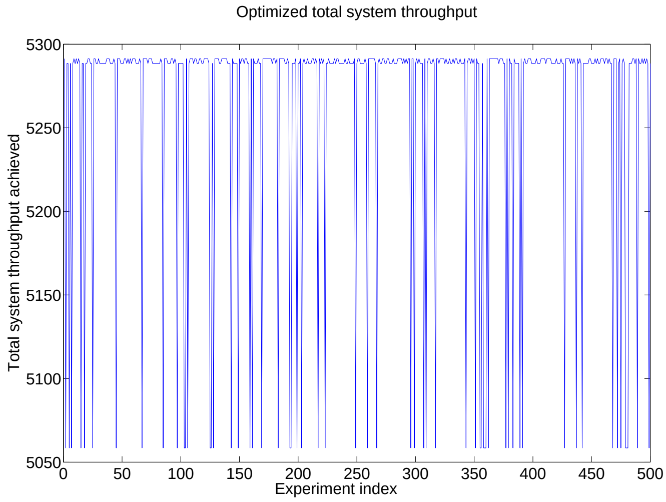

FIG. 4.2. *Maximized total system throughput achieved by (single) condensation method for 500 different initial feasible vectors (Example 4). Each point represents a different experiment with a different initial power vector.*

We have thus far discussed a power control problem (4.3) where the objective function needs to be condensed. The method is also applicable if some constraint functions are signomials and need to be condensed. For example, consider the case of differentiated services where a user expects to obtain a predicted QoS relatively better than the other users. We may have a proportional delay differentiation model where a user who pays more tariff obtains a delay proportionally lower as compared to users who pay less. Then for a particular ratio between any user i and j, σ_{ij}, we have

$$\frac{\overline{D}_i}{\overline{D}_j} = \sigma_{ij}, \tag{4.4}$$

which, by (3.5), is equivalent to

$$\frac{1+\mathrm{SIR}_j}{(1+\mathrm{SIR}_i)^{\sigma_{ij}}} = 2^{(\lambda_j - \sigma_{ij}\lambda_i)T/\Gamma}. \tag{4.5}$$

The denominator on the left hand side is a ratio between posynomials raised to a positive power. Therefore, the double condensation method can be readily used to solve the proportional delay differentiation problem because the function on the left hand side can be condensed to a monomial, and a monomial equality constraint is allowed in GP.

Example 5A. We consider the wireless cellular network in Example 4 with an additional constraint $\frac{\overline{D}_1}{\overline{D}_3} = 1$. The arrival rates of each user at base station is measured and input as network parameters into (4.5). Figures 4.3 and 4.4 show the convergence towards satisfying all the QoS constraints including the DiffServ constraint. As shown on the figures, the convergence is extremely fast with the power allocations very close to the optimal power allocation by the 8th GP iteration.

Example 5B. A slightly larger cellular system with 6 users is then studied. The number of users is still small enough for exhaustive search to be conducted and

Method	System throughput	P_1^*	P_2^*	P_3^*
Exhaustive search	6626 kbps	0.65	0.77	0.79
Single condensation	6626 kbps by solving 17 GPs	0.65	0.77	0.78
Added DiffServ constraint	6624 kbps by solving 17 GPs	0.68	0.75	0.68

TABLE 4.1

Examples 4 and 5A. Row 2 shows Example 4 using the single condensation on system throughput maximization, and Row 1 shows the optimal solutions found by exhaustive search. Row 3 shows Example 5A with an additional DiffServ constraint $D_1/D_3 = 1$, which also recovers the new optimal solution of 6624 *kbps as verified by exhaustive search.*

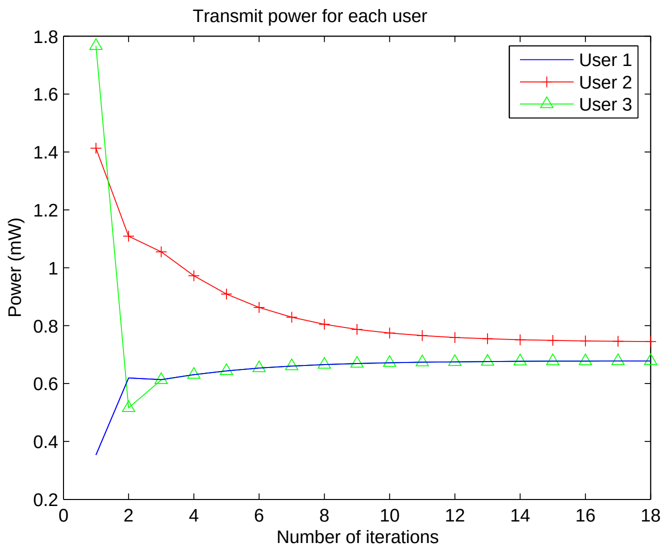

FIG. 4.3. *Convergence of power variables (Example 5A).*

to establish the benchmark of globally optimal power control. Performance of SP condensation method for 300 different initial feasible power allocations is shown in Figure 4.5. With an extremely few number of exceptions, SP condensation returns the globally optimal power allocation. By using a more relaxed but sufficiently accurate exit condition $\epsilon = 10^{-3}$, the average number of GP iterations needed is reduced to 11 even though the problem size doubles compared to Example 5A.

The optimum of power control produced by the condensation method may be a local one. The following new heuristic of solving a series of SPs (each solved through a series of GPs) can be further applied to help find the global optimum. After the original SP (4.3) is solved, a slightly modified SP is formulated and solved:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \prod_{i=1}^{N} \frac{1}{1+\mathrm{SIR}_i} \le t, \\ & t \le \frac{t_0}{\alpha}, \\ & \textit{Same set of constraints as problem (4.3)}, \end{array} \tag{4.6}$$

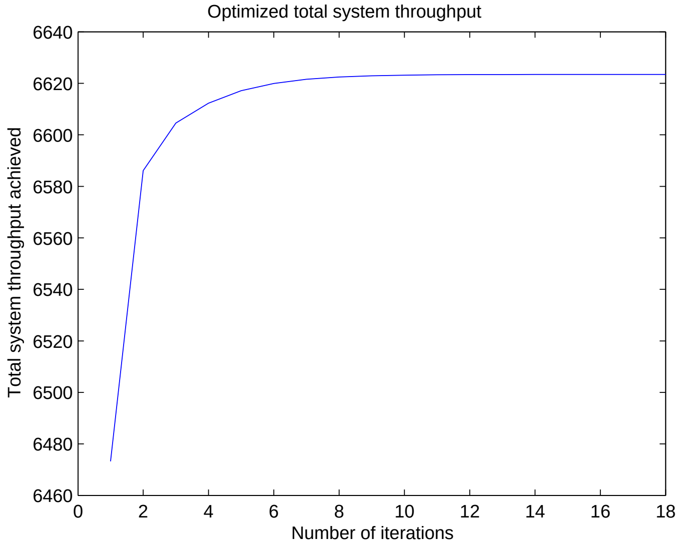

FIG. 4.4. *Convergence of total system throughput (Example 5A).*

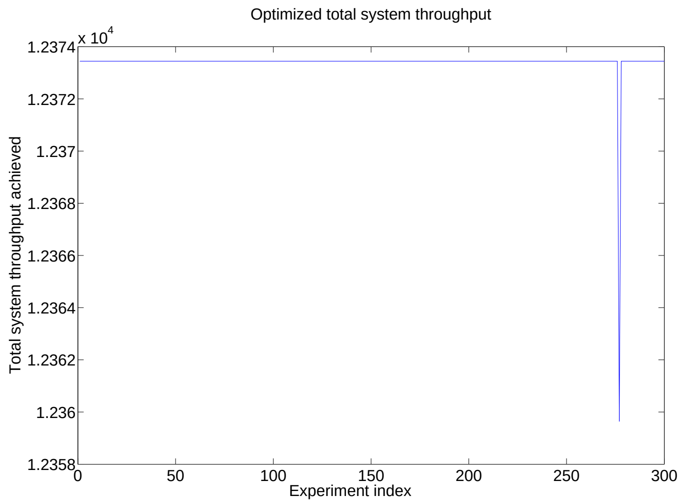

FIG. 4.5. *Maximized total system throughput achieved by the (single) condensation method for 300 different initial feasible vectors (Example 5B).*

where α is a constant slightly larger than 1. At each iteration of a modified SP, the previous computed optimum value is set to constant t_0 and the modified problem (4.6) is solved to yield an objective value that is better than the objective value of the previous SP by at least α. The auxiliary variable t is introduced so as to turn the problem formulation into SP in $(\mathbf{P}, t)$. The starting feasible $\mathbf{P}$ for each modified SP is picked at random from the feasible set, if any, of the modified SP. If we already obtained the global optimal solution in (4.3), then (4.6) would be infeasible, and the iteration of SPs stops.

Example 4 Continued. The above heuristic is applied to the rare instances of Example 4 where solving SP returns a locally optimal power allocation, and is found to obtain the globally optimal solution within 1 or 2 rounds of solving additional SPs (4.6).

5. Distributed Implementation. Another limitation for GP based power control is the need for centralized computation if a GP is solved by interior point methods. The GP formulations of power control problems can also be solved by a new method of distributed algorithm for GP that is first presented in this chapter (see also [19, 7]). The basic idea is that each user solves its own local optimization problem and the coupling among users is taken care of by message passing among the users. Interestingly, the special structure of coupling for the problem at hand (essentially, all coupling among the logical links can be lumped together using interference terms) allows one to further reduce the amount of message passing among the users [19]. Specifically, we use a dual decomposition method to decompose a GP into smaller subproblems whose solutions are jointly and iteratively coordinated by the use of dual variables. The key step is to introduce auxiliary variables and to add extra equality constraints, thus transferring the coupling in the objective to coupling in the constraints, which can be solved by introducing 'consistency pricing'. We illustrate this idea through an unconstrained minimization problem followed by an application of the technique to wireless network power control.

5.1. Distributed algorithm for GP. Suppose we have the following unconstrained standard form GP in $\mathbf{x} \succ 0$:

$$\text{minimize} \quad \textstyle\sum_i f_i(x_i, \{x_j\}_{j \in I(i)}) \tag{5.1}$$

where x_i denotes the local variable of the ith user, $\{x_j\}_{j\in I(i)}$ denote the coupled variables from other users, and f_i is either a monomial or posynomial. Making a change of variable $y_i = \log x_i, \forall i$ in the original problem, we obtain

$$\text{minimize} \quad \textstyle\sum_i f_i(e^{y_i}, \{e^{y_j}\}_{j\in I(i)}).$$

We now rewrite the problem by introducing auxiliary variables y_{ij} for the coupled arguments and additional equality constraints to enforce consistency:

$$\begin{array}{ll} \text{minimize} & \sum_i f_i(e^{y_i}, \{e^{y_{ij}}\}_{j\in I(i)}) \\ \text{subject to} & y_{ij} = y_j, \forall j \in I(i), \forall i. \end{array} \tag{5.2}$$

Each ith user controls the local variables $(y_i, \{y_{ij}\}_{j\in I(i)})$. Next, the Lagrangian of (5.2) is formed as

$$\begin{aligned} L(\{y_i\}, \{y_{ij}\}; \{\gamma_{ij}\}) &= \sum_i f_i(e^{y_i}, \{e^{y_{ij}}\}_{j\in I(i)}) + \sum_i \sum_{j\in I(i)} \gamma_{ij}(y_j - y_{ij}) \\ &= \sum_i L_i(y_i, \{y_{ij}\}; \{\gamma_{ij}\}) \end{aligned}$$

where

$$L_i(y_i, \{y_{ij}\}; \{\gamma_{ij}\}) = f_i(e^{y_i}, \{e^{y_{ij}}\}_{j\in I(i)}) + \Bigg(\sum_{j: i\in I(j)} \gamma_{ji}\Bigg) y_i - \sum_{j\in I(i)} \gamma_{ij} y_{ij}. \tag{5.3}$$

The minimization of the Lagrangian with respect to the primal variables $(\{y_i\}, \{y_{ij}\})$ can be done simultaneously (in a parallel fashion) by each user. In the more general case where the original problem (5.1) is constrained, the additional constraints can be included in the minimization at each L_i.

In addition, the following master dual problem has to be solved to obtain the optimal dual variables or consistency prices $\{\gamma_{ij}\}$:

$$\max_{\{\gamma_{ij}\}} \quad g(\{\gamma_{ij}\}) \tag{5.4}$$

where

$$g(\{\gamma_{ij}\}) = \sum_i \min_{y_i, \{y_{ij}\}} L_i(y_i, \{y_{ij}\}; \{\gamma_{ij}\}).$$

Note that the transformed primal problem (5.2) is convex, and the duality gap is zero under mild conditions; hence the Lagrange dual problem indeed solves the original standard GP problem. A simple way to solve the maximization in (5.4) is with the following update for the consistency prices:

$$\gamma_{ij}(t+1) = \gamma_{ij}(t) + \delta(t)(y_j(t) - y_{ij}(t)). \tag{5.5}$$

Appropriate choice of the stepsize $\delta(t) > 0$ leads to stability and convergence of the dual algorithm [7].

Summarizing, the ith user has to: i) minimize the function L_i in (5.3) involving only *local* variables upon receiving the updated dual variables $\{\gamma_{ji}, j : i \in I(j)\}$ (note that $\{\gamma_{ij}, j \in I(i)\}$ are local dual variables), and ii) update the local consistency prices $\{\gamma_{ij}, j \in I(i)\}$ with (5.5).

5.2. Applications to power control. As an illustrative example, we maximize the total system throughput in the high SIR regime with constraints local to each user. If we directly applied the distributed approach described in the last subsection, the resulting algorithm would not be very practical since it would require knowledge by each user of the interfering channels and interfering transmit powers, which would translate into a large amount of message passing. To obtain a practical distributed solution, we can leverage the structures of power control problems at hand, and instead keep a local copy of each of the *effective received powers* $P_{ij}^R = G_{ij}P_j$ and write the problem as follows (using the log change of variable $\tilde{x} = \log x$):

$$\begin{array}{ll} \text{minimize} & \sum_i \log\left(G_{ii}^{-1}\exp(-\tilde{P}_i)\left(\sum_{j\neq i}\exp(\tilde{P}_{ij}^R) + \sigma^2\right)\right) \\ \text{subject to} & \tilde{P}_{ij}^R = \tilde{G}_{ij} + \tilde{P}_j, \\ & \text{Constraints local to each user.} \end{array} \tag{5.6}$$

The partial Lagrangian is

$$L = \sum_i \log\left(G_{ii}^{-1}\exp(-\tilde{P}_i)\left(\sum_{j\neq i}\exp(\tilde{P}_{ij}^R) + \sigma^2\right)\right) + \sum_i\sum_{j\neq i}\gamma_{ij}\left(\tilde{P}_{ij}^R - \left(\tilde{G}_{ij} + \tilde{P}_j\right)\right),$$

from which the dual variable update is found as

$$\begin{aligned} \gamma_{ij}(t+1) &= \gamma_{ij}(t) + \delta(t)\left(\tilde{P}_{ij}^R - \left(\tilde{G}_{ij} + \tilde{P}_j\right)\right) \\ &= \gamma_{ij}(t) + \delta(t)\left(\tilde{P}_{ij}^R - \log G_{ij}P_j\right). \end{aligned} \tag{5.7}$$

Each user has to minimize the following Lagrangian (with respect to the primal variables) subject to the local constraints:

$$L_i\left(\tilde{P}_i, \left\{\tilde{P}_{ij}^R\right\}_j; \{\gamma_{ij}\}_j\right) = \log\left(G_{ii}^{-1}\exp(-\tilde{P}_i)\left(\sum_{j\neq i}\exp(\tilde{P}_{ij}^R)+\sigma^2\right)\right)+\sum_{j\neq i}\gamma_{ij}\tilde{P}_{ij}^R-\left(\sum_{j\neq i}\gamma_{ji}\right)\tilde{P}_i. \tag{5.8}$$

Some practical observations are in order:

- For the minimization of the local Lagrangian, each user only needs to know the term $\left(\sum_{j\neq i}\gamma_{ji}\right)$ involving the dual variables from the interfering users, which requires some message passing.
- For the dual variable update, each user needs to know the effective received power from each of the interfering users $P_{ij}^R = G_{ij}P_j$ for $j \neq i$, which in practice may be estimated from the received messages, hence no explicit message passing is required for this.

With this approach we have avoided the need to know all the interfering channels G_{ij} and the powers used by the interfering users P_j. However, each user still needs to know the consistency prices from the interfering users via some message passing. This message passing can be reduced in practice by ignoring the messages from links that are physically far apart, leading to suboptimal distributed heuristics.

Example 6. We apply the distributed algorithm to solve the above power control problem for three logical links with $G_{ij} = 0.2, i \neq j$, $G_{ii} = 1, \forall i$, maximal transmit powers of 6mW, 7mW and 7mW for link 1, 2 and 3 respectively. Figure 5.1 shows the convergence of the dual objective function which is also the global optimal total throughput of the links. Figure 5.2 shows the convergence of the two auxiliary variables in link 1 and 3 towards the optimal solutions.

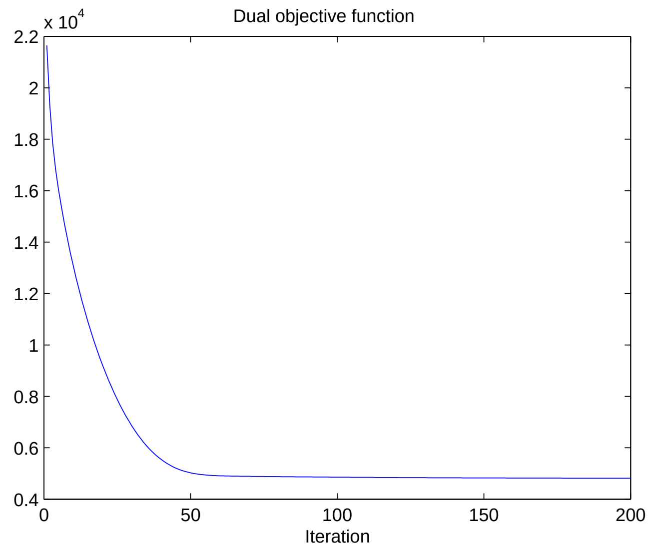

Fig. 5.1. *Convergence of the dual objective function through distributed algorithm (Example 6).*

6. Conclusions. Power control problems with nonlinear objective and constraints may seem to be difficult to solve for global optimality. However, when SIR is much

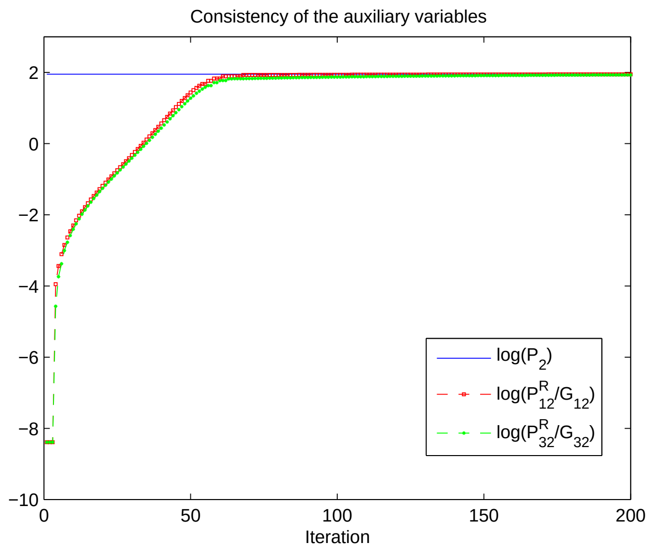

FIG. 5.2. *Convergence of the consistency constraints through distributed algorithm (Example 6).*

larger than 0dB, GP can be used to turn these problems, with a variety of possible combinations of objective and constraint functions involving data rate, delay, and outage probability, into intrinsically tractable convex formulations. Then interior point algorithms can efficiently compute the globally optimal power allocation even for a large network. Feasibility and sensitivity analysis of GP naturally lead to admission control and pricing schemes. When the high SIR approximation cannot be made, these power control problems become SP and may be solved by the heuristic of condensation method through a series of GPs. Distributed optimal algorithms for GP-based power control in multihop networks can also be carried out through message passing.

Sections 4 and 5 present very recent advances overcoming the major limitations of the original GP-based power control methods in [14]. Several interesting research issues remain to be further explored: reduction of SP solution complexity by using high-SIR approximation to obtain the initial power vector and by solving the series of GPs only approximately (except the last GP), combination of SP solution and distributed algorithm for distributed power control in low SIR regime, and application to optimal spectrum management in DSL broadband access systems with interference-limited performance across the tones and among competing users sharing a cable binder.

Acknowledgement. We would like to acknowledge collaborations on GP-based resource allocation with Stephen Boyd at Stanford University and Arak Sutivong at Qualcomm, and discussions with Wei Yu at the University of Toronto.

This work has been supported by Hertz Foundation Fellowship, Stanford Graduate Fellowship, NSF Grants CNS-0417607, CNS-0427677, CCF-0440443, and NSF CAREER Award CCF-0448012.

REFERENCES

[1] M. Avriel, Ed. *Advances in Geometric Programming,* Plenum Press, New York, 1980.

[2] N. Bambos, "Toward power-sensitive network architectures in wireless communications: Concepts, issues, and design aspects." *IEEE Pers. Comm. Mag.*, vol. 5, no. 3, pp. 50-59, 1998.

[3] C. S. Beightler and D. T. Philips, *Applied Geometric Programming,* Wiley 1976.

[4] S. Boyd, S. J. Kim, L. Vandenberghe, and A. Hassibi, "A tutorial on geometric programming", *Stanford University EE Technical Report*, 2005.

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

[6] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: Jointly optimal congestion control and power control," *IEEE J. Sel. Area Comm.,* vol. 23, no. 1, pp. 104-116, Jan. 2005.

[7] M. Chiang, "Geometric programming for communication systems", To appear in *Foundations and Trends of Communications and Information Theory*, 2005.

[8] M. Chiang and S. Boyd, "Geometric programming duals of channel capacity and rate distortion," *IEEE Trans. Inform. Theory,* vol. 50, no. 2, pp. 245-258, Feb. 2004.

[9] M. Chiang and A. Sutivong, "Efficient nonlinear optimization of resource allocation", *Proc. IEEE Globecom,* San Francisco, CA, Dec. 2003.

[10] M. Chiang, A. Sutivong, and S. Boyd, "Efficient nonlinear optimizations of queuing systems," *Proc. IEEE Globecom,* Taipei, ROC, Nov. 2002.

[11] R. J. Duffin, "Linearized geometric programs," *SIAM Review,* vol. 12, pp. 211-227, 1970.

[12] R. J. Duffin, E. L. Peterson, and C. Zener, *Geometric Programming: Theory and Applications,* Wiley, 1967.

[13] G. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence", *IEEE Trans. Veh. Tech.*, vol. 42, no. 4, 1993.

[14] D. Julian, M. Chiang, D. ONeill, and S. Boyd, "QoS and fairness constrained convex optimization of resource allocation for wireless cellular and ad hoc networks." *Proc. IEEE Infocom*, New York, NY, June 2002.

[15] S. Kandukuri and S. Boyd, "Optimal power control in interference limited fading wireless channels with outage probability specifications", *IEEE Trans. Wireless Comm.*, vol. 1, no. 1, pp. 46-55, Jan. 2002.

[16] K. O. Kortanek, X. Xu, and Y. Ye, "An infeasible interior-point algorithm for solving primal and dual geometric programs," *Math. Programming,* vol. 76, pp. 155-181, 1996.

[17] D. Mitra, "An asynchronous distributed algorithm for power control in cellular radio systems", *Proceedings of 4th WINLAB Workshop* Rutgers University, NJ, 1993.

[18] Yu. Nesterov and A. Nemirovsky, *Interior Point Polynomial Methods in Convex Programming,* SIAM Press, 1994.

[19] D. Palomar and M. Chiang, "Distributed algorithms for coupled utility maximization," Working draft, Princeton University, 2005.

[20] C. Saraydar, N. Mandayam, and D. Goodman, "Pricing and power control in a multicell wireless data network", *IEEE J. Sel. Areas Comm.*, vol. 19, no. 10, pp. 1883-1892, 2001.

[21] C. Sung and W. Wong, "Power control and rate management for wireless multimedia CDMA systems", *IEEE Trans. Comm.* vol. 49, no. 7, pp. 1215-1226, 2001.

[22] R. Yates, "A framework for uplink power control in cellular radio systems", *IEEE J. Sel. Areas Comm.*, vol. 13, no. 7, pp. 1341-1347, 1995.